

Sparse Representation of Signals: from astrophysics to real-time data analysis for fusion plasmas and system optimization analysis for ITER and TCV

D.Testa¹, H.Carfantan², M.Albergante³, P.Blanchard¹, S.Bourguignon⁴, A.Fasoli¹, A.Goodyear⁵,
A.Klein⁶, J.B.Lister¹, T.Panis³, and JET contributors*

EUROfusion Consortium, JET, Culham Science Centre, Abingdon, OX14 3DB, UK

- 1) Ecole Polytechnique Fédérale de Lausanne (EPFL), Swiss Plasma Center (SPC), Lausanne, CH
- 2) Université de Toulouse, Institut de Recherche en Astrophysique et Planétologie, Centre National pour la Recherche Scientifique, Toulouse, France
- 3) formerly at the Centre de Recherches en Physique des Plasmas (CRPP), EPFL, Lausanne, CH
- 4) University of Nice Sophia Antipolis, CNRS, Observatoire de la Cote d'Azur, F-06304 Nice, France
- 5) Culham Centre for Fusion Energy, Culham Science Centre, Abingdon, United Kingdom
- 6) formerly at the Plasma Science and Fusion Centre, Massachusetts Institute of Technology, Boston, USA

e-mail contact of corresponding author: duccio.testa@epfl.ch

Abstract.

Efficient, real-time and automated data analysis is one of the key elements for achieving scientific success in complex engineering and physical systems, of which two examples are the JET and ITER tokamaks.

One problem which is common to these fields is the determination of *pulsation modes* from irregularly sampled time-series. To this end, there is a wealth of signal processing techniques that are being applied to post-pulse and real-time data analysis in such complex systems. Here we wish to present a review of the applications of a method based on the *Sparse Representation of Signals*, using examples of the synergies that can be exploited when combining ideas and methods from very different fields, such as astronomy and astrophysics and thermonuclear fusion plasmas.

Examples of this work in astronomy and astrophysics are the analysis of pulsation modes in various classes of stars and the orbit determination software of the Pioneer spacecrafts. Two examples of this work in thermonuclear fusion plasmas are the detection of magneto-hydrodynamic instabilities, which is now performed routinely in JET in real-time on a sub-millisecond time-scale, and the studies leading to the optimization of the magnetic diagnostic system in ITER and TCV.

These questions have been solved formulating them as inverse problems, despite the fact that these applicative frameworks are extremely different from the classical use of *Sparse Representations*, on both the theoretical and computational points of view. Requirements, prospects and ideas for the signal processing and real-time data analysis applications of this method to routine operation of ITER will also be discussed.

Finally, a very recent development has been an attempt at the application of this method to the deconvolution of the measurement of electric potential performed during a ground-based survey of a proto-Villanovian necropolis in central Italy.

PACS Classification scheme: 07.05.Dz, 07.05.Kf, 52.35.Bj, 52.55.Fa, 52.70.Ds.

(1) Introduction.

Efficient, real-time and automated data analysis is one of the key elements for achieving scientific success in complex engineering and physical systems, of which two examples are the JET and ITER tokamaks. There is a wealth of signal processing techniques that are being applied to data analysis in such complex systems, and here we wish to present a review of some examples of the synergies that can be exploited when combining ideas and methods from different fields, such as Astronomy and Astrophysics (A&A) and thermonuclear fusion plasmas.

One problem which is common to these subjects is the determination of pulsation modes from irregularly sampled temporal and spatial series [1]. Historically, this problem has been addressed combining methods based on various forms of the Fourier Transforms (FT) for the time-series analysis, and using methods essentially based around the Lomb-Scargle Periodograms (LSP) [2-5] for the spatial-series analysis. For the latter, much work has been performed to improve on the limitations of the original periodogram methods, essentially in the field of A&A. This general measurement problem is further complicated in thermonuclear fusion plasmas, and specifically in large-scale tokamak and stellarator devices, by the (often very) low number of measurement points in the spatial domain, which is due to in-vessel engineering and installation constraints, leading to a number of mathematical difficulties. Therefore, analysis method based on the spatial Nyquist criterion cannot in general be used because of the effect of aliasing, particularly if intermediate to small wavelengths need to be resolved. For the specific case of magneto-hydro-dynamic (MHD) instabilities in tokamak devices, this deficiency has prompted the development and the successful application of various analysis methods, such as the Singular Value (SVD) [6, 7] and wavelet [8] decompositions, the Wigner [9], Choi-Williams [10, 11] and Hilbert [12] Transforms, and a generalization of the LSP specifically adapted to stellarator devices [12].

A particular sub-class of MHD analysis problems is that of understanding the behaviour of instabilities that are essential for controlling the stability of magnetically confined thermonuclear plasmas. Specifically, the problem of automated real-time detection of MHD modes has now become one of the most important aspects for machine protection and control and optimization of plasma discharges in thermonuclear fusion experiments. The method routinely used for this analysis involves sampling a (usually) rather small set of input signals, such as measurements of magnetic, temperature and density fluctuations, which in most cases are un-evenly sampled in the spatial domain. Appropriate processing of such a set of input data facilitates the detection of the different components in a multi-harmonics spectrum. Furthermore, when the data contains some spatial periodicities, these can be readily used to enhance or eliminate the detection of certain components. A real-time algorithm can then generate a global alarm that is sent to the plant. Under certain specified and pre-determined

operational conditions, this may then trigger a feedback control mechanism. For some examples of these activities, the Readers are referred to Chapter3 and Chapter7 and references therein in [13] and Chapter2 and references therein in [14].

One drawback of most of the current MHD detection and control methods is that they can only detect modes when they have become *unstable* (with a growth rate $\gamma_{\text{GROWTH}}/\omega > 0$, where $\omega > 0$ is the mode's angular frequency), i.e. when they may have already had some detrimental effect on the actual plasma operation and performance. On the other hand, an alternative and innovative method is in use on the JET tokamak [15]. This diagnostic technique combines the active excitation (via a set of in-vessel antennas) of magnetic field perturbations which have a very small amplitude at the plasma edge (maximum intensity $|\delta B_{\text{DRIVEN}}| < 100\text{mG}$, i.e. typically $\sim 10^5$ times smaller than the value of the toroidal magnetic field in JET, $B_{\text{TOR}} \sim 1\text{T}$ to $B_{\text{TOR}} \sim 4\text{T}$) with synchronous real-time detection of the resonant plasma response to such antenna-driven perturbations. This method then allows detecting MHD modes when they are still *stable* (with a damping rate $\gamma_{\text{DAMP}}/\omega > 0$), i.e. before they could have affected the discharge, which is evidently a much more satisfying situation for plasma control and machine protection.

However, none of the methods described above can be efficiently used for the decomposition of a stable spectrum of MHD modes with the aim to measure their damping rate, because of their mathematical limitations and computational requirements, particularly if real-time, sub-millisecond calculations are needed, and when the measurement spectrum is frequency-degenerate, i.e. contains multiple spatial components at frequencies which are separated by less than the damping. On the other hand, a method based on the *Sparse Representation of Signals* (SRS), as implemented in the *SparSpec* code (freeware available at: <http://www.ast.obs-mip.fr/article123.html>) [16, 17] has been demonstrated to efficiently and correctly perform the post-pulse [18] and real-time [19, 20] blind and automated signal decomposition of data which are un-evenly sampled in the spatial domain using a (very) small number of measurement points.

Finally, following the successful application of the SRS to the analysis of actual measurements of different classes of MHD instabilities on the JET tokamak, we have used this method as an “inverse tool” to predict the measurement performance of the ITER array of high-frequency (HF) magnetic probes [21-24]. This has allowed us to comparatively test the measurement performance of different system designs, thus allowing us to propose an optimized diagnostic design [25, 26].

This paper aims at presenting a review as complete as possible of the mathematical properties of analysis methods based on the SRS, particularly focussing on the *SparSpec* code, and of current applications of such methods to A&A and fusion plasma problems, completed with an outlook to problems that could be tackled in the future using this method. As we want the Readers to be presented

clearly and rapidly with the main results of our work, while still being able to then delve into the details if willing to, we present most of the background and reference material separately from the main text using self-contained Appendices.

Starting from existing literature [27], in Section-2 we briefly review the mathematical foundations of the SRS and of the *SparSpec* code (a more detailed review is then presented as the stand-alone Appendix-A). Analysis methods based on the SRS and the *SparSpec* code itself have been used in different A&A applications, and Section-3 provides some examples of these analyses [16, 28-31]. Drawing then from various previously published results [32-40], Section-4 presents a review of the applications of the *SparSpec* algorithm to the analysis of various MHD instabilities observed on the JET tokamak (more details are then presented as the stand-alone Appendix-B). This section is then completed with some previously unpublished results obtained with the *SparSpec* code when analysing magnetic fluctuations corresponding to Tearing Modes (TM) with multiple helicities on the TCV tokamak. Section-5 then presents a review of the application of a method based on the SRS to the optimization of the design of the ITER HF magnetic diagnostic system, as previously discussed in [25, 26] (more details are then presented as the stand-alone Appendix-C). A new topic presented in this section reports on the optimization of the in-vessel positioning of a set of new 3D HF magnetic sensors recently installed in the TCV tokamak. Section-6, as a rather unorthodox conclusion to this review, illustrates the first attempts at a completely new development in the applications of the SRS, which also realizes a youngster's idea of the main author of this paper (DT, when he was still working as an archaeologist during his summers), namely obtaining an efficient and inexpensive interpretation of the ground-based measurements of electric potential made during surveys of one archaeological site in central Italy, the proto-Villanovian necropolis of Poggio della Pozza (dated circa 1200 BC: the main author of this paper (DT) was in fact in charge of the original excavation of this site in 1988, during which the measurements of electric potential were made). Finally, Section-7 presents the conclusions of this work and some prospects for future applications to ITER and other complex engineering and scientific systems such as the Square Kilometre Array (SKA) telescope [41].

(2) Sparse Representations of Signals and the SparSpec code: a brief mathematical review.

The detection and the characterization of oscillations existing in time series are questions raised by many data analysis problems. In A&A observations, the acquisition is often subject to an irregular process of measurement, in particular for ground-based observations: objects of interest can be periodically unobservable due to the alternation of day and night or seasons, and poor weather conditions cause gaps in temporal coverage. In fusion plasmas, the spatial acquisition can similarly be based on irregular sampling, for instance due to engineering and financial constraints. In general

terms, and for any experiment, the sampling time points and spatial positions may be (both) irregularly spaced due to observational constraints. This kind of sampling makes the problem of oscillation frequencies detection and estimation particularly difficult.

Considering for simplicity the general conjugated variables *time* and *frequency* (which in fusion plasmas could represent also toroidal (poloidal) *angles* and toroidal (poloidal) *mode numbers*), this irregular sampling can be seen as the application of an irregular Dirac comb $w(t)$ to the original signal $y_s(t)$ and can be well understood in the Fourier domain:

$y_s(t) = \begin{cases} \sum_{p=1}^P y(t_p)\delta(t-t_p) & \xrightarrow{\text{Fourier Transform}} Y_s(f) = \int_{-\infty}^{\infty} y_s(t)e^{-2j\pi ft} dt = \sum_{p=1}^P y(t_p)e^{-2j\pi ft_p} \\ y(t) \times \underbrace{\sum_{p=1}^P \delta(t-t_p)}_{w(t)} & \xrightarrow{\text{Fourier Transform}} Y_s(f) = Y(f) \star \underbrace{\sum_{p=1}^P e^{-j2\pi ft_p}}_{W(f)} \end{cases} \quad (1)$	(1)
---	-----

In eq.(1), $y(t_p)$ are the individual measurements taken at all the time points t_p (for $p=1, \dots, P$), from which the irregularly sampled signal $y_s(t)$ is constructed through the filtering via the Dirac delta function $\delta(t-t_p)$, and $Y_s(f)$ is the Fourier Transform (FT) in time of $y_s(t)$. $Y_s(f)$ then corresponds to the convolution of the FT of the original signal $Y(f)=FT(y(t))$ with the spectral windows $W(f)$, which is the FT of the irregular Dirac comb $w(t)$, $W(f)=FT(w(t))$.

In the theoretical infinite regular sampling case, the spectral window is the usual Dirac comb and the Fourier transform of the sampled data corresponds to a periodised version of the original signal's FT. This property leads to the well-known Nyquist-Shannon theorem [42] which is not valid in the irregular sampling case.

Hence, the analysis problem becomes that of obtaining a deconvolution of the spectral line data $Y(f)$ from the spectral window $W(f)$. The mathematical modelling for this problem is relatively simple: as the original signal is constituted of a sum of pure frequencies, each data point $y(t_p)$ is expressed as a weighted sum of complex sinusoids, the so-called *atoms*:

$y(t_p) = \sum_{l=1}^L c_l e^{2j\pi \nu_l t_p} + \varepsilon_p,$	(2a)
--	------

where ε_p is the error on the measurement, c_l and ν_l are the complex amplitudes and frequency, and L is the total number of spectral components. The formulation of eq.(2a) presents two problems: first, it is non-linear with respect to the (continuum of) frequencies ν_l , and second, L is unknown a-priori.

The problem described by eq.(2a), which amounts to fitting multiple complex sinusoids to the input data, is a very general signal processing problem which arises in many fields of physics. Such a spectral analysis problem from irregularly sampled data is very common in A&A, where time series acquisition usually suffer from incomplete temporal coverage, in particular periodic gaps caused by the Earth's rotation and revolution, and a-periodic interruptions due to the weather. Many methods have been proposed in the field of A&A to improve the analysis of such irregularly sampled time series, based on generalizations of the LSP [2, 3] and of Data-Compensated Discrete FT [43]. When dealing with data which contain oscillations at several frequency, iterative procedures are generally used [5, 44]. Such methods, however, are inadequate when there are several temporal frequencies and too few measurements. Specific methods have also been developed for short data strings to analyse strictly periodic signals (fundamental frequency ν_0 and harmonics frequencies at an exact multiple of ν_0), such as Phase Dispersion Minimization [45] and string length method [46].

A major simplification [47] of eq.(2a) can be obtained using a discretization of the frequency axis into the known components $f_k=(k/K)*f_{MAX}$, with $k=[-K, \dots, K]$, where f_{MAX} is much larger than the largest frequency component that can reasonably be present in the measurements, leading to:

$y(t_p) = \sum_{k=-K}^K x_k e^{2j\pi f_k t_p} + \varepsilon_p,$	(2b)
---	------

The problem is now linear with respect to the mode amplitude x_k , but we must deal with an even larger number of still unknown $\{x_k\}$ and now known $\{f_k\}$, as one may take that $K \gg L$ to achieve a high resolution analysis. However, the estimation of the spectral lines $\{x_k, f_k\}$ can be greatly simplified by using the *sparsity* of $\{x_k\}$ and $\{f_k\}$, i.e. imposing that the $\{x_k\}$ and $\{f_k\}$ have only a small number of non-zero components. Such a problem can be tackled through the principle of SRS.

Formally, SRS [17, 48, 49] are representations that account for all information in the input data $y(t)$ with a linear combination of a small number of elementary signals (sine waves, Diracs, ...) called *atoms* that belong to a selected family (a *dictionary*) which contains many such elementary signals. The atoms set is a redundant family, i.e. it does not form a basis as the number of atoms ($2K+1$, see eq.(2b)) exceeds the dimension P of the signal space, so any signal can be represented by more than one combination of different atoms. Among all these various possible combinations, the one with the smallest number of atoms is the *Sparse Representation of the Signal*. The sparsity of $\{x_k\}$ can be quantified with the L_0 (pseudo-) norm, which corresponds to the number of non-zero components in $\{x_k\}$: $\|\mathbf{x}\|_0 = \#\{k, |x_k| \neq 0\}$. The *Sparse Approximation of Signals* [50-52] (SAS) is the version of the SRS method adapted to noisy data. Theoretically, the problem tackled with the SAS can also be written as the minimizer of the criterion:

$J_0(x) = \ \mathbf{y} - W\mathbf{x}\ ^2 + \lambda \ \mathbf{x}\ _0,$	(3)
---	-----

where λ is a penalization parameter related to the noise level. However, to minimize this criterion, one must perform a combinatorial optimization, i.e. sift through all possible combinations of elementary signals, which becomes rapidly intractable as the datasets become larger. Two different methods have been proposed to get round this problem. The first one, often called a *Greedy Pursuit* algorithm, iteratively adds atoms to the initial approximation of the signal to improve such approximation [53]. The second one, often called a *Convex Relaxation* scheme, replaces the L_0 pseudo-norm in eq.(3) with another penalization term, generally based on the L_1 -norm $\|\mathbf{x}\|_1 = \sum_k |x_k|$, such that the criterion may be minimized more easily, particularly when considering optimization of the use of CPU time.

In our work we follow this convex relaxation approach, classically using the L_1 -norm, instead of the L_0 pseudo-norm shown in eq.(3). Hence we obtain the criterion:

$J_1(x) = \ \mathbf{y} - W\mathbf{x}\ ^2 + \lambda \ \mathbf{x}\ _1 = \ \mathbf{y} - W\mathbf{x}\ ^2 + \lambda \sum_{k=-K}^K (x_k).$	(4)
--	-----

It can then be easily shown that the criterion of eq.(4) is convex, therefore has no local minima, but, as the number of unknowns may be larger than the number of data points, this criterion is not strictly convex, i.e. the global minimum cannot be a-priori guaranteed to be unique. Moreover, minimizing eq.(4) does not necessarily lead to the same solution as minimizing eq.(3), i.e. sufficient conditions for the equivalence between the L_0 pseudo-norm and L_1 -norm need to be satisfied (see for instance [53, 54, 55]).

Many computationally efficient algorithms have been developed to optimize the solution of eq.(4), some of which can be made compatible with the requirements for real-time analysis of JET data. While for real-valued unknown $\{\mathbf{x}_k\}$ this problem can be written as a classical Quadratic Program, for complex-valued unknown $\{\mathbf{x}_k\}$ it corresponds to a Second-Order Cone Program [49]. An algorithm based on an iterative Block Coordinate Descent procedure has been previously proposed [16, 17], and implemented in the *SparSpec* code. This procedure consists of performing successive one-dimensional minimization steps with respect to each complex-valued unknown \mathbf{x}_k , where each one-dimensional minimization has an explicit solution. This algorithm is very efficient and a correct solution can be typically found in less than 1ms using the rather modest computational resources available to process real-time JET data [19, 20, 27].

Considering now the application of the SRS and SAS methods to magnetically confined thermonuclear fusion plasmas in a tokamak device, the analysis is based on magnetic and turbulence measurements, and typically starts with an initial Fourier decomposition of the data in the time/frequency domain to obtain the individual frequency components $\psi(\omega)$. In a tokamak the plasma column has, to a first approximation, 2D boundary conditions along the toroidal direction and on the poloidal plane, perpendicular to the toroidal direction. The spatial structure of the instabilities is determined by further decomposing each frequency component in its toroidal (n) and poloidal (m) harmonics: $\psi(\omega) = e^{-j\omega t} \sum_{n,m} A_{mn} e^{jn\phi} e^{jm\theta}$. Here ϕ and θ are the toroidal and poloidal angle coordinates, respectively, and we have used the fact that in tokamak geometry one single toroidal component with a given n usually has multiple poloidal components due to toroidicity and various other geometrical effects. The aim of mode number analysis is to determine the value of $\{n, m\}$ of the magnetic instabilities present in the plasma and to estimate their amplitude from data acquired with P detectors un-evenly positioned at angles ϕ_p (θ_p), $p = \{1, \dots, P\}$ being the suffix labelling the individual sensors used for the measurement.

For generality and consistency with the original A&A notation, in the tokamak plasma fusion problem the Fourier conjugated variables can still be called $\{time, frequency\}$, which can then be the actual toroidal $\{\phi, n\}$ or the poloidal $\{\theta, m\}$ conjugated angle and mode-number variables. This means that the mathematical formulation can be equivalently used for calculating the n - (toroidal) and m - (poloidal) mode numbers by using the relevant sensor geometry. For the determination of the poloidal mode numbers in tokamak geometry, one has to remember that we should consider explicitly the so-called θ_* -correction [56-58] to the sensors' position, so as to run the mode number decomposition analysis using the correct, i.e. equilibrium-dependent, sensor geometry.

When applied to thermonuclear plasma physics, the problem described by eq.(2b) has some additional requirements with respect to the A&A problem, which is also described by eq.(2b), even if its solution can still be obtained using eq.(4) in both cases. First, the data are complex-valued, which implies that the Fourier transform of the data does not satisfy the Hermitian property $\hat{y}(-\nu) = \hat{y}^*(\nu)$ as in the spectral analysis of real-valued data. Obviously, the complex-valued data have to be analysed together, conserving the phase relation between in-phase (I) and quadrature (Q) components, and not independently. Second, the mode numbers can only take positive or negative integer values, while in the general spectral analysis problem, frequencies take real values. This is a favourable property as the model (2b) works on a discretized frequency grid. For A&A problems, a very fine discretization of the frequency grid is required so that real valued frequencies are not too distant from the nearest frequency on the grid. Third, in the real time applications we consider for JET (and for ITER), a set of data is acquired every 1ms, therefore the spectral analysis must be completed in an automated

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016
manner in the short time between each measurement acquisition. Finally, in terms of amplitude estimations, it has been shown [16, 17] that minimizing eq.(4) leads to an under-estimation of the amplitudes of the detected mode numbers due to the L_1 -norm penalization term. Thus, an a-posteriori Least-Square (LS) re-estimation of these amplitudes is usually performed for post-pulse analysis in a second step within the calculations, after the modes have been actually detected in real-time.

The Readers are referred to the stand-alone Appendix-A for a complete, more detailed presentation of the mathematical background leading to the SRS method and the implementation of the *SparSpec* code for the analysis of A&A and fusion data.

(3) Applications of the SRS method and of the *SparSpec* code to the analysis of A&A data.

Analysis methods based on the SRS and using the *SparSpec* code have been applied to various A&A problems. Some examples that we will briefly discuss in this Section are the analysis of the radial velocity curves of the Herbig Ae star HD 104237 [16, 59], the pulsation modes of the red supergiant HV2576 in the Large Magellanic Cloud [28], the orbit determination software for the analysis of the anomaly in the trajectories of the Pioneer 10 and Pioneer 11 spacecrafts [29], the estimate of the p-mode frequencies of the solar twin 18 Scorpii [30], and the search for long term periodicities of the X-ray binaries observed with the Swift Burst Alert Monitor [31].

(3.1) Analysis of the radial velocity curves of the Herbig Ae star HD 104237.

The A&A problem consists in the analysis of time-series: these can be, for instance, light curves or radial velocity measurements, which are subject to observational constraints, such as day/night alternation and meteorological conditions. The A&A ground-based measurements are therefore always obtained through irregular sampling. An example of such a data set is provided in fig1, showing the observation for the radial velocity curve of the Herbig Ae star HD 104237, obtained over five observing nights of high resolution spectroscopy at the South African Astronomical Observatory during April 1999 [59, 60].

As for many such pre-main sequence binary stars, the main pulsation modes are dominated by the low-frequency orbital movement caused by the system multiplicity, which could be removed thanks to a second set of observations during April and May 2000. Using a pre-whitening method together with a frequency analysis via the PERIOD98 package (available at www.astro.univie.ac.at/~dsn/) [61], it was possible to detect five dominant oscillations modes between 28.5/days and 35.6/days. These pulsation modes are shown in the top frame of fig2 as the dominant peaks in the FT analysis of the data.

The spectral window of the measurement time series is shown in the bottom frame of fig2, highlighting the very large side lobes at ± 1 day caused by the daily gaps in the observations. The *SparSpec* code was applied to the analysis of the data for HD 104237, and very similar results were obtained for the four main FT frequencies, except for the low frequency pulsations associated to the orbital motion, which shows the robustness of *SparSpec* towards such low-frequency perturbations. Moreover, the pulsation associated to the very large side lobes at ± 1 day could be easily removed from the FT spectrum. However, the fifth frequency detected by *SparSpec* at $f_5=34.6/\text{days}$ [16] differs quite significantly from the one originally detected at $f_5=33.862/\text{days}$ [59]. To determine the correctness of either finding, simulations were run using *SparSpec* and the pre-whitening method using the HD 104237 data time sampling and the two sets of spectral components previously obtained with the *SparSpec* and the PERIOD98 FT analysis. This demonstrated that *SparSpec* would have been able to correctly find the f_5 component from the PERIOD98 FT analysis [59], had that been in the measurement dataset, while the pre-whitening method would have failed to find the f_5 component from *SparSpec*. This shows that the five main pulsation frequencies identified with the *SparSpec* algorithm is a more reliable result than that obtained with the PERIOD98 FT analysis.

(3.2) The pulsation modes of the red supergiant HV2576 in the Large Magellanic Cloud.

Red supergiants are rather massive He-burning stars, with a mass between 10 and 30 times the mass of the Sun. One such example is the Harvard Variable HV2576, a red supergiant in the Large Magellanic Cloud. This star has a very complex light variation: when analysed using a simple linear sine fitting method, it shows a periodicity of around 530 days [62]. This single pulsation mode however does not fit very well the measurements taken from a database irregularly spanning 8 years of astronomical observations. When fitting the original photometric measurements with several other algorithms, such as those based on the PERIOD04 package [63], the Phase Dispersion Minimization method [64] and *SparSpec*, two distinct oscillation periods appear [28]. The first one corresponds to a pulsation mode with a period of 525 days, very close to the original single-mode result [62]. The second pulsation mode has a period of 261 days, its presence helping to provide a much better fit to the measurements. Moreover, this second pulsation mode is slowly varying in time, indicating the presence of huge convection cells that interplay with the first two pulsation modes or the presence of a much longer periodicity in this red supergiant.

(3.3) The analysis of the Pioneer anomaly.

The Pioneer anomaly refers to the difference between the trajectories of the Pioneer-10 and Pioneer-11 spacecrafts computed using the standard relativistic model for the gravitational force and their

actual trajectories as observed through irregular time sampling with Doppler tracking by the NASA Deep Space Network. These differences only became clearly observable after the spacecrafts had moved beyond 20 AU on their way out of the solar system. This difference has been described as due to a constant anomalous red-shifted acceleration directed towards the Sun of around $(8.74 \pm 1.33) \times 10^{-8} \text{cm/s}^2$ [65]. For many years, and before rather mundane engineering causes related to the anisotropic radiation pressure produced by the spacecrafts heat losses could be found [66], the Pioneer anomaly was the subject of a large number of studies as a possible indication of the requirement for a correction of the standard relativistic model for the gravitational force.

One such study [30] employed the *SparSpec* algorithm, together with many other methods, to analyze the residuals of the fit to the Doppler tracking data, after processing them with the secular term provided by [66], with the aim of finding pulsation modes that could then be linked to physical causes not present in the standard relativistic model for the gravitational force. The *SparSpec* analysis indicates the presence of significant pulsation modes in the residual to the fit of the Pioneer anomaly with periods corresponding to half a sidereal day, one sidereal day and half a sidereal year. These pulsation modes were then included in the orbit determination software for the Pioneer spacecrafts to better track its position over time [30], and were considered to be consistent with the effects on the propagation of light from the spacecraft to the observer on Earth due to metric corrections in the framework of general relativity or to an incorrect modelling of the solar corona.

(3.4) The estimate of the p-mode frequencies of the solar twin 18 Scorpii.

Solar twins is a terminology used to indicate stars that are spectroscopically very similar to the Sun. These objects have been the focus of a significant number of photometric and spectroscopic studies in recent years, with the interest in these astronomical objects stemming from trying to answer two main questions: first, what is the “best” solar twin, a quest that could then indicate the location of a system capable of supporting exoplanets similar to the Earth, and second, is the Sun really unique?

One of the main analysis tools for solar twins is astero-seismology, which can provide insights into the un-observable stellar interior from the measurement of pulsating modes in their light emission spectra. As an example, the measured frequencies of pressure-driven modes (p-mode frequencies) provide information on the internal pressure fluctuations in the solar twin, the dynamics of these modes being determined by the internal hydro-dynamical stability and thus the radial profile of the speed of sound: p-modes are very common in Sun-like stars, such as Alpha-Centauri. Similarly, gravity-driven (g-)modes are produced by buoyancy, and thus provide information on the radial profile of the mass density: these modes are essentially confined to the core of the star’s convection system, which converts thermal energy into the kinetic energy of the modes, and are often observed

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016 in white dwarfs. Finally, surface (f-)modes are the equivalent of tidal waves on the surface of the ocean, providing insights on the perturbation at the surface of the star due to centrifugal effects or the presence of a binary sister star nearby. Detailed and accurate analysis of multiple periodicities in their light emission spectra through irregularly sampled time series has therefore become a major tool for the understanding of the internal structure of solar twins.

The *SparSpec* code has been applied, together with other analysis methods, to the deconvolution of the emission spectra of the solar twin 18 Scorpii [30], a solitary star in the Scorpius constellation at approximately 14 parsecs from the Earth [67]. For this study, 100 irregularly-sampled time series were constructed using an arbitrary number of input frequencies so as to match the observations, and thus deduce the accuracy of the Bayesian method used for the analysis of the photometric measurements, a Maximum A Posteriori (MAP) approach obtained through a Markov-Chain Monte Carlo estimation, where the likelihood is replaced by a Posterior Probability Density. Figure 3 shows the results of this analysis: of the original 52 pulsation modes detected with MAP in the frequency range between $\sim 2000\mu\text{Hz}$ and $\sim 3800\mu\text{Hz}$, only 21 of them sitting in the frequency range between $\sim 2800\mu\text{Hz}$ and $\sim 3400\mu\text{Hz}$ are deemed to be sufficiently reliable when comparing with the results from *SparSpec* and CLEAN [68], and therefore only these modes are used in the ensuing asteroseismology studies of 18 Scorpii. The identification of each pulsation mode is then accomplished straightforwardly by comparing to well-established solar pulsation modes.

(3.5) The search for long-term periodicities in X-rays binaries.

The detection of multiple pulsation modes in the light curves emitted by X-ray binaries is a very important tool to understand the properties of these systems. Periodic variations, ranging from a few hours to hundreds of days, have indeed been observed in the light curves of several low-mass and high-mass X-ray binaries. These sources have been monitored for over 10 years now using two different systems: the Swift Burst Alert Monitor [69] (S-BAT), which observes more than 50% of the sky each day, and the Rossi X-rays Timing Explorer [70] (RXTE), part of the All-Sky-Monitor system, both providing long term light curves of several hundreds of sources.

A selection of data from the S-BAT and RXTE systems have been analysed in [31] using the *SparSpec* code, and multiple orbital and super-orbital periodicities have been detected. While the orbital motion is a natural candidate for several pulsating modes, super-orbital periodicities have also been detected, with the likely explanation being the presence of an unstable, pre-cessing and tilted accretion disk. Using the same analysis method (*SparSpec*) to compare the pulsation modes more recently obtained using the S-BAT with those previously obtained from the RXTE data, this in itself being a feat not

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016 often accomplished in A&A work, the work reported in [31] was able to discover new periodicities and to confirm previously reported one.

(4) Applications of a Sparse Representation method and of the SparSpec algorithm to the analysis of magnetic fluctuations in JET tokamak plasmas.

The main application of *Sparse Representation* methods and of the *SparSpec* code on JET has been the real-time and post-pulse analysis of Alfvén Eigenmodes [71-74]. Alfvén Eigenmodes (AEs) are a particularly important example of real-time mode detection and tracking in thermonuclear fusion experiments for two essential reasons. First, these waves are a natural Eigenmode of any magnetically confined plasma: their frequency F_{AE} is simply proportional to the ratio between the magnetic field and the square root of the plasma mass:

$F_{AE} = (\text{multiplier}) \times \frac{B_{TOR}}{4\pi R_{RES} q_{RES} \sqrt{m_p \sum_i n_i A_i}},$	(5)
---	-----

and thus represents the balance between the tension force of the ambient magnetic field lines and the plasma inertia. In eq.(5) B_{TOR} is the toroidal magnetic field, n_i and A_i are the density and atomic mass of all ion species, m_p is the proton mass, $q_{RES}(r_{RES})=(2m+1)/2n$ is the value of the safety factor at the mode resonant position $R_{RES}=R_0+r_{RES}$, where R_0 is the magnetic axis position and r the minor radius. The quantity *multiplier* defines which class of AEs is being investigated: *multiplier*=1 is used for Toroidal AEs (TAEs) and *multiplier*=2 for ellipticity-induced AEs. The analysis of the dispersion relation of AEs can thus provide unique information on the plasma isotopic composition, the safety factor profile and the toroidal rotation frequency via the Doppler shift in the AE mode frequency for different toroidal mode numbers [75-77]. Second, the fusion-born alpha particles (α s) have a supra-thermal speed at birth that is typically well above the Alfvén speed in the usual thermonuclear tokamak plasma conditions. Resonant interaction with AEs is the first wave-particle interaction encountered by the α s during their thermalization process: hence, this mechanism for phase-space and spatial diffusion needs to be appropriately monitored and controlled to guarantee good confinement of the α s themselves [78, 79].

While in JET the measurements of unstable AEs, i.e. with a negative imaginary component $\gamma < 0$ of the mode frequency $\omega = \omega_{AE} + i\gamma$, are obtained using standard high-frequency magnetic diagnostic systems, the measurements of stable AEs, i.e. with a positive $\gamma > 0$, are obtained using the so-called Alfvén Eigenmodes Active Diagnostic (AEAD) system [15], as its original (and still currently predominant) aim is indeed that to drive and detect AEs. This diagnostic system works on the principle of active excitation of low-amplitude magnetic perturbations using in-vessel antennas, with

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016

maximum driven intensity at the plasma edge $|\delta B_{\text{DRIVEN}}| < 100 \text{mG}$, i.e. typically $\sim 10^5$ times smaller than the value of the toroidal magnetic field in JET, $B_{\text{TOR}} \sim 1 \text{T}$ to $B_{\text{TOR}} \sim 4 \text{T}$. Active excitation is then combined with synchronous real-time detection of the resonant plasma response to such antenna-driven perturbations, which provides the capability of measuring only the plasma response at the frequency corresponding to the antenna excitation. The AEAD real-time controller, the Alfvén Eigenmodes Local Manager (AELM), works on a millisecond time scale and constitutes one essential and furthermore worldwide unique component of the JET Real Time Data Network. The measurement of the mode characteristics, such as the frequency, amplitude, toroidal mode number and damping rate, are obtained in real-time through calculations performed on a sub-millisecond time scale by the AELM software. This data is then passed to the Real Time Signal Server [80], which in principle allows implementing a real-time control and feedback system for the modes detected with the AELM by measuring the distance from the marginal stability limit $\gamma/\omega=0$, and calling for a reaction of the plant when the plasma is approaching the limit during the discharge. More details on the technical implementation of the AELM hardware infrastructure and software can be found in [27].

The *Sparse Representation* method and the *SparSpec* code have been extensively used to obtain results on the dependency of the mode frequency and damping rate for stable AEs on various background plasma parameters, and these results have been presented previously [32-38]. Hence, in this Section we focus our attention to a brief review of the analysis capabilities of the *SparSpec* code and on newer data that may open novel applications of such method on JET (and ITER), specifically in view of the on-going upgrade of the AEAD system [81-83]. The Readers can find additional details on the use of the *SparSpec* code and its accuracy for the analysis of magnetic fluctuation data in tokamak fusion experiments in the stand-alone Appendix-B.

(4.1) Real-time and post-pulse measurement of the TAE mode frequency, amplitude and damping rate using the SparSpec algorithm.

Figure4 shows an overview of the measurement of the damping rate, mode frequency and mode amplitude for TAEs with different toroidal mode numbers. The measurements were obtained in JET using the AEAD system for the He4 discharge #79216, and these results were obtained using the post-pulse implementation of the *SparSpec* algorithm. In addition to the TAE data, fig4 also shows some representative plasma background parameters (electron density and temperature, measured with a high-resolution Thomson Scattering diagnostic system; safety factor, obtained combining EFIT [84] reconstruction with Motional Stark Effect (MSE) and polarimetry measurements), the value of the antenna driving frequency and the value of the central frequency of the $n=1$ TAE gap computed as in eq.(5) with $R_{\text{RES}}=3\text{m}$, $q_{\text{RES}}=1.5$, $\sum_i n_i A_i = 4n_{e0}$, and using the real-time and post-pulse measurement of

the toroidal magnetic field and of the electron density line-integrated along a chord passing through the plasma centre. Figure4 shows that many different toroidal mode numbers are present in the frequency-degenerate spectrum of stable TAEs, and these modes have amplitude and damping rate that can easily vary by up to a factor 10 at any given time point. Hence, an accurate real-time mode number discrimination is clearly an essential ingredient for obtaining damping rate and (at least relative) amplitude data that could meaningfully be used for plasma control purpose.

The main differences between the real-time and post-pulse implementations of the *SparSpec* algorithm are described in details in [27], and are basically due to the different computing resources that are available to perform these two sets of analyses (real-time: 1GHz PowerPC with 512MB RAM; post-pulse: 2.4GHz laptop with 8GB of RAM). Moreover, for real-time calculations there is a *hard* CPU time limit of 850 μ sec, as the AELM works on a 1kHz clock rate: the results of missing a deadline (i.e. not completing the calculations within the allocated time frame) are classed as a failure, leading to a full stop of the experiment in order to avoid any potential damage to the AEAD system or to the JET machine itself. Post-pulse processing does not suffer of this CPU time limit, hence allows for a much greater flexibility and scope in the analysis.

Figure5 shows the comparison between the real-time (RT) and post-pulse (PP) measurements of the mode frequency, amplitude and damping rate for some representative toroidal mode numbers for the JET discharge #77417. The post-pulse analysis was performed using $\lambda_{\text{NORM}}=0.65$, a normalized quantity in the range $\lambda_{\text{NORM}}=\lambda/\lambda_{\text{MAX}}=0\rightarrow 1$ related to the penalization parameter λ of eq.(4), and $|f_{\text{MAX}}|=150$, the size of the mode-number dictionary, whereas for the real-time analysis we set $\lambda_{\text{NORM}}=0.85$ and $|f_{\text{MAX}}|=20$ to satisfy the corresponding CPU and RAM limits. The data were evaluated using the RT and PP implementation of the *SparSpec* algorithm, and for presentation we selected both low- n and high- n modes. We note that the RT and PP measurement of the mode frequency for the different mode numbers shown in fig5 almost exactly overlap, with a difference that is typically less than 100Hz: this is essentially due to the accuracy of the synchronous detection system. Similarly, the damping rate measurement follows almost exactly the same trends in real-time and post-pulse, and the discrepancy in their absolute value is usually well below 15% when the temporal evolution of the mode was well tracked in real-time. On the other hand, the measurement of the absolute value of the mode amplitude is only correct in real-time within a factor 3 to 10, although the trends are sufficiently well reproduced. This is essentially due to the fact that the LS renormalization of the output amplitudes required by *SparSpec* cannot be implemented in real-time due to the limitations in the available CPU and RAM resources.

In summary, the representative TAE measurements shown in fig4 and fig5 demonstrate that the *SparSpec* algorithm fully satisfies the requirement for accurate mode detection and discrimination

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016 not only post-pulse, where in principle “unlimited” computational capabilities are available, but also, and more importantly, in real-time when using more limited computational resources. The next step in JET would be to upgrade the AELM hardware so that even more accurate calculations could be performed in real-time, and then implement a feedback loop where appropriate actuators are activated when the damping rate approaches the marginal stability limit $\gamma/\omega=0$ while the mode amplitude exceeds a certain threshold. As an example of possible actuators, and specifically for the case of AEs, a change by less than 3% in the edge elongation (κ_{95}) in the range $1.45 < \kappa_{95} < 1.55$ has been experimentally observed for JET plasmas to cause an increase in the mode damping from below 0.3%, hence very close to the marginal stability limit, to above 1%, hence much further away from the marginal stability limit, with no other effect on the overall plasma confinement [32, 98]. Such a change is obtained through a variation by less than 5% in the voltage applied to some of the shaping coils, which could then be used as an actuator to control the stability of AEs. Other possible actuators for the control of the stability of AEs are the launchers for the Electron Cyclotron Resonance Heating (ECRH) and Current Drive (ECCD) systems, as recently demonstrated in experiments on DIII-D [99].

(4.2) Measurement of the mode location of unstable TAEs.

The *SparSpec* algorithm can also be used to determine the toroidal and poloidal (m) mode number of unstable TAEs, driven by a population of energetic ions. An example of these measurements is shown in fig6 for the JET discharge #49384, where we want to compare the mode radial location as determined from cross-correlation analysis and from the resonant condition $q_{RES}(r_{RES})=(2m+1)/2n$.

For these studies, we compare the post-pulse *SparSpec* toroidal and poloidal mode number results (fig6a) with those obtained using a Least Square Fit (LSF) algorithm (fig6b). While with *SparSpec* we obtain multiple components at any given time and frequency point, the LSF algorithm provides us only with what is calculated to be the dominant (toroidal, poloidal) mode number. The *SparSpec* analysis was performed using $\lambda_{NORM}=0.65$ and $|f_{MAX}|=150$ for the toroidal mode number analysis, and $\lambda_{NORM}=0.35$ and $|f_{MAX}|=300$ for the poloidal mode number analysis, respectively. Figure6c shows the mode radial location determined as the Eigenfunction peak using $q_{RES}(r_{RES})=(2m+1)/2n$ and cross-correlation analysis between one magnetic probe located and the plasma edge and ECE and reflectometry measurements viewing various radial positions. Note that to determine the q_{RES} value, we only use the dominant component provided by the *SparSpec* analysis.

To first approximation, we find that both the *SparSpec* and LSF algorithms provide an evaluation of r_{RES} which is consistent with the cross-correlation data. However, at a closer inspection we actually see that the trend, indicated by the cross-correlation measurements, of r_{RES} moving towards the magnetic axis as the current profile relaxes and the value of the q -profile drops across the plasma

cross-section, is only well reproduced when using the *SparSpec* q_{RES} data. Conversely, the LSF data produce (even if relatively small) jumps in q_{RES} that are not consistent with the cross-correlation data. The reason for this difference is straightforward: whereas the LSF algorithm effectively forces a best (in the LS sense) fit of the magnetic measurements with one single component, the *SparSpec* algorithm allows multiple components. The dominant one is only selected a-posteriori to obtain q_{RES} , and this turns out to be a more accurate processing method.

(4.3) Measurement of turbulence suppression by fusion-born alpha particles.

The Deuterium-Tritium Experiment (DTE1) performed in JET in 1997 [85] produced the world record fusion power $P_{FUS} \approx 16 \text{ MW}$, with a record fusion energy gain $Q_{DT} \approx 0.65$ maintained over about half the plasma energy confinement time τ_E . One of the main purposes of this experiment was that to verify the plasma self-heating by the fusion-born α s: this process requires the α s first to thermalize on the electrons, on a time scale $\tau_{\alpha e}$ that is in general comparable with τ_E , and then the electrons are required to heat the ions through energy equi-partition, occurring over a time scale τ_{ei} that is around five to ten times longer than τ_E . This mechanism for the plasma self-heating by the fusion-born α s was fully verified during the DTE1 experiment [86, 87]. However, and totally unexpectedly from a theoretical point of view, at the time of the DTE1 experiment it was also noted that under certain experimental conditions a thermal ion heating was obtained that was much larger than what could be predicted, and furthermore occurring over time scales even shorter than τ_E . This unexpected observation of an anomalous ion heating in the presence of a minority population of fusion-born α s has only been very recently explained in terms of turbulence suppression in the Ion Temperature Gradient (ITG) channel by the α s themselves [39, 40]. This explanation has been obtained by combining methods that have only recently become available, such as the *SparSpec* algorithm for the spectral analysis of the magnetic and turbulence measurements and the GENE code [88] for the numerical simulation of turbulence in the ion and electron channels in the presence of multiple and non-thermal ion species.

Figure7 shows a summary of the spectral measurements of the turbulence in the ion acoustic frequency range, presented as function of the toroidal mode number for three discharges with different densities of α s, and for two phases in the discharge: before and after full thermalization of the α s. These results were obtained using the post-pulse implementation of the *SparSpec* algorithm, with $\lambda_{NORM} = 0.15$ and $|f_{MAX}| = 500$. The very small value of λ_{NORM} and high value of $|f_{MAX}|$ are both needed as the turbulence spectra are incoherent, in principle could extend up to very large toroidal mode numbers (i.e., down to a wavelength smaller than the ion Larmor radius), and have an amplitude

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016 which may sit just above the background noise level. Data for negative and positive mode numbers (in the plasma moving frame, i.e. with the spectrum corrected for the background toroidal rotation) correspond to turbulence in the electron and ion channels, respectively Trapped Electron Modes (TEM) and Ion Temperature Gradient (ITG) driven modes. For the discharge #41069, which does not have α_s , the data have been analysed at the same two time points used for the discharges #42847 and #43011, which had α_s . Whereas in #41069 the turbulence spectra are very similar at these two time points, in #42847 and #43011 we note a large increase in the TEM turbulence as the α_s thermalize. More importantly, there is a factor ~ 2 to ~ 5 suppression of turbulence in the ITG channel when the α_s have not yet thermalized, with the larger reduction occurring in the discharge #42847, which has a larger fraction of α_s . Simulations performed with the GENE code indicate that the intensity $|\delta B_{MEAS}|$ and growth rate γ_{ITG} turbulence in the toroidal mode number range $65 < n < 120$, which is identified in the simulations as ITG modes, both decrease as the α_s start thermalizing on the background plasma, as indicated by the turbulence measurements. This then allows the core ion temperature T_{i0} to increase on a time scale which is comparable to the energy confinement time $\tau_E \sim 0.8$ sec, as measured, but is much faster than the α_s slowing-down time on the ions ($\tau_{\alpha i} \sim 4$ sec) and the electron-ion energy equipartition time ($\tau_{ei} \sim 5$ sec). Again consistently with the turbulence and ion temperature measurements, the increase in T_{i0} drives an increase in $|\delta B_{MEAS}|$ and γ_{ITG} as the fusion born α_s start to fully thermalize, which prevents a further increase in T_{i0} itself.

(4.4) Analysis of Tearing Modes with multiple helicities in the TCV tokamak.

One example of application of the *SparSpec* algorithm on the TCV tokamak is the analysis of Tearing Modes (TM) at multiple mode frequencies and with multiple helicities m/n , and the ensuing estimation of the island width from the measured magnetic data at the plasma edge. One such case is the TCV discharge #48799 where two TMs are observed experimentally: the first a pure $m/n=2/1$ mode at around 4kHz, the second a $m/n=4/2$ mode at around 8kHz, with a significant sideband $m/n=5/2$ component, providing approximately 1/3 of the mode energy (evaluated as $\propto |\delta B_{MEAS}|^2$).

For these TMs, the island width has been evaluated using magnetic measurements obtained through Mirnov coils mounted on the vacuum vessel wall following the method presented in [89]. First, we evaluate the different (m,n) helical harmonic components of the perturbed surface current density along the resonant flux-surface $\delta \mathbf{j}_{SUR} \sim \sin(m\theta_* - n\phi) \mathbf{B}/RB_\theta$, where θ_* and ϕ are the poloidal (in straight-field line coordinates) and toroidal angles, respectively, \mathbf{B} and B_θ are the total magnetic (vector) field and its poloidal component, respectively, and R is the major radius position of the resonant q -surface determined by $q_{RES} = m/n$. This 3D problem can then be treated with a simpler 2D model by

considering only the toroidal component of $\delta \mathbf{j}_{\text{SUR}}$. This allows evaluating the perturbation $\delta B_{\theta, \text{MEAS}} = (b_{\text{MS}} + b_{\text{MV}} c_{\text{VS}}(\omega)) I_{\text{S}}$ in the poloidal magnetic field at the sensors' location, where b_{MS} and b_{MV} represent the coupling coefficients between the different toroidal filaments in the perturbed helical surface current I_{S} (subscript "S"), the vessel eddy current I_{V} (subscript "V") and the sensors (subscript "M"), respectively, and $c_{\text{VS}}(\omega)$ represents the frequency-dependent vessel response to the perturbation caused by I_{S} . The "best" mode at any one frequency, or the "best" superposition of modes at different frequencies, can then be identified fitting the model poloidal magnetic field perturbation to the Mirnov coil measurements. The perturbed helical flux at each plasma position x_{P} produced by the perturbed helical current at the q-surface resonant position x_{S} can then be obtained from this "best" (superposition of) mode(s), and it is given by $\delta \psi(x_{\text{P}}) = M_{\text{PS}} I_{\text{S}}(x_{\text{S}}) |\mathbf{B}| / B_{\phi}$, with $I_{\text{S}} = \sum_k \alpha_k I_{\text{S}}(m_k, n_k, x_{\text{S}})$, where M_{PS} is the mutual inductance between the toroidal current filaments located at x_{P} and x_{S} and B_{ϕ} is the toroidal magnetic field component. The equilibrium flux is modulated around the resonant surface by $\delta \psi(x_{\text{P}})$: the X-points (respectively O-points) of the islands are then associated with the maxima (minima) of this modulation. The island width is then simply given by half the separation between the two X-points.

Figure 8 shows three different estimation of the island width for the TCV discharge #48799 using the magnetic measurements and the procedure described above. The first estimate is obtained when considering only a pure $m/n=2/1$ mode at around 4kHz (no significant sidebands were detected at this frequency using *SparSpec*), the second summing up this mode and a pure $m/n=4/2$ mode at around 8kHz, and the third including the contribution of the sideband $m/n=5/2$ component at 8kHz obtained through the *SparSpec* analysis. We note that by including this last term we obtain a larger estimate for the island width, increasing from $\sim 1.8\text{cm}$ to $\sim 2.2\text{cm}$ over the time window of interest. Although relatively small, as it amounts to only about $\sim 20\%$ of the single-helicity width, this increase in the island width estimation for TM with multiple helicities could not have been detected without allowing for multiple modes at the same frequency, and it is this latest method which is now currently used in analyses of TM onset and stabilization for TCV plasma [90-92].

(5) Optimization of the design of the ITER HF magnetic diagnostic system.

The ITER HF magnetic diagnostic system is intended to provide measurements of MHD modes and magnetic fluctuations with magnitude as low as $|\delta B_{\text{MEAS}}| \sim 10^{-4}\text{G}$ (as measured at the position of the sensors) and up to frequencies $> 300\text{kHz}$, with toroidal and poloidal mode numbers up to $|n|=30$ and $|m|=60$ [22-24]. Figure 9a shows the baseline design for this system as in 2009, built around 2 main arrays of $2 \times 18 = 36$ equi-spaced sensors each on the low-field side of the machine for toroidal mode number detection (indicated by the filled dots in fig 9a) and 6 main arrays of 16 un-evenly spaced

sensors each for poloidal mode number detection (indicated by the filled squares in fig9a), covering the entire poloidal cross-section but for the divertor region. The toroidal and poloidal mode number detection systems can be improved in the ITER original design layout, initial provisions already having been made for this, by adding a number of high-resolution mini-arrays on the horizontal (for toroidal analysis) and vertical (for poloidal analysis) edges of some of the equatorial ports on the low-field side, as shown in fig9b.

We have performed [25, 26] the baseline analysis and an attempt at optimization of the ITER HF magnetic diagnostic system using an approach based on the SRS method. The *SparSpec* algorithm has been applied to a model dataset of input modes for various implementations of the ITER HF magnetic sensor geometry for n(m)-number detection. This analysis is performed by scanning various parameters for the input mode spectrum: number of components with their relative random amplitude, relative phase and mode number, and standard deviation in the background Gaussian (white) noise. The ITER measurement requirements and the expected measurements' errors and tolerances are explicitly considered in our method to define the *correct* and the *wrong* detection of the input modes. We then consider a measure of the projected financial costs required for each implementation of the ITER HF magnetic diagnostic system. The ratio between the confidence level in the measurement performance, i.e. how correctly the input mode spectrum is detected, and the financial costs necessary to achieve this performance, is then taken as the measure of the cost-normalized measurement performance. By then averaging over all the scans performed to study a given implementation of the ITER HF magnetic diagnostic system, we obtain an assessment of the overall system performance, one where we have integrated physics and budgetary requirements: the highest ratio defines the cheapest (financially) diagnostic system that allows obtaining a satisfactory measurement performance. The Readers can find additional details on the use of the *SparSpec* code and the method we have employed for the analysis of the measurement performance of the ITER HF magnetic diagnostic system in the stand-alone Appendix-C.

It is now important to introduce the nomenclature that we use for this analysis. We use the wording *geometry* to define a specific method to select the number and position of the magnetic sensors. For instance, one geometry is made up with 36 un-evenly spaced sensors, and a second geometry still has 36 un-evenly spaced sensors in total, but of these 12 are installed in one high-resolution array in one of the equatorial ports. For each geometry, the actual position of each sensor is selected either ad-hoc (for evenly spaced sensors), or through a pseudo-random algorithm (for un-evenly spaced sensors) that takes into account all potential installation constraints, such as zones where sensors cannot be located (for instance the divertor region in the case of a poloidal array). This is particularly important for the geometries used for poloidal analysis, as only the equatorial port can be used to add a single

high-resolution array, and no sensors can be positioned in the divertor region or in the lower and upper mid-plane ports. Because of these installation constraints, the position of each sensor cannot be truly randomly selected. Each geometry is then represented by a number of different *sensors' arrangements*. Each geometry using un-evenly spaced sensors is represented by at least 10 arrangements of sensors, corresponding to the different pseudo-random number realizations used to construct the sensors' position. For geometries using evenly-spaced sensors, for instance those used for toroidal analysis, different arrangements can be obtained when changing the equatorial port(s) where the high-resolution array(s) are installed. In total, we consider 19 different geometries for the ITER HF magnetic diagnostic system, as summarized in Table1.

geometry	sensor spacing type	high-resolution arrays	mode number analysis
ES1	evenly spaced sensors	none	toroidal + poloidal
ES2	evenly spaced sensors	1x5 sensors	toroidal + poloidal
ES3	evenly spaced sensors	1x7 sensors	toroidal + poloidal
ES4	evenly spaced sensors	1x12 sensors	toroidal + poloidal
ES5	evenly spaced sensors	3x5 sensors	toroidal only
US1	un-evenly spaced sensors	none	toroidal + poloidal
US2	un-evenly spaced sensors	1x5 sensors	toroidal + poloidal
US3	un-evenly spaced sensors	1x7 sensors	toroidal + poloidal
US4	un-evenly spaced sensors	1x12 sensors	toroidal + poloidal
US5	un-evenly spaced sensors	3x5 sensors	toroidal only
BT1	baseline design toroidal system	none	toroidal only
BT2	baseline design toroidal system	1x5 sensors	toroidal only
BT3	baseline design toroidal system	1x7 sensors	toroidal only
BT4	baseline design toroidal system	1x12 sensors	toroidal only
BT5	baseline design toroidal system	3x5 sensors	toroidal only
BT6	baseline design poloidal system	none	poloidal only
BT7	baseline design poloidal system	1x5 sensors	poloidal only
BT8	baseline design poloidal system	1x7 sensors	poloidal only
BT9	baseline design poloidal system	1x12 sensors	poloidal only

Table1. Overview of the main characteristics of the 19 different geometries used for the baseline analysis and optimization of the ITER HF magnetic diagnostic system.

Ten of these geometries are labelled as ES# and US# to indicate assemblies with a varying number of evenly (ES) and un-evenly (US) spaced sensors, as follows: #1: no high-resolution sensors; #2: adding 1x5 high-resolution sensors; #3: adding 1x7 high-resolution sensors; #4: adding 1x12 high-

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016 resolution sensors; #5: adding 3x5 high-resolution sensors. The BT labels indicate the baseline design for toroidal mode number measurements with 2x18 evenly spaced sensors, and the configurations obtained adding to this baseline assembly [1x5, 1x7, 1x12, 3x5] high-resolution sensors, and the baseline design for poloidal mode number measurements with 16 un-evenly spaced sensors, and the configurations obtained adding to this baseline assembly [1x5, 1x7, 1x12] high-resolution sensors. The {ES1→ES4} and {US1→US4} geometries can be used for both toroidal and poloidal mode number analysis, as only one equatorial port is required for the high-resolution sensors. Conversely, the ES5 and US5 geometries can only be used for toroidal analysis, as multiple high-resolution arrays are used.

Figure10a and fig10b show some representative examples of the geometries used for the toroidal and poloidal mode number measurement arrays. For comparison purposes, we also show the spectral window $W(v)$ for the various geometries shown in fig10(a,b), v being the toroidal (n) or poloidal (m) mode number, respectively. The spectral window for all these geometries is well behaved, i.e. it does not show any peaks >0.8 in the mode number range of interest, which would very much complicate the analysis [24, 25]. Hence, having passed the preliminary test of a well-behaved $W(v)$, we have then defined four additional and complementary criteria to assess the cost-normalized measurement performance of any given geometry. For each test, we performed 50'000 simulation runs for each of the selected 19 geometries. We have used an input spectrum $S_{IN}(x_n)$

$$S_{IN}(x_n) = \left[\sum_{k=-f_{MAX}}^{k=+f_{MAX}} A_k \exp(j\nu_k x_n + j\delta_k) + \sigma_{SIG} \times (r_{1k} + jr_{2k}) \right] + \sigma_{MEAS}(x_n) \times (r_{3n} + jr_{4n}). \quad (6)$$

containing between 3 and 7 modes with known randomly chosen and normalized amplitudes in the range $0.05 \leq A_k \leq 1.00$, relative phases $0.00 \leq \delta_k \leq 1.95 * \pi$ and random choice of integer mode numbers up to $|n| \leq 30$ ($|f_{MAX}|=150$) and $|m| \leq 60$ ($|f_{MAX}|=300$) for the toroidal and poloidal mode number analysis, respectively, and scanning the standard deviation in the background noise in the range $0.00 \leq \sigma \leq 0.30$. In eq.(6) $x_n \in [0, 2\pi]$ is the position of the n -th sensor, and the quantities $\sigma_{SIG} \in [0, 1]$ and $\sigma_{MEAS}(x_n) \in [0, 1]$ represent the standard deviation in the background noise on each spectral component and on the measurement itself at each sensor, respectively. These quantities have a fixed and unique value for each simulation as they can in principle be measured directly on the system when installed. The quantities $\{r_{1k}, r_{2k}, r_{3n}, r_{4n}\}$ are random numbers chosen from a uniform distribution in the interval $[0.0 \rightarrow 1.0]$; note that the random seed used for $\{r_{1k}, r_{2k}\}$ can be different from the one used for $\{r_{3n}, r_{4n}\}$. With this approach, the noise has independent and un-correlated complex components satisfying the circularity property. In general, σ_{SIG} and σ_{MEAS} can be different and, more importantly, σ_{MEAS} can have different values for different sensors. Intuitively, σ_{SIG} can be associated to background noise

from the plasma, for instance due to un-coherent turbulence; conversely, σ_{MEAS} is associated with “engineering” errors, such as tolerances on the position and alignment of the sensors, calibration errors, and various effects such as cross-talk, drifts, offset, signal pick-up and bit-noise in the cabling and electronics.

The main results of the baseline analysis and an attempt at optimization of the ITER HF magnetic diagnostic system are shown separately for the four selected tests in the following sub-sections. We will then conclude this section with the analysis of the expected measurement performance of the HF magnetic diagnostic system used for toroidal mode number analysis on the TCV tokamak, which has recently been upgraded with the installation of 3 new 3D sensors made using the Low-Temperature Co-fired Ceramic technology [93].

(5.1) Noise test analysis.

The first test aims at assessing whether one particular geometry is more prone than the others to mistakenly “recognize” white noise as being a high- $n(m)$ mode. Therefore, we consider an input data set made only of white Gaussian noise of known variance, and we determine the 95% and 99% confidence level for not detecting any *true* mode.

Figure 11 shows a summary overview of this analysis. We find that the best performing geometry has ~ 30 un-evenly spaced sensors, but needs around ~ 40 equi-spaced sensors. For an even higher number of sensors the cost function increases much more rapidly than the confidence level for noise rejection, i.e. the measurement performance of the system becomes much less cost-efficient. We also find that the reduction in the cost-normalized confidence level for noise rejection is much sharper for an equi-spaced geometry as the number of sensors increases above its optimum value. We conclude that geometries made with sub-assemblies with spatial periodicities are inherently more prone to incorrect detection of high- $n(m)$ modes than those using un-evenly distributed sensors. For the baseline toroidal and poloidal designs, the best cost-normalized performance is obtained adding one array of 7 high-resolution sensors to the nominal ITER design, for a total of 43 and 23 sensors, respectively. However in both cases the resulting confidence level values =0.907 (toroidal) and =0.924 (poloidal) are still below the best values =0.935 which is obtained with 25 un-evenly spaced + 1x5 high-resolution sensors.

(5.2) False alarms analysis.

For the second test, we consider the statistics of correctly recognizing the given input real modes, to which white Gaussian noise of known variance is added, vs. the occurrence of *false alarms*, i.e. modes being detected which are not in the input dataset (i.e. detected modes have a different mode number

than the input ones) or for which the difference between the input and detected amplitude is larger than the set tolerance for the corresponding mode number. The geometries giving the higher number of correctly detected modes and the lower values of *false alarms* are then the best choices for actual in-vessel installation.

Figure12 shows a summary of this analysis. We find again that not only the fraction of *false alarms* is lower for the geometries using un-evenly spaced sensors, but it also reaches a local minimum for a lower number of sensors. The addition of one high-resolution array of seven sensors clearly improves the measurement performance of the baseline ITER geometries against detection of false alarms for toroidal and poloidal mode number analysis, at the expense, however, of a larger number of sensors for a higher false alarm fraction.

(5.3) Resilience to the loss of sensors.

For the third test, we consider the resilience in the measurement performance of all the test geometries against the loss of sensors through faults, considering the nominal cases of [10%, 20%, 30%] loss. A measure for this resilience is provided by evaluating the confidence level in achieving the same measurement performance of a sensors' arrangement that has all its sensors when some sensors have been lost: the higher the confidence level over all possible permutations of lost sensors and input spectrum variations, the more resilient is that sensors' arrangement against the loss of sensors. To assess the results of this test, we (logically, but somewhat arbitrarily) choose to define the values ≥ 0.85 , ≥ 0.75 and ≥ 0.65 as the thresholds in the confidence level associated to a [10%, 20%, 30%] loss of sensors, respectively, that indicate that a certain geometry still satisfies the measurement performance requirements even when not all sensors are available. Note that the system costs do not enter the evaluation of this particular test, as these costs are defined once and for all when the system is built and do not change if any number of sensors is lost at a later stage during the machine lifetime. Figure13 shows a summary of this analysis. The threshold values for acceptance of this test are explicitly shown by the horizontal (magenta) lines to guide the eye. Considering the example of a nominal 10% loss of sensors, we find that the nominal ITER geometry for toroidal mode number detection does not satisfy the requirements for the resilience in the measurement performance because of its intrinsic spatial periodicity, and only adding at least one array of 12 high-resolution sensors can correct this problem. For toroidal mode number analysis, an assembly with 25 un-evenly spaced sensors in total, comprising 3x5 high-resolution arrays, satisfies the requirements for resilience in the measurement performance even for a 30% loss of sensors. It is not possible to satisfy the requirements for resilience in the measurement performance against loss of sensors with the nominal ITER geometry for poloidal mode number detection, even when adding up to 12 high-resolution sensors.

This is due to the low number of sensors =16 in the nominal design. For poloidal mode number analysis, an assembly with 30 to 35 un-evenly spaced sensors in total, comprising one array of 12 high-resolution sensors, satisfies the requirements for resilience in the measurement performance even against a nominal 30% loss of sensors.

(5.4) Tolerance to installation and calibration errors and uncertainties.

For the fourth test, we consider two elements, namely that (a) the position of each individual sensor is not absolutely fixed, as given in the installation drawings, but that there is a given volume where that sensor will be located, as actually installed in-vessel, and (b) there will be uncertainties in the end-to-end calibration of the transfer function for each sensor, causing errors in the determination of the relative amplitude and phase of the signals from multiple sensors [94, 58]. To practically understand this last term, consider that the relative phase shift $\Delta\Phi$ between the measurements obtained from two sensors is only due to a calibration error. We can then set $\Delta\Phi=v\Delta\zeta$, where v is the mode number and ζ the corresponding angular coordinate: for any mode number, the “wrong” relative phase shift corresponds to a “calculated” sensor separation which is not the actual one.

These two elements effectively add an additional free parameter, i.e. a “tolerance” on the nominal position of each sensor as given by an in-vessel survey. We can then assess whether one particular geometry is more sensitive than the others to the exact position of each sensor by “numerically moving” the initial location of each sensor to achieve the best measurement performance. If a displacement outside this positional tolerance is needed to improve the overall measurement performance, this tells us that we must then change the nominal in-vessel position of (some of) the sensors, which in turns implies that the initial geometry was not optimized.

Defining pos_{ITER} as the initial and pos_{OPT} as the optimized final sensor position, we construct the average shift in the sensor position $SensorShift$ for a total of NN sensors as:

$SensorShift = \frac{1}{NN - 1} \sqrt{\sum_{k=1}^{k=NN} (pos_{OPT} - pos_{ITER})^2} .$	(7)
--	-----

We then (again logically but somewhat arbitrarily) consider that each individual geometry satisfies the ITER measurement requirements if the overall sensors’ displacement is $\langle SensorShift \rangle \leq 2.5deg$ with a standard deviation $std(SensorShift) \leq 1.5deg$ (or larger, provided their sum is $\leq 3.5deg$) when averaged over all arrangements representing that particular geometry.

Figure14 shows a summary of this analysis. For toroidal mode number detection, the baseline ITER geometry satisfies the optimization requirements only when adding at least 1x7 high-resolution sensors. This gives a total of 43 sensors, whereas an assembly with 25 un-evenly spaced sensors in

total, comprising 3x5 high-resolution arrays, already satisfies those requirements. The nominal ITER geometry for poloidal mode number analysis does not satisfy the optimization requirements. Conversely, an assembly with 30 un-evenly spaced sensors in total, comprising at least 7 high-resolution sensors, satisfies them.

(5.5) Overall evaluation of the measurement performance.

Having performed the four tests mentioned above, we can now proceed to an overall evaluation of the measurement performance of the ITER HF magnetic diagnostic system for the test geometries analysed in this work. To this end, we define a threshold value for each of the individual tests, telling us if such a test has been passed. In that case, we give a value =1 to the (cost-normalized) confidence level in achieving the target measurement requirements. If the test has not been passed, the confidence level is then reduced proportionally to the distance from the set threshold value. The results from the four tests are then averaged: this defines the cost-normalized confidence level in achieving the ITER measurement requirements. The threshold values defining the acceptance of a test as function of the toroidal and poloidal mode numbers for different class of HF instabilities are presented in Table2, using the same units and conventions of the corresponding Sections 5.1 to 5.4. For the noise rejection tests presented in Section 5.1, we indicate the fraction of detected modes due to noise. For the false alarms tests presented in Section 5.2, we s the fraction of wrongly detected modes. For the tests on the resilience to the loss of sensors presented in Section 5.3, we indicate the confidence level in achieving the same measurement performance of the complete array when some sensors are lost. For the positional optimization tests presented in Section 5.4, we indicate the maximum allowed sensor shift and its standard deviation.

	Machine Protection	Basic Control	Advanced Control	Physics Studies
Mode Numbers	$n \leq 3, m \leq 5$	$n \leq 5, m \leq 10$	$3 \leq n \leq 10, 5 \leq m \leq 20$	$10 \leq n \leq 20, 20 \leq m \leq 30$
Noise Rejection	0%	0%	10%	30%
False Alarms	0%	0%	10%	30%
Sensor Loss: 10%	95%	95%	85%	70%
Sensor Loss: 20%	90%	90%	75%	60%
Sensor Loss: 30%	85%	85%	65%	50%
Sensor Shift	$(1.5 \pm 0.5) \text{deg}$	$(1.5 \pm 0.5) \text{deg}$	$(2.0 \pm 1.0) \text{deg}$	$(2.5 \pm 1.5) \text{deg}$

Table2: the threshold values used to define acceptance of the tests described in Section 5.1 to 5.4, for the different classes of HF instabilities in the corresponding mode number range.

Figure15 shows the results of this analysis for some of the test geometries analysed in the previous sub-sections. For clarity, the three cases of [10%, 20%, 30%] loss of sensors are considered

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016 separately, and we also separate the analysis for individual (groups of) mode numbers, as in Table2. There are two main reasons for this approach, which are graphically illustrated in fig15.

The ITER measurement requirements for HF instabilities [21-23] specify four main topics, namely machine protection, basic and advanced control, and physics studies, classified according to the mode numbers. Modes required for machine protection are those with toroidal mode number $|n| \leq 3$ and poloidal mode number $|m| \leq 5$, which correspond to basic instabilities such as precursors for sawteeth, Edge Localized Modes and disruptions, Neoclassical Tearing Modes (NTMs) and Resistive Wall Modes: detection of these modes will call for a hard feedback controlled reaction aimed at protecting the integrity of the machine. Modes required for basic control have similar mode numbers in the range $|n| \leq 5$ and $|m| \leq 10$ and correspond to less dangerous variants of the modes sought for machine protection, so that their detection will call for a softer feedback controlled reaction. Modes required for advanced control usually have $3 \leq |n| \leq 10$ and poloidal mode number $5 \leq |m| \leq 20$, corresponding to Alfvén Cascades and AEs. Finally, modes with higher toroidal and poloidal mode numbers up to $|n| \leq 20$ and $|m| \leq 30$ are classified of interest for dedicated physics studies. Even higher mode numbers are not subject to any detailed measurement specification, and do not enter the assessment of the measurement performance. The classification of these groups of instabilities as function of their mode numbers is somewhat arbitrary and partially overlapping, as for instance global, low $m/n=3/2$ AEs are usually much more benign modes than $m/n=3/2$ NTMs, so that the former always falls into the basic control class, whereas the latter may under certain experimental conditions fall into the machine protection class. This is reflected by the (green) vertical lines in fig15, separating the mode numbers into classes: these vertical lines can be “moved” to reflect changing physical understanding and measurement specifications.

Similarly, we somewhat arbitrarily choose to define that an acceptable value for the confidence level in the measurement performance of a given sensors’ arrangement is $=0.85$ when all sensors are still available, and we also set this value to be the same for all classes of HF instabilities, i.e. a flat value independent of the mode number. We then reduce this threshold to again a flat value $=0.70$ and $=0.65$ for toroidal and poloidal mode number analysis, respectively, when considering a nominal 30% loss of sensors. This is reflected by the horizontal (magenta) lines in fig15, which can also be “moved” to reflect changing requirements in the desired confidence level. By combining our wishes for the confidence level for the different classes of HF instabilities, we can then determine whether a particular sensors arrangement satisfies the ITER measurement requirements.

Figure15a shows the summary results of this analysis for some representative geometries usable for toroidal mode number detection. We find that only by adding one array of 12 high resolution sensors can we use the baseline 2009 ITER design with 2×18 equi-spaced sensors, but only when all 48

sensors are present: this is simply due to the intrinsic periodicity of the baseline array. However, already a 10% loss makes this modified baseline geometry unable to satisfy the requirements in the measurement performance, unless we install 3x5 high-resolution arrays in well-separated equatorial ports, for instance ports [#3, #10, #14]. This therefore makes a total of >50 sensors in each array used for toroidal mode number detection. When using un-evenly spaced sensors, we can reduce their total number down to 30 if we install 3x5 high-resolution arrays in the same well-separated equatorial ports. If these ports are too close-by, for instance ports [#8, #9, #10], then we are unable to satisfy the measurement performance requirements with 30% loss of sensors, particularly for higher mode numbers. By installing high-resolution array too close by, we reduce the spatial coverage of the measurements: hence we became very sensitive to which sensors are actually lost through fault. If only one equatorial port is available for HF magnetic measurements, then our best geometry is not as performing and needs at least 35 sensors, including an 1x7 high resolution array, or 40 sensors, if we can only have an 1x5 high resolution array.

Figure15b shows the summary results of this analysis for some representative geometries usable for poloidal mode number detection. We find that not even by adding one array of 12 high resolution sensors can we use the baseline ITER design with 16 un-evenly spaced sensors: this is simply due to the intrinsic low number of sensors in the baseline array. Similar results are obtained when considering two alternative geometries with 32 un-evenly spaced sensors in total, comprising an high-resolution array of 12 and 7 sensors, respectively. Only a geometry with 35 un-evenly spaced sensors in total, comprising one array of 12 high resolution ones, can satisfy the measurement performance requirements for physics studies, i.e. up to $|m| \leq 30$, and this also considering the case of a 30% sensor loss. If we relax this requirement and take into account only modes relevant for advanced control, i.e. up to $|m| \leq 20$, we find two other geometries to be acceptable: the first one still has 35 sensors in total, including one array of 7 high resolution ones, whereas the second requires 45 sensors in total as it only uses one array of 5 high resolution sensors.

In summary, our analysis has demonstrated that the more robust sensor geometry is the un-evenly spaced one, i.e. one without periodicities in the sensors' spacing. For the foreseeable input mode spectra for ITER, a truly un-evenly distributed geometry is the more resilient to the loss of sensors, furthermore being much less sensitive to false alarms caused by background noise in the input spectrum. Conversely, a geometry made up of equi-spaced sub-assemblies has the lowest resilience to the loss of sensors, and the highest sensitivity to false alarms, even if the initial number of sensors is larger than that needed to obtain the required spatial Nyquist number. High-resolution arrays (located in well separated ports for toroidal mode number analysis) are very useful to reduce the total number of sensors required for installation. Finally, our optimization tests indicate that a separation

smaller than 2deg to 3deg between adjacent sensors is not necessarily beneficial, even for high-n(m) detection, as random phase shifts due to background noise mask the “true” phase shifts for the closest sensors, which in turns makes it more difficult to detect high-n(m) modes with a sufficiently high confidence level. Keeping in mind the ITER measurement requirement for the HF magnetic diagnostic system, these results mean that it is indeed possible to find an optimum compromise between the need for redundancy, calling for the use of arrays of the largest possible size, the in-vessel installation constraints, which calls for the least possible number of sensors, and the need for having a solution for the n- and m-number analysis which is unique (i.e. irrespective of the number of sensors), robust (i.e. keeping the same accuracy irrespective of the input spectrum) and resilient against the loss of faulty sensors up to a specified value.

(5.6) Analysis of the measurement performance of the low-field-side toroidal array on TCV.

A new set of HF, 3D magnetic sensors based on combining the Low Temperature Co-fired Ceramic (LTCC) and the classical thick-film technologies have been designed and manufactured in-house for installation on the Tokamak à Configuration Variable (TCV) [93]. Five LTCC-3D sensors were initially foreseen to be installed along the toroidal direction on the low-field side (LFS) of TCV to optimize the measurement performance of this toroidal array while limiting the number of elements for the first phase of installation on TCV. However, due to problem that occurred during installation with one of the in-vessel attachment and support elements, only three sensors could be installed and therefore it became important to understand which of the original five positions should be sacrificed in order to maintain the highest possible measurement performance.

The original LFS toroidal array on TCV has 16 equi-spaced Mirnov-1D sensors (one on each sector of the TCV tokamak) and a 17th not equi-spaced sensor that is installed mid-way between the 15th and 16th equi-spaced sensors. The five LTCC-3D sensors were originally intended to be installed at not equi-spaced position in sectors [#2, #10, #12, #14, #16]. Figure16 shows the spectral window for all the possible selection of three out of five LTCC-3D sensors. We find that the current LFS toroidal array on TCV exceeds the “bad detection limit” [25, 26] of a spectral window $W=0.85$ for $|n|=16$, which represents the absolute measurement limitation of this system, twice the Nyquist limit $|n|=8$. Had it be possible to install the five LTCC-3D sensors, this limit would have been removed, as $W(|n|=16)=0.61$ and $W(|n|=32)\sim 0.79$ also would have remained below the value $W=0.85$, but also $W<0.25$ for all other toroidal components, i.e. below the spectral noise limit [25, 26]. By installing only three sensors, we find that choosing sectors [#2, #14, #16] we have $W(|n|=16)=0.72$ and $W(|n|=32)\sim 0.84$, marginally below the bad detection limit. This configuration is therefore chosen for

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016
installation, as all other selection of three sectors out of five give rise to $W(|n|=32) > 0.87$, marginally above the bad detection limit.

(6) Electrical survey of the proto-Villanovian necropolis in Poggio della Pozza (Italy).

A completely new development on possible applications of an analysis method based on the SRS has come to light very recently following a discussion with some childhood friends of one of the Authors (DT), namely if it was conceivable to develop a cheap and user-friendly method to analyse the data obtained during an electrical prospection made in 1988 of the Proto-Villanovian necropolis of Poggio della Pozza [95]. This site is dated circa 1200 BC and is located around 50km north-east of Rome. One important feature of this site is the use of a *ziro* (one example is shown in figure 17) as the cinerary urn. The *ziro* is a huge carved-out stone egg (approximately 2m tall, 1m wide), set typically 1m to 2m underground in an isolated position. Inside the *ziro*, there is a ceramic urn, a *canopo*, which contains the dead's ashes and a few objects. Sometimes smaller *ziri* are present around a larger one, which is usually considered to indicate the close-by burial of different family members. Although very rewarding when found, it is in fact very time consuming to find a *ziro*, as these objects are usually sparsely located in a field of excavation that can in principle be very large.

The site of Poggio della Pozza (first the necropolis and later also the close-by village) has been extensively excavated by Universities and by an Italian volunteer archaeological association from the early 1980's to the mid-1990's, and one of the Authors (DT) of this paper has been in charge of the excavations conducted during the summer of 1988. During this period, an electrical prospection was performed to try to locate the presence of the *ziri* by measuring the anomalies that such a bulky stone object would produce in the electrical potential. This technique is sometimes used in archaeological excavations, and possibly one of the most successful examples is the 3D electrical tomography survey of an insula in Regio III in Pompei [96]. The idea behind this technique is very simple [97], as shown in fig 18: the ground resistance is affected by the presence of a sufficiently large object underground, the AC current flow driven with an array of current electrodes is then distorted around this object, and therefore the potential contours measured with an array of potential electrodes will also be distorted and will not show anymore the usual 2D (planar at ground level) circular pattern.

While the measurement technique is in principle simple and sufficiently inexpensive, the analysis of these electrical prospection data is actually very complex and requires dedicated and very expensive purpose-built software that is not available to a volunteer group in Italy. An attempt has therefore been made to extend the model leading to the SRS to the analysis of such archaeological data, by assuming that equi-potential circular *equilibrium* contours are affected by the presence of a small number of spatially separated *sources* (namely: buried objects which affect the underground electrical

potential through changes in the local resistivity), which create localised *instabilities* in the equilibrium signal. The analogy with the SRS method is as follows: in the absence of instabilities, the measurements of electrical potential will show a (circular) periodicity related to the spacing of the electrodes, which corresponds to the spectral window of the system. The spectral window is changed by the presence of underground sources: by moving the electrodes, a map of the localisation of the sources could in principle be reconstructed. Then, following the same approach used in Section 5.4 for the analysis of the ITER HF magnetic system, the position of these underground sources, assuming a standard size and depth for the ziro, is varied to match as best as possible the ground measurement of electrical potential, assuming an otherwise uniform ground resistance.

Figure19 show the results of this attempt for the analysis of the electrical prospection data of the proto-Villanovian necropolis in Poggio della Pozza: the position of the majority of the larger ziri that were already found up to 1988 is actually sufficiently well reproduced, typically within a 3m radius. Only in a couple of cases there is no direct correspondence within a radius of 10m between the position of a ziro as determined from the analysis of the electrical survey performed in 1988 and as actually found during the excavations. In a few more cases there is such a correspondence, but only when selecting a radius between 3m and 10m, and this usually occurs close to the sparse trees that were growing on the site at the time. These trees may have large underground roots that locally affect the ground water retention, hence likely invalidating the assumption of a ground resistance that is otherwise uniform but for the presence of a ziro. Based on the results of this analysis, we are hopeful that during forthcoming excavation campaigns, possibly scheduled for the summers of 2016 or 2017, we should be able to locate more rapidly the other larger ziri still buried underground in the non-excavated area of this site.

(7) Summary, conclusions and an outlook to future work.

We have used recent techniques of signal processing in astronomy and astrophysics, based on the *Sparse Representations of Signals*, to solve current questions arising in thermonuclear fusion plasmas. Two examples are the detection of magneto-hydrodynamic instabilities, which is now performed routinely in JET in real-time on a sub-millisecond time-scale, and the studies leading to the optimization of the magnetic diagnostic system in ITER and TCV. These questions have been solved formulating them as inverse problems, despite the fact that these applicative frameworks are extremely different from the classical use of *Sparse Representations*, on both the theoretical and computational points of view. Very recently, an attempts has also been made to apply this method to the analysis of the data obtained during an electrical prospection of an archaeological site in central Italy.

The advantages of using a *Sparse Representation* method, and particularly algorithms based on the *SparSpec* code, are the high speed at which automated calculations can be performed, coupled with the relative ease with which the results can be understood. As an example, using Matlab R14 on a 2GHz laptop with 1024MB of RAM, the complete post-pulse analysis of the mode number and damping rate data collected during one full JET discharge with the AEAD system takes about 200sec of CPU time when performed using the *SparSpec* algorithm, compared to in excess of 2'000sec of CPU time when performed using an SVD algorithm similar to those presented in [6, 7]. Despite the many hardware limitations of the current AELM system on JET, only with *SparSpec* a similar analysis can be performed in real time on a sub-millisecond time base, as other available algorithms based on SVD methods would clearly exceed the CPU and RAM limits. Whereas it is true that for ITER the real-time hardware resources will be much improved, it will also be clearly beneficial to use them judiciously, i.e. as efficiently and as little as possible.

Similarly, the optimization analysis for the ITER HF magnetic diagnostic system takes advantage of one of the main features of *Sparse Representation* methods, namely the relative ease with which physics-based tests can be turned without further supervision into precise mathematical properties and specifications for any diagnostic system. Combined with the numerical efficiency of *SparSpec*, and using the methods described in this work, we find that we can fully analyse the cost-normalized measurement performance of one arrangement of magnetic sensors typically within 12 hours of CPU time using Matlab R14 on a 2GHz laptop with 1024MB of RAM.

For forthcoming large engineering and scientific projects, such as ITER and the SKA, it is clear that efficient and automated data analysis, in “real-time” wherever possible, will be of paramount importance. For these projects, “real-time” means that the data analysis will have to be performed over a time scale much faster than those over which the experimental measurements will change, so that a control reaction of some actuators may need to be called upon. Whereas for ITER and perhaps one gigabyte of data for one particular sub-system, real-time means calculations performed on a sub-millisecond time scale, for the SKA collecting hundreds of terabytes of data for one single image the relevant time scale is of the order of a week, i.e. the time it may take to re-deploy some of the optics for some of the telescopes to obtain a more accurate image of the same view of the universe. Therefore, the experience with JET data has clearly indicated that, due to their speed, accuracy and automated operation, *Sparse Representation* methods are perfectly adapted to achieve the goal of obtaining real-time measurements with an accuracy satisfying the desired requirements in future complex engineering and scientific devices.

Acknowledgements.

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission. This work was also partly supported by the Swiss National Science Foundation. The Authors would like to thank the various members, past and present, of the CRPP, MIT and JET staff that have contributed to the design, installation, commissioning and operation of the Alfvén Eigenmode Active Diagnostic system over more than 10 years of experiments at JET. The Authors would also like to acknowledge the useful discussions we have had during the execution of some of this work with Dr. George Vayakis of the ITER organization. The Authors would also like to thank Dr. Ming Yang of the National Observatory of Athens for his assistance and permission to reproduce some of the contents of his work on the Harvard Variable HV2576, Dr. Micael Bazot of the Center for Space Science of New York University in Abu Dhabi for his assistance and permission to reproduce some of the contents of his work on the p-mode frequencies of the solar twin 18 Scorpii, **and Dr. Neil Linford of English Heritage for his assistance and permission to reproduce some of the contents of his work on the use of electrical prospection techniques on archaeological sites.**

Finally, DT would also like to personally acknowledge the contributions of various members and childhood friends of the Gruppi Archeologici d'Italia for their work in the proto-Villanovian site of Poggio della Pozza, and for trying to obtain permission for a new campaign of excavation that could validate the analysis of the 1988 electrical survey data reported here.

References.

1. F.Marvasti, *Non-Uniform Sampling: Theory and Practice*, 2001, Springer-Verlag.
2. N.R.Lomb, *Astrophysics and Space Science* **39(2)** (1976), 447.
3. J.D.Scargle, *The Astroph. Journal* **263** (1982), 835.
4. D.Roberts, J.Lehar, J.Dreher, *The Astron. Journal* **93** (1987), 968.
5. G.Foster, *The Astron. Journal* **109** (1995), 1889.
6. J.S.Kim et al., *Plasma Phys. Control. Fusion* **41** (1999), 1399.
7. M.Hole, L.Appel, *Plasma Phys. Control. Fusion* **49** (2007), 1971.
8. F.M.Poli et al., *Plasma Phys. Control. Fusion* **50** (2008), 095009.
9. E.P.Wigner, *Phys. Rev.* **40** (1932), 749.
10. H.Choi, W.Williams, *IEEE Trans. Acoust. Speech Signal Process* **37** (1989), 86.
11. N.E.Huang et al., *Proc. Royal Society London* **A454** (1998), 903.
12. S.Zegenhagen, A.Werner et al., *Plasma Phys. Control. Fusion* **48** (2006), 1333.
13. Various Authors, *Progress in the ITER Physics Basis 2007*, Nucl. Fusion **47**, 2007.
14. E.J.Strait, E.D.Fredrickson, J-M.Moret, M.Takechi, *Fus. Sci. Tech.* **53** (2008), 304.
15. A.Fasoli et al., *Phys. Rev. Lett.* **75** (1995), 645.
16. S.Bourguignon, H.Carfantan, T.Böhm, *Astron. Astroph.* **462** (2007), 379.
17. S.Bourguignon, H.Carfantan, J.Idier, *IEEE Journal of Selected Topics in Sig. Proc.* **1** (2007), 4.
18. A.Klein, H.Carfantan, D.Testa et al., *Plasma Phys. Control. Fusion* **50** (2008), 125005.
19. D.Testa et al., *Fus. Eng. Des.* **86** (2011), 381.
20. D.Testa et al., *Europhysics Letters* **92** (2010), 50001.
21. For a general overview of the ITER measurement requirements for the HF magnetic diagnostic system, see Table-2 in the Design Description Document DDD 5.5.A, ITER document reference “N55DDD101-06-12W0.3”.
22. G.Vayakis et al., *Rev. Sci. Instrum.* **74** (2003), 2409.
23. G.Vayakis et al., *Rev. Sci. Instrum.* **83** (2012), 10D712.
24. A.J.H.Donné et al., *Nucl. Fusion* **47** (2007), S337.
25. D.Testa et al., *Fusion Science Technology* **57(3)** (2010), 208; <http://epubs.ans.org/?a=9468>.
26. D.Testa et al., *Fusion Science Technology* **57(3)** (2010), 208; <http://epubs.ans.org/?a=9469>.
27. D.Testa, et al., *Plasma and Fusion Research* **9** (2014), 1405003.
28. M.Yang and B.W.Jiang, Proceedings International Astronomical Union Symposium No. 252, 6-11 April 2008, Sanya, Hainan Island, China; doi: 10.1017/S1743921308022990.
29. A.Lévy et al., *Space Sci Rev* **151** (2010), 105; doi: 10.1007/s11214-009-9596-4.
30. M.Bazot et al., *Astron. Astroph.* **44** (2012), A106; doi: [10.1051/0004-6361/201117963](https://doi.org/10.1051/0004-6361/201117963).
31. R.Poggiani, *Proc.Science* (SWIFT 10) 157, 2014.

32. D.Testa et al., *Nucl. Fusion* **50** (2010), 084010; doi: [10.1088/0029-5515/50/8/084010](https://doi.org/10.1088/0029-5515/50/8/084010).
33. D.Testa et al., *Nucl. Fusion* **51** (2011), 043009; doi: [10.1088/0029-5515/51/4/043009](https://doi.org/10.1088/0029-5515/51/4/043009).
34. T.Panis et al., *Nucl.Fusion* **52** (2012), 023013; doi: [10.1088/0029-5515/52/2/023013](https://doi.org/10.1088/0029-5515/52/2/023013).
35. T.Panis et al., *Nucl. Fusion* **52** (2012), 023014; doi: [10.1088/0029-5515/52/2/023014](https://doi.org/10.1088/0029-5515/52/2/023014).
36. D.Testa et al., *Nucl. Fusion* **52** (2012), 094006; doi: [10.1088/0029-5515/52/9/094006](https://doi.org/10.1088/0029-5515/52/9/094006).
37. D.Testa et al., *Nucl. Fusion* **55** (2015), 123010; doi: [10.1088/0029-5515/55/12/123010](https://doi.org/10.1088/0029-5515/55/12/123010).
38. D.Testa, M.Albergante, *Europhysics Lett.* **97** (2012), 35003; doi: [10.1209/0295-5075/97/35003](https://doi.org/10.1209/0295-5075/97/35003).
39. D.Testa, M.Albergante, *Nucl. Fusion* **52** (2012), 083010; doi: [10.1088/0029-5515/52/8/083010](https://doi.org/10.1088/0029-5515/52/8/083010).
40. C.L.Carilli, S.Rawlings, *New Astronomy Review* **48** (2004), 979; <http://dx.doi.org/10.1016/j.newar.2004.09.001>; see SKA website <http://www.skatelescope.org/>.
41. C.E.Shannon, *Communication in the presence of noise*, Proc. Institute of Radio Engineers **37** (1949), 10; reprinted in: *Proc. IEEE* **86** (1998), 447.
42. J.Lafler, T.D.Kinman, *The Astrophys Journal Suppl.* **11** (1965), 216.
43. D.F.Gray, K.Desikachary, *The Astrophys Journal* **181** (1973), 523.
44. M.M.Dworetsky, *Mon. Not. R. Astron. Soc.* **203** (1983), 917.
45. P.Stoica, R.L.Moses, B.Friedlander, T.Söderström, *IEEE Trans. ASSP* **37** (1989), 378.
46. M.D.Sacchi, T.J.Ulrych, C.Walker, *IEEE Trans. Signal Proc.* **46** (1998), 31.
47. D.L.Donoho, *Jour. Comm. on Pure & Applied Math.* **59** (2006), 797.
48. D.M.Malioutov, *A sparse signal reconstruction perspective for source localization with sensor arrays*, Master Thesis (2003), MIT.
49. S.S.Chen, D.L.Donoho, M.ASaunders, *SIAM Journal on Scient. Comp.* **20** (1998), 33.
50. J-J.Fuchs, *IEEE Trans. Inf. Theory* **50** (2004), 1341.
51. J.A.Tropp, *IEEE Trans. Inf. Theory* **52** (2006), 1030.
52. J.A.Tropp, *IEEE Trans. Inf. Theory* **50** (2004), 2231.
53. P.Moulin, J.Liu, *IEEE Trans. Inf. Theory* **45** (1999), 909.
54. J-J.Fuchs, *IEEE Trans. Inf. Theory* **51** (2005), 3601.
55. V.Merezhkin, *Soviet JETP* **4** (1978), 152.
56. O.Kluber et al., *Nucl. Fusion* **31** (1991), 907.
57. D.Testa, A.Fasoli, E.Solano, *Review of Scient. Instrum.* **74**, (2003), 1694.
58. T.Böhm et al., *Astron. Astroph.* **427** (2004), 907; doi: [10.1051/0004-6361:20041227](https://doi.org/10.1051/0004-6361:20041227).
59. A.Fumel, T.Böhm, *Astron. Astroph.* **540** (2012), A108; doi: [10.1051/0004-6361/201117817](https://doi.org/10.1051/0004-6361/201117817).
60. M.Sperl, *Comm. in Asteroseismology* **111** (1998), 1.
61. P.R.Wood, *Memorie Società Astronomica Italiana* **77**, (2006), 76.
62. P.Lenz and M.Breger, *Comm. in Asteroseismology* **146** (2005), 53.
63. R.F.Stellingwerf, *The Astrophys Journal* **224** (1978), 953.

64. J.Anderson et al., *Phys. Rev. D* **65** (2002), 082004.
65. S.G.Turyhsev et al., *Phys. Rev. Lett.* **108** (2012), 241101.
66. G.Cayrel de Strobel, *Astron. Astroph. Review* **7** (1996), 243; doi: [10.1007/s001590050006](https://doi.org/10.1007/s001590050006).
67. For details on the SWIFT mission and its burst alert monitor see: <http://swift.gsfc.nasa.gov/>.
68. For details on the Rossi X-rays Timing Explorer see: <http://xte.mit.edu/>.
69. C.Z.Cheng, L.Chen, M.S.Chance, *Ann. Phys.* **161** (1985), 21.
70. R.Betti, J.P.Freidberg, *Phys. Fluids B* **3** (1991), 1865.
71. M.S.Chu et al., *Phys. Fluids B* **4** (1992), 3713.
72. A.D.Turnbull et al., *Phys. Fluids B* **5** (1993), 2546.
73. A.Fasoli et al., *Phys. Plasmas* **7** (2000), 1816.
74. S.Sharapov et al., *Phys. Letters A* **28** (2001), 127.
75. A.Fasoli et al., *Plasma Phys. Control. Fusion* **44** (2002), 159.
76. W.W.Heidbrink, *Phys. Plasmas* **9** (2002), 2113.
77. G.Vlad, S.Briguglio, G.Fogaccia, F.Zonca, *Nucl. Fusion* **46** (2006), 1.
78. Q.A.King, H.Brelen, *An Experimental Control Facility at JET*, JET-P(98)24 (1998), Joint European Torus, Abingdon (UK) (weblink: <http://www.iop.org/Jet/fulltext/JETP98024.pdf>).
79. P.Woskov et al., *New Digital Control System for the JET Alfvén Eigenmode Active Spectroscopy Diagnostic*, GP8.00129, Proceedings 54th Annual Meeting of the APS Division of Plasma Physics, October 29th –November 2nd 2012, Providence (USA).
80. T.Debelle et al., *Toroidal Alfvén Eigenmode Amplifier Control at JET Using Commercial FPGA and PXI Platform to Study Plasma Instabilities*, 11th International Symposium on Fusion Nuclear Technology (ISFNT), 16-20 September 2013, Barcelona (Spain).
81. A.Fasoli, P.Blanchard, D.Testa, *New capabilities of the JET Alfvén Eigenmode antenna system*, ITPA Topical Group Meeting on Energetic Particles, Culham (UK), 22-25 April 2013.
82. L.Lao et al., *Nuclear Fusion* **25** (1985), 1611.
83. M.Keilhacker for the JET team, *Nucl. Fusion* **39** (1999), 209.
84. P.R.Thomas et al., *Phys. Rev. Lett.* **80** (1998), 5548.
85. P.R.Thomas for the JET team, *Nucl. Fusion* **39** (1999), 1619.
86. F.Jenko et al., *Phys. Plasmas* **7** (2000), 1904.
87. M.Schittenhelm et al., *Nucl. Fusion* **37**(9) (1997), 1255.
88. E.Lazzaro et al., *Nucl. Fusion* **55** (2015), 093031; doi: [10.1088/0029-5515/55/9/093031](https://doi.org/10.1088/0029-5515/55/9/093031).
89. S.Nowak et al., *NTM Onset by Central EC Power Deposition in FTU and TCV Tokamaks*, Paper EX/P2-54, 25th IAEA Fusion Energy Conference, San Petersburg, Russia, 13-18 October 2014.
90. E.Lazzaro et al., *Triggerless onset and “natural” rotation on NTMs in TCV tokamak*, paper EX/P4-32, 24th IAEA Fusion Energy Conference, San Diego, USA, 08-13 October 2012.

91. D.Testa et al., *Fus. Eng. Des.* **96-97** (2015), 989-992; [doi:10.1016/j.fusengdes.2015.05.065](https://doi.org/10.1016/j.fusengdes.2015.05.065).
92. R.F.Heeter et al., *Review Scient. Instrum.* **71** (2000), 4092.
93. V.d'Ercole et al., *Ripresa degli scavi nella necropoli proto-Villanoviana di Poggio della Pozza in Allumiere*, Atti del Secondo Incontro di Studi, Centro Studi di Preistoria e Archeologia PPE II(1) (Milano, 1995), 177.
94. R.Alaia et al., *Journal Geophys. Eng.* **5** (2008), 67; doi: 10.1088/1742-2132/5/1/007.
95. N.Linford, *Rep. Prog. Phys.* **69** (2006), 2205; stacks.iop.org/RoPP/69/2205.
96. D.Testa, A.Fasoli, *Nucl. Fusion* **41** (2001), 809; iopscience.iop.org/article/10.1088/0029-5515/41/7/101.
97. M.A.VanZeeland et al., *Nucl. Fusion* **49** (2009) 065003; doi:10.1088/0029-5515/49/6/065003.

List of main symbols used in equations.

<i>Symbol</i>	<i>Explanation</i>
$y(t_p)$	measurements (y) made at the time points t_p
T_s	sampling time
W	spectral window
ε_p	the error on the measurement
c_l	(complex) mode amplitude
ν_l	mode frequency
$\{f_k\}$	discretized mode frequency
$\{x_k\}$	mode amplitude for the discretized mode frequency $\{f_k\}$
λ	penalization parameter related to the noise level.
$\{\phi, n\}$	toroidal angle and (conjugated) toroidal mode number
$\{\theta, m\}$	poloidal angle and (conjugated) poloidal mode number
F_{AE}	Frequency of Alfvén Eigenmodes
B_{TOR}	toroidal magnetic field
R_{RES}	mode resonant position
q_{RES}	safety factor at the mode resonant position
$S_{IN}(\nu_n)$	input signal S_{IN} at the position (ν_n) of each sensor
A_k, δ_k	randomized amplitude (A_k) and relative phase (δ_k)
σ_{SIG}	standard deviation in the background noise on each spectral component
$\sigma_{MEAS}(\nu_n)$	standard deviation on the measurement itself at each sensor

Appendix-A: Sparse Approximation for line spectra estimation from irregularly sampled data in astronomy, astrophysics and thermonuclear fusion.

The A&A problem consists in the analysis of time-series, which can be, for instance, light curves or radial velocity measurements. These measurements are subject to observational constraints, such as day/night alternation and meteorological conditions. The A&A ground-based measurements are therefore always obtained through irregular sampling. This kind of sampling makes the problem of oscillation frequencies detection and estimation particularly difficult. In A&A data analysis (as for analysis of magnetic measurements in thermonuclear fusion plasmas), the main objective is that of looking for periodicities. For the case of variable stars, and multiple star systems, there are several oscillation modes, some of which are related to the stars' orbits and have to be filtered out when oscillations in other quantities are sought. This leads to the estimation of spectral lines from the data. The irregular sampling can be seen as the application of an irregular Dirac comb $w(t)$ to the original signal $y(t)$ and can be well understood in the Fourier domain. Ideal sampling of the signal $y(t)$ at time points t_p , $p=\{1, \dots, P\}$ through the Dirac comb $w(t_p)=\delta(t-t_p)$ leads to the convolution of the spectrum being sought $Y(f)$ by the *spectral window* $W_s(f)$ to produce the signal $Y_s(f)$:

$$y_s(t) = y(t) \times \sum_{p=1}^P \delta(t - t_p) \xrightarrow{FT} Y_s(f) = Y(f) * \mathcal{W}_s(f)$$

$$\text{with } \mathcal{W}_s(f) = \sum_{p=1}^P e^{j2\pi f t_p}.$$

Here $\delta(t)$ stands for the Dirac delta function. For regular sampling $t_p=(p-1)T_s$, where T_s is the sampling time, and the spectral window $W_s(f)$ reduces to a periodized trigonometric (sin) function

$$\mathcal{W}_s(f) = e^{j\pi(P-1)fT_s} \frac{\sin(\pi f P T_s)}{\sin(\pi f T_s)}$$

with a main lobe of width $2/(pT_s)$. With irregular sampling, the window does not have simple properties, but it should be noted however that the width of the main lobe remains of the order of the inverse of the duration of the observation. Moreover, it can be shown that a periodic lack of data generates secondary lobes in $|W_s(f)|$ at the corresponding frequency, whose amplitude increases with the proportion of gaps in the temporal coverage.

Note that the periodicity of $W_s(f)$, which is at the origin of spectral aliasing in the regular sampling case (the well-known Nyquist-Shannon theorem [42]), is no longer valid with irregular sampling. The irregularity of sampling is in this sense an advantage, allowing the estimation of the spectrum on a much wider frequency range [A1]. In return, the definition of the largest frequency for the spectrum is a complex issue. A practical solution consists in choosing a value in accordance with the physics

of data, ensuring that the spectral window does not present any pseudo-periodicity in the band considered for the analysis.

If the signal is considered to be a sum of M frequencies ν_m with amplitudes a_m such that $y(t) = \sum_{m=1}^M a_m e^{j2\pi\nu_m t}$, it has a line spectrum $Y(f) = \sum_{m=1}^M a_m \delta(f - \nu_m)$, and the spectrum of the irregularly sampled signal simply writes as $Y_s(f) = \sum_{m=1}^M a_m \mathcal{W}_s(f - \nu_m)$, which is a weighted sum of shifted spectral windows. The estimation problem is therefore all the more complex when the spectral window has a complex shape.

FigureA1 shows such a particularly difficult artificial data set similar to the one proposed by Foster ([5], data set A), which consists of three sinusoids with periods =370 days, =230 days and =100 days and amplitudes =3.0, =2.828 and =3.0, respectively. A constant value of 10 is added. An initial data set was generated with 200 points sampled every 10 days. The final data sets shows gaps of 100 days every 365 days and gaps of 10 days every 30 days. To make the problem a bit trickier, a fourth sinusoid was added with period =122.5 days and amplitude =3.0, such that the side-lobes caused by the annual gaps (for the periods =370 days and =122.5 days) superimpose at a period =184 days ($1/122.5 - 1/365 = 1/370 + 1/365 = 1/184$), generating a high false peak in the Fourier spectrum. White Gaussian noise with standard deviation $\sigma=0.30$ was also added. Due to the initial regularity of the sampling (sampling frequency $f_s=0.1/\text{day}$) the spectral windows is periodic with a period = f_s . High secondary lobes can then also be seen at frequencies which are multiple of $1/365=2.7 \times 10^{-3}/\text{day}$, corresponding to the annual gaps period. Therefore, the Fourier spectrum of the data presents several local maxima at frequencies that do not correspond to oscillations.

In general terms, for a set of M frequencies $\mathbf{v}=[\nu_1, \dots, \nu_M]$, the data can be modelled as:

$$y(t_p) = \sum_{m=1}^M a_m e^{j2\pi\nu_m t_p} + \varepsilon_p \Leftrightarrow \underline{\mathbf{y}} = \Gamma(\underline{\mathbf{v}})\underline{\mathbf{a}} + \underline{\boldsymbol{\varepsilon}}, \quad (\text{A1})$$

where $\underline{\mathbf{y}}=[y_1(t_1), \dots, y_P(t_P)]$ contains the data sampled at the time points t_p , $\underline{\mathbf{a}}=[a_1, \dots, a_M]^T$ the amplitudes, $\Gamma(\mathbf{v})$ is a $P \times M$ matrix with elements $\Gamma_{mp}=\exp(j2\pi\nu_m t_p)$ and $\boldsymbol{\varepsilon}=[\varepsilon_1, \dots, \varepsilon_P]$ is a random perturbation term. For the sake of simplicity, we consider independent and identically distributed (IID) perturbations, Gaussian, centred, with a variance σ^2 . Note however that all the developments presented here, which are based on the likelihood of the data, can be extended without any major difficulty to the case of non-IID Gaussian perturbations, a case frequently encountered in A&A, for example due to the variation of the atmospheric conditions or to the observation of the same object with several instruments.

When the frequencies are known, the maximum likelihood (ML) estimation of their amplitudes \mathbf{a} corresponds to minimize the LS criterion $\min_{\underline{\mathbf{a}}} \|\underline{\mathbf{y}} - \Gamma(\underline{\mathbf{v}})\underline{\mathbf{a}}\|^2$, which leads to the solution $\hat{\underline{\mathbf{a}}} =$

$(\Gamma(\underline{\mathbf{v}})^H \Gamma(\underline{\mathbf{v}}))^{-1} \Gamma(\underline{\mathbf{v}})^H \underline{\mathbf{y}}$, where the subscript H stands for the Hermitian transpose operator. When the number K of frequencies is known but not the values of these frequencies, the ML estimation leads to minimize the criterion $J_{ML}(\underline{\mathbf{v}}, \underline{\mathbf{a}}) = \|\underline{\mathbf{y}} - \Gamma(\underline{\mathbf{v}})\underline{\mathbf{a}}\|^2$, which is a very difficult optimization problem in general due to the large number of local minima of this criterion. Note that, thanks to the previous expression of $\hat{\underline{\mathbf{a}}}$ this amounts to minimize with respect to the frequencies \mathbf{v} the criterion:

$$J_{ML}(\underline{\mathbf{v}}) = \left\| \underline{\mathbf{y}} - \Gamma(\underline{\mathbf{v}}) \left(\Gamma(\underline{\mathbf{v}})^H \Gamma(\underline{\mathbf{v}}) \right)^{-1} \Gamma(\underline{\mathbf{v}})^H \underline{\mathbf{y}} \right\|^2.$$

Particular cases of this problem are the single frequency estimation, which leads to estimate the frequencies as those which maximize the periodogram:

$$\mathbf{P}(\nu) = \frac{1}{P} \left| \sum_{p=1}^P y(t_p) e^{j2\pi\nu t_p} \right|^2 = \frac{1}{P} |Y_s(\nu)|^2.$$

In the case of a real valued single frequency, writing the matrix $\Gamma(\mathbf{v})$ in terms of sine and cosine functions instead of exponentials, this leads to maximize the LSP [2]. Introducing a time delay $\tau(\nu) = \frac{1}{4\pi\nu} \text{Arctan}(\sum_p \sin(4\pi\nu t_p) / \sum_p \cos(4\pi\nu t_p))$ to invert a diagonal matrix equivalent to $\Gamma(\underline{\mathbf{v}})^H \Gamma(\underline{\mathbf{v}})$, this can be written as:

$$\mathbf{P}_L(\nu) = \frac{(\sum_{p=1}^P y(t_p) \cos(2\pi\nu(t_p - \tau(\nu))))^2}{\sum_{p=1}^P \cos^2(2\pi\nu(t_p - \tau(\nu)))} + \frac{(\sum_{p=1}^P y(t_p) \sin(2\pi\nu(t_p - \tau(\nu))))^2}{\sum_{p=1}^P \sin^2(2\pi\nu(t_p - \tau(\nu)))}$$

Note that in the case of more than one frequencies, if the convolution of the original line spectrum with the spectral window is an interesting interpretation of the classical periodogram, it is not valid for the LSP. Moreover, as illustrated figA1, the maxima of the periodogram do not correspond to the true location of the frequencies. Therefore in the multi-frequencies case, the periodogram cannot be used directly to estimate the frequencies.

A lot of work has been performed in A&A to tackle this difficult multi-frequencies case and most of it can be interpreted in the framework of Sparse Approximations. The difficulty to minimize the least squares criterion $J_{ML}(\mathbf{v})$, even when the number of frequencies is known, is mainly due to the non-linearity of the model with respect to the frequencies ν_m , which leads to a large amount of local minima in the criterion. The problem is generally much more difficult in general, when the number of frequencies is unknown.

Instead of searching for a small number M of frequencies ν_m , one can consider a multi-frequencies model, with a large number K of frequencies regularly discretized between 0 and a maximum frequency f_{MAX} . The model is then written as:

$y(t_p) = \frac{1}{\sqrt{P}} \sum_{k=-K}^K x_k e^{j2\pi f_k t_p} + \varepsilon_p \Leftrightarrow \underline{\mathbf{y}} = W \underline{\mathbf{x}} + \underline{\boldsymbol{\varepsilon}}$	(A2)
--	------

Here the frequencies ν of the multi-frequencies model given in eq.(A1) are searched in the set $f_k=(k/K)f_{\text{MAX}}$, with $k \in \{-K, \dots, K\}$, with corresponding amplitudes x_k and W is a $P \times (2K+1)$ matrix with elements $W_{pk}=(1/\sqrt{P})\exp(i2\pi f_k t_p)$.

For large enough K , both models are equivalent if one adds the constraints that $x_k=0$ for most of the indices k , which correspond to a *sparsity* prior. In contrast to the previous model, for which the unknown parameters are the order M , the M frequencies and the M corresponding amplitudes, the new model, considering fixed frequencies and a known order K , becomes linear with respect to the only unknowns x_k . The introduction of the factor $1/\sqrt{P}$ in the matrix W helps to have its columns \mathbf{w}_k of unit norm, a property which will be useful later.

Abundant works concerning *denoising* and *compression* during the last two decades have relied on the *sparsity* hypothesis which assumes that, for certain classes of signals and images, the essential information is concentrated in a small number of coefficients, expressed in an appropriate transformed space [A2]. Under a linear model such as that of eq.(A2), where W represents this transformation, often redundant (\mathbf{x} typically being of larger dimension than \mathbf{y}), then the sparsest approximation of the data is sought, that is to say, with the least non zero components, that correctly fits the data:

$\min_{\underline{\mathbf{x}}} \ \underline{\mathbf{x}}\ _0 \text{ such that } \ \underline{\mathbf{y}} - W\underline{\mathbf{x}}\ ^2 < \tau,$ $\text{where } \ \underline{\mathbf{x}}\ _0 = \text{Card}\{k \mid x_k \neq 0\}.$	(A3)
---	------

This large dimension combinatorial optimization problem can generally not be solved exactly and two classes of sub-optimal approaches are often considered:

- *greedy methods*, building a sparse solution by selecting components in an iterative manner [53];
- *relaxation methods* of the problem towards a numerically computable approximation: the most commonly used substitute the pseudo-norm L_0 -norm by the L_1 -norm: $\|\underline{\mathbf{x}}\|_1 = \sum_k |x_k|$ [52].

The characterization of the solutions obtained by these approaches in relation to the one of eq.(A3) has been the subject of numerous works and sufficient conditions of equivalence of the solutions have been proposed, based on near-orthogonality measures of W . For example, the solutions obtained by

greedy methods [53] and by L₁-norm relaxation [51] methods correspond to the solution of eq.(A3) in the noiseless case if it satisfies $\|\underline{x}\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mu(W)}\right)$ where the *mutual coherence* is defined by $\mu(W) = \max_{k \neq l} |\underline{w}_k^H \underline{w}_l|$. Other sufficient conditions have been established using *restricted isometry constants* [A3] or the *Exact Recovery Coefficient* [52, 53].

For line spectra estimation, the structure of the dictionary is imposed by the multi-frequencies model and the considered configuration: sampling time points t_p and frequency grid f_k . It can be shown that the above mentioned sufficient conditions are generally not satisfied. In particular, one has that $|\underline{w}_k^H \underline{w}_l| = \frac{1}{p} \left| \mathcal{W}_s \left(\frac{k-l}{K} f_{\max} \right) \right|$, thus the *mutual coherence* is equal to the maximum value (except at 0) of the spectral window. Thus, for the problem of figA1, where the highest secondary lobe is higher than 1/3, the greedy and relaxed approaches are guaranteed to find the correct frequencies only if it has a single frequency ($\|\underline{x}\|_0 = 1$), which is of low interest in practice.

Greedy algorithms, such as those presented in [A4], construct a sparse solution in an iterative manner. A number of variants have been proposed, whose performances can be evaluated from a perspective of compromise between the quality of the solution and their computational complexity. If the interpretation of these methods in the context of sparse representations is fairly recent, they were used much earlier for astronomical data analysis, as early as the 1970's.

Greedy methods construct a solution, for which at the n^{th} iteration the signal \mathbf{y} is approximated with the sum of n columns (called *atoms*) of the matrix W (called *dictionary*), indexed by Ω_n :

$$\hat{\mathbf{y}}_n = \sum_{k \in \Omega_n} x_k \underline{w}_k = W_{\Omega_n} \underline{\mathbf{x}}_n \quad \text{with } W_{\Omega_n} = [\underline{w}_k]_{k \in \Omega_n} \quad \text{and } \underline{\mathbf{x}}_n = [x_k]_{k \in \Omega_n}$$

Based on a zero initialization such that $\hat{\mathbf{y}}_0 = 0$, each iteration thus consists of:

- the selection of an atom \underline{w}_k from the approximation of the previous iteration, defining the new support $\Omega_n = \Omega_{n-1} \cup \{k\}$;
- the amplitudes estimation $\underline{\mathbf{x}}_n$ of the selected atoms;
- the update of the approximation $\hat{\mathbf{y}}_n$ and of the residual error $\underline{\mathbf{r}}_n = \mathbf{y} - \hat{\mathbf{y}}_n$.

The algorithm is stopped when none of the atoms is any longer significant in the residual $\underline{\mathbf{r}}_n$. The stopping rule therefore controls the number of components of the solution, thus is difficult to tune in practice. However, when the noise distribution is known, as for the previous simulated data where the noise is centered, Gaussian with known variance σ^2 (then $\|\underline{\mathbf{e}}\|^2$ follows a χ^2 distribution with N degrees of freedom), a stopping rule can be built based on statistical bases, such as $\|\underline{\mathbf{r}}_n\|^2 / \sigma^2 \leq \xi$, where $\Pr(u \leq \xi \mid u \sim \chi_N^2) = \eta$, for a probability η fixed at 95%.

The *prewhitening* technique [68] which is equivalent to the *Matching Pursuit* (MP) algorithm [A4] selects at each iteration the atom having the highest correlation with the residual, and estimates the amplitude associated with the correlation coefficient:

$$(i) k = \operatorname{argmax}_k |\underline{\mathbf{w}}_k^H \underline{\mathbf{r}}_{n-1}| \quad \text{and} \quad (ii) x_k = \underline{\mathbf{w}}_k^H \underline{\mathbf{r}}_{n-1}$$

This approach is very sensitive to interferences between the different frequencies as $R_k = |\underline{\mathbf{w}}_k^H \underline{\mathbf{r}}_{n-1}|$ corresponds to the periodogram of the residual for frequency f_k . Thus, on critical spacing of frequencies, such an approach can perform an erroneous selection, generating the propagation of errors in subsequent iterations. FigureA2 shows the results obtained with the MP algorithm on the data of figA1, presenting several false alarms due to an erroneous selection at the 3rd iteration of the algorithm. One way to improve the estimation step (ii) of the MP is to perform an orthogonal projection of the data on all the selected atoms, which leads to the Orthogonal Matching Pursuit (OMP) algorithm, with estimation step (ii) $\underline{\mathbf{x}}_n = \operatorname{argmin}_{\underline{\mathbf{x}}_n} \|\underline{\mathbf{y}} - W_{\Omega_n} \underline{\mathbf{x}}_n\|^2$. However, as the selection step is not modified, such algorithm is still sensitive to erroneous selections, as can be seen in figA2.

The idea leading to the *CLEAN* method [A5, 4], to decrease the sensitivity to erroneous selection, is to build the approximation accounting only for a fraction (typically 10%) of the amplitude associated to a selected atom. Therefore, an atom can be selected at a large number of iterations and the estimation is not easy to grasp, as illustrated figA2. To this end, astronomers generally convolve the estimated frequencies with an ideal spectral window, with only one primary lobe. Even if it is more difficult to interpret in terms of frequency estimation, such heuristic clearly improve the spectrum representation, even if the effect of the false detection is still visible.

A real improvement can be made by greedy algorithms which try to minimize the least squares criterion in the selection step, such as the Orthogonal Least Squares (OLS) [A6] for which (i) $k = \operatorname{argmin}_k \operatorname{argmin}_{\underline{\mathbf{x}}_n} \|\underline{\mathbf{y}} - W_{\Omega_n} \underline{\mathbf{x}}_n\|^2$ with $\Omega_n = \Omega_{n-1} \cup \{k\}$, with the same estimation step than the OMP. The Continuation Single Best Replacement (CSBR) algorithm [A7], although based on a different principle, can be interpreted as a greedy algorithm which proposes, at each iteration, either to select a new atom, either to suppress a previously selected atom, based on the minimization of the least squares criterion penalized with the L_0 -norm. As can be seen in figA2, such approach clearly improves the detection results, but with a much higher computation cost.

Contrary to greedy algorithms which increase the number of detected frequency at each iteration, the aim of global approaches is to estimate the whole frequencies at once, relaxing the L_0 -norm in eq.(A3) to obtain a simpler optimization problem. Typically, the L_0 -norm is relaxed with a L_1 -norm, which leads to minimize a L_1 -norm penalized least squares criterion:

$\hat{\underline{\mathbf{x}}}_\lambda = \operatorname{argmin}_{\underline{\mathbf{x}}} J_\lambda(\underline{\mathbf{x}}) \quad \text{with } J_\lambda(\underline{\mathbf{x}}) = \ \underline{\mathbf{y}} - W\underline{\mathbf{x}}\ ^2 + \lambda\ \underline{\mathbf{x}}\ _1.$	(A4)
--	------

Indeed, used since the 1970s in spike train deconvolution [A8], penalization with the L_1 -norm is known to enforce the sparsity of the solution. It is nowadays the standard method for sparse approximations, also known as *Basis Pursuit Denoising*, and has been the subject of numerous works, both theoretical and applied [53, A2]. Note that such a criterion is convex, so is guaranteed to admit no local minima and can be minimized easily.

In the context of frequencies estimation, such an approach has some specific features:

- It is essential to apply the L_1 -norm on the complex values amplitudes, and not independently on their real and imaginary parts [16]. Indeed, if a spectral line is present at the frequency f_k , both the real and imaginary parts of the associated amplitude are generally non-zero.
- The uniqueness of the solution can be shown if it has less than $P/2$ non-zero components, based on the *Unique Representation Property* (URP) of the matrix W (any set of P columns of W are linearly independent). Actually, it can be shown that for a fixed K and given acquisition time points t_p the values of f_{MAX} for which W does not satisfy the URP are isolated points [17].
- As the data \mathbf{y} are real valued, the solution must satisfy the Hermitian symmetry property, which means that for each frequency f_k detected with amplitude a_k , the frequency $-f_k$ should also be detected with amplitude a_k^* (the complex conjugate of a_k): the minimizer of J_λ is guaranteed to satisfy this property [17].
- It can be shown [17] that, if $\underline{\mathbf{R}} = W^H(\underline{\mathbf{y}} - W\hat{\underline{\mathbf{x}}}_\lambda)$ denotes the Fourier Transform of the residual computed at frequencies f_p , the solution satisfies the following condition:

$$\begin{cases} \forall k \text{ such that } x_k = 0, & |R_k| < \lambda, \\ \forall k \text{ such that } x_k \neq 0, & R_k = \lambda e^{i \arg(x_k)}. \end{cases}$$
 This characterization of the optimality makes it possible to strictly test the convergence of the optimization procedure. It also provides an interesting information on the meaning of the regularization parameter λ , which is the maximum value reached by the periodogram of the estimation residual.
- The L_1 penalization helps to obtain a sparse spectrum and therefore to detect frequencies in the data corresponding to the non-zero coefficients in $\hat{\underline{\mathbf{x}}}_\lambda$. However the estimation of the associated amplitudes is biased [51]. Moreover, it should be noted that in practice, a spectral line is sometimes detected by the two closest frequencies on the grid. In addition to the minimization of J_λ , an a posteriori frequencies correction, for example by simply taking the barycenter of the two neighboring frequencies weighted by their estimated amplitudes, and a least squares re-estimation of their amplitudes, allow to improve their accuracy as can be seen in figA3.

Finally, a large number of algorithms have been proposed in the context of sparse approximations to minimize the L_1 -norm penalized least squares criterion. A large majority takes the form of projected gradient algorithms, alternating between a stage of descent, guided by the gradient of the quadratic part of the criterion, and a thresholding step [A9]. These methods have been developed in relation with large-size problems (in particular, for images), where sparsity is often expressed in a widely incoherent dictionary, and the product of vectors with matrices W and W^H can be computed with efficient algorithms (Fast Fourier Transform, Orthogonal Wavelet Transform, ...). These properties condition the complexity of optimization: the closer the dictionary will be to orthogonality, the more significant will be the descent step and the more effective gradient-computation based approaches will be. In contrast, for frequencies estimation, the dictionary is extremely correlated due to the required high frequency resolution and the shape of the spectral window, no fast algorithm is available associated with W , but the size of the data remains limited due to their one-dimensional nature. The above mentioned algorithms can then appear to be pathologically slow. It should also be noted that the inability to use a formulation with a L_1 -norm involving real variables excludes any quadratic programming-based optimization approach as well as the powerful *homotopic* methods [A10]. Conversely, an algorithmic scheme such as *Iterative Coordinate Descent*, alternating minimizations with respect to every complex variable x_k , proved particularly effective for this problem [17]. Indeed, if this strategy proves to be generally inefficient, it allows several properties of the criterion to be exploited: (i) each minimization has a low computational-cost analytical solution; (ii) the solution being sparse, the number of non-zero components drops quickly during the iterations, hence a selective scan of the coordinates can then be performed, frequently updating the non-zero components; (iii) the non-differentiable part of the criterion being separable in the unknowns, this strategy is guaranteed to converge to the minimal value of the criterion. Such characteristics are fully taken into account in the *SparSpec* software [16].

Considering now the particular application of *SRS* methods to magnetically confined thermonuclear fusion plasmas in a tokamak, the analysis is based on magnetic and turbulence measurements, and typically starts with an initial Fourier decomposition of the data in the time/frequency domain to obtain the individual frequency components $\psi(\omega)$. In a tokamak the plasma column has, to a first approximation, 2D boundary conditions along the toroidal direction and on the poloidal plane, perpendicular to the toroidal direction. The spatial structure of the magnetic instabilities is determined by further decomposing each frequency component in its toroidal (n) and poloidal (m) harmonics:

$$\psi(\omega) = e^{-i\omega t} \sum_{n,m} A_{nm} e^{in\phi} e^{im\theta}. \text{ Here } \phi \text{ and } \theta \text{ are the toroidal and poloidal angle coordinates, respectively,}$$

and we have used the fact that in tokamak geometry one single toroidal component with a given n usually has multiple poloidal components due to toroidicity and various other geometrical effects.

The aim of mode number analysis is to determine the value of $\{n, m\}$ of the magnetic instabilities present in the plasma and to estimate their amplitude from data acquired with P detectors un-evenly positioned at angles ϕ_p (θ_p), $p=\{1, \dots, P\}$ being the suffix labelling the individual sensors used for the measurement.

Considering now for simplicity of notation the specific case of toroidal mode number analysis, each measurement $y(t_p)$ can be mathematically modelled with a slight variation of eq.(A1):

$y(\phi_p) = \sum_{l=1}^L \alpha_l e^{jn_l \phi_p} + \varepsilon_p,$	(A5)
--	------

where n_l and α_l are the unknown mode numbers and amplitudes, respectively, L is the unknown number of modes and ε_p corresponds to the noise on the data for the given p -th sensor, and periodic boundary conditions in ϕ have been used. Thus, the mode detection problem is strictly equivalent mathematically to the A&A spectral analysis problem.

Evaluating the amplitudes α_l and the mode numbers n_l of multiple modes in a multi-harmonic spectrum is a very difficult problem, even if the number of modes in the input spectrum is actually known a-priori. The usual way to tackle this problem is performing a best LS fitting of the input data. However, this criterion has many local minima for real valued spectral peaks [48, 52], hence in principle requiring a combinatorial exploration for integer-valued mode numbers n_l , and an a-posteriori thresholding scheme to differentiate the “correct” from the “wrong” solutions, in many aspects similar to the SVD-based method described in [7]. This is a very CPU-time intensive process and cannot possibly be adapted for real-time applications on the sub-millisecond time scale required for the real-time analysis of the JET (and ITER) measurements. An alternative solution consists of providing an estimate for the amplitudes of all possible mode numbers in the range $\{-K, \dots, K\}$ (where $|K|$ is much larger than the maximum mode number that can be conceivably present in the input spectrum), at the same time enforcing that most of these modes actually have a null amplitude, i.e. a utilizing a *Sparse Approximation*.

The mode detection (i.e. the spectral analysis) problem is particularly difficult in the case of tokamak plasma physics as the data is un-evenly sampled and sparse, because of unavoidable installation constraints on the measurement devices. It can be shown that the difficulty of the spectral analysis problem is closely related to properties of the spectral window, such as the height and positions of its secondary lobes. Indeed, the mathematical problem described by eq.(A5) can be expressed equivalently in the Fourier spatial domain as:

$$Y(\nu) = W(\nu) * \sum_{l=1}^L \alpha_l \delta(\nu - n_l) + E(\nu) = \sum_{l=1}^L \alpha_l W(\nu - n_l) + E(\nu), \quad (\text{A6})$$

where ν is the spatial frequency, $Y(\nu)$ and $E(\nu)$ corresponds to Fourier transform with respect to the angular position ϕ of the data and the noise, the symbol “*” is the convolution operator and $W(\nu)$ is the spectral window of the sampling scheme. Thus, if $W(\nu)$ has a high secondary lobe (with an amplitude close to 1) at a frequency ν_0 , a mode number n will produce in the Fourier transform $Y(\nu)$ a maximum at $\nu_1=n$ and a secondary maximum at $\nu_2=n\pm\nu_0$, with an amplitude proportional to $W(\nu_2 - \nu_1) = W(|\nu_0|)$. This means that it will be difficult to distinguish the actual mode n from the two “aliases” modes at $n\pm\nu_0$, even in the absence of noise in the data. If we consider for simplicity that obtaining the “true” mode number n_0 obeys a normal probability distribution with variance $\sigma^2(n_0)$, i.e. $P_{TRUE}(n = n_0) \propto \exp\left[-(n = n_0)^2 / \sigma^2(n_0)\right]$, then we also find (using a best fit of the data that led to the analysis presented in [25, 26]) that the probability $P_{FALSE}(n=n_0)$ for a *false detection* of n as n_0 due to the secondary lobes in the spectral window when using the *SparSpec* algorithm is $P_{FALSE}(n=n_0) \propto \sigma^2(n_0) \times \sum_i (W(n_i) + W(|n_i|=\nu_0)) / W(n_0)$, where the sum is intended on all combinations of n_i and ν_0 such that $n_i = n_0 \pm \nu_0$.

In thermonuclear (tokamak) plasmas these lobes are due to regularities in the sampling (for instance when using a spacing larger than the Nyquist condition) and to the low number of sensors. This situation is further compounded by the failure of sensors over time, a problem that cannot easily be rectified due to restricted in-vessel access. As an example, the spectral window for two families of JET high-frequency magnetic sensors is shown in figA4, comparing the data for the original complete set of 11 sensors that could be used in 1997 for toroidal mode number analysis, and then for the 7 and remaining 4 sensors in that set which could be used for real-time and post-pulse analysis of MHD instabilities in 2013 and as of today, respectively. Note that the original dominant $\nu_0 = \pm 10$ secondary lobe has now been supplemented from 2013 by an even higher secondary lobe at $\nu_0 = \pm 4$, which is much more difficult to deal with as the most interesting n -number range is actually within $|n|=1$ and $|n|=10$. Moreover, the failure of three additional sensors between 2013 and the end of 2015 means that now there is an even more problematic secondary lobe at $n = \pm 1$ with $W(\nu) = 0.62$.

When applied to thermonuclear plasma physics, the problem described by eq.(A6) has some additional requirements with respect to the A&A problem described by eq.(A1), even if its solution can still be obtained using eq.(A4). First, the data are complex-valued, implying that the Fourier transform of the data does not satisfy the Hermitian property $\hat{y}(-\nu) = \hat{y}^*(\nu)$ as in the spectral analysis

of real-valued data. Obviously, the complex-valued data have to be analysed together, conserving the I/Q phase relation between them, and not independently. Second, the mode numbers n_l can only take positive or negative integer values, while in the general spectral analysis problem frequencies take real values. This is a favourable property as the model of eq.(A1) works on a discretized frequency grid. For A&A problems, a very fine discretization of the frequency grid is required so that real valued frequencies are not too distant from the nearest frequency on the grid. However, as for A&A data, the maximum mode number is unknown. Therefore, one has to account for a mode number in the range $\{-K, \dots, K\}$. Note however that a posterior estimation of the out-grid estimation of the detected frequencies can be performed, i.e. using a barycentric estimation of the neighbour frequencies (as we are interested in all the $|x_k| \neq 0$ components and not in the approximation of the signal as $y \sim \sum_k W_k x_k$) [16]. Third, in the real time applications we consider for JET (and for ITER), a set of data is acquired every 1ms, therefore the spectral analysis must be completed in an automated manner in the short time between each measurement acquisition.

Therefore, the choice of the family of atoms is critical in the *Sparse Representations* (and *Sparse Approximations*) of signals as, with an appropriate choice, these atoms might be well adapted to the signal to be analysed and might lead to a matrix \mathbf{W} with good analytical and numerical properties. For these reasons, the matrix \mathbf{W} is often chosen as a family of relatively uncorrelated atoms, such as wavelets, Diracs, pure sine waves, etc ..., but cannot be chosen arbitrarily as it is not guaranteed to have the required good properties, as we will show later. For the spectral analysis problem that we analyse here the atoms are driven by the problem itself and therefore we use $w_k = \exp(2i\pi f_k t_p)$, for $p=\{1, \dots, P\}$. Moreover, the sparsity of the components x_k , hence the L_1 -norm, has to be computed on the modulus of the complex amplitudes x_k , while the *Sparse Approximation* problem is generally studied, only for real-value amplitudes. Finally, note that $|x_k| \neq |\operatorname{Re}(x_k)| + |\operatorname{Im}(x_k)|$, so imposing the sparsity on the complex modulus is radically different in terms of the model than sparsity applied separately on the real and imaginary components [16].

For the analysis presented here, the atoms are imposed by the model setup in eq.(A5) to be pure complex exponential waves, $\overline{\overline{W}} = \left\{ \exp(in_k \phi_p) \right\}_{k,p}$, for $p=\{1, \dots, P\}$ and $k=\{1, \dots, M\}$, with $n_k = k - K + 1$ and $M = 2K + 1$. Due to the irregular sampling, the atoms are strongly correlated. Indeed, it can be shown that $\left| \overline{w_k}^H \overline{w_l} \right| = W(n_k - n_l)$, so that it corresponds to regular samples of the spectral window. As $W(v)$ may take values greater than 1/3 (as shown in figA4), the previous sufficient condition guarantees exact detection only if the signal consists of a single mode number. Nevertheless, it has been shown from many simulations and analysis of measurements using

comparisons between different numerical methods that such a solution generally gives very satisfactory results in terms of detection, even in the case of multiple modes [18, 25, 26, 48]. Moreover, for irregular sampling, uniqueness of the global minimizer is almost surely guaranteed if it has less than $P/2$ non-zero components, where P is the data size [16, 25, 26].

In terms of amplitude estimations, it has been shown [16, 17] that minimizing eq.(A4) leads to an under-estimation of the amplitudes of the detected mode numbers due to the L1-norm penalization term. Thus, an a-posteriori LS re-estimation of these amplitudes is usually performed for post-pulse analysis in a second step within the calculations, after the modes have been actually detected. Their amplitudes are computed by minimizing the least square criterion $\|\mathbf{y}-\mathbf{W}_{\text{DET}}\mathbf{x}_{\text{DET}}\|^2$ where only the non-zero amplitudes of the optimization step are preserved in \mathbf{x}_{DET} . Note that this a-posteriori amplitude estimation step is not an absolute necessity for the real-time analysis, as its main objective is to detect the actual modes, their mode numbers and frequency width, and not to precisely estimate their absolute amplitudes, a scaled value being sufficient for this purpose.

Many numerical algorithms are available to minimize criteria such as those of eq.(A4) for *Sparse Approximations*. While for real-valued unknown $\{\mathbf{x}_k\}$ this problem can be written as a classical Quadratic Program, for complex-valued unknown $\{\mathbf{x}_k\}$ it corresponds to a Second-Order Cone Program [49]. An algorithm based on an iterative Block Coordinate Descent procedure has been previously proposed [16, 17], and implemented in the *SparSpec* code. This procedure consists of performing successive one-dimensional minimization steps with respect to each complex-valued unknown \mathbf{x}_k , where each one-dimensional minimization has an explicit solution. This algorithm is very efficient and a correct solution can be typically found in less than 1ms using the rather modest computational resources available to process real-time JET data [19, 20, 27].

A real-time implementation of the proposed modes detection method requires not only an efficient optimization algorithm to minimize eq.(A4) but also, and even more importantly for a frequency-degenerate spectrum, an efficient automated tuning of the penalization parameter λ . The penalization parameter λ is related to the noise level [16] and requires an appropriate tuning, since it increases the penalty for those solutions which invoke a larger number of modes. The first order necessary and sufficient optimality conditions for convex non-differentiable functions (often known as the Karush-Kuhn-Tucker optimality conditions [A11-A13]), provide a physical interpretation for λ : (a) for

$\lambda > \lambda_{\text{MAX}} = \max_k \left[\left| \overline{w}_k^H \left(\overline{y} - \overline{\overline{W}} x_{\text{MIN}} \right) \right| \right] = \max_k \left[|Y|(n_k) \right]$, the minimizer \mathbf{x}_{MIN} of eq.(A4) is identically

zero, i.e. the unique solution has no detected modes; (b) for a given λ , the minimizer \mathbf{x}_{MIN} of eq.(A4)

satisfies $\max_k \left(\left| \overline{w}_k^H \overline{r} \right| \right) = \max_k \left[|R|(n_k) \right] < \lambda$, where $\overline{r} = \overline{y} - \overline{\overline{W}} x_{\text{MIN}}$ is called the residual (data minus

paper PPCF2016-V3: topical review SparSpec, for proofs September 2016 the model corresponding to the estimated modes). Hence λ can be interpreted as the maximum peak amplitude allowed in the FT modulus of the residual, and choosing λ to be a fraction $\lambda_{\text{NORM}} \in [0, 1]$ of the maximum of the FT of the data $\lambda = \lambda_{\text{NORM}} \times \max(|\mathbf{W}^H \mathbf{y}|)$, ensures the FT of the residual \mathbf{r} to be lower up to this fraction relative to the maximum of the data FT. Hence knowledge of the noise level in the measurements helps to determine the optimum value for λ_{NORM} to be used for real-time and post-pulse analysis of the magnetic fluctuation data.

Additional references for Appendix-A.

1. L.Eyer, P.Bartholdi, *Astronomy and Astrophysics Supplement Series* **135(1)** (1999), 1.
2. S.Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, Academic Press (2008).
3. E.J.Candès, J.K.Romberg, T.Tao, *Communications on Pure and Applied Mathematics* **59(9)**, (2006), 1207.
4. S.Mallat, Z.Zhang, *IEEE Transaction on Signal Processing* **41(12)** (1993), 3397.
5. J.A.Högbom, *Astronomy and astrophysics Supplement* **15** (1974), 417.
6. T.Blumensath, M.A.Davies, *On the difference between Orthogonal Matching Pursuit and Orthogonal Least Squares*, Technical report, Edinburgh University (2007).
7. C.Soussen et al., *IEEE Transactions on Signal Processing* **63(13)**, (2015), 3301.
8. H.L.Taylor, S.C.Banks, J.F.McCoy, *Geophysics* **44(1)**, (1979), 39.
9. I.Daubechies et al., *Communication on pure and applied mathematics* **57(11)**, (2004), 1413.
10. M.R.Osborne, B.Presnell, B.A.Turlach, *IMA Journal of Numerical Analysis* **20(3)** (2000), 389.
11. H.W.Kuhn, A.W.Tucker, *Nonlinear Programming*, Proceedings of 2nd Berkeley Symposium (Berkeley: University of California Press, 1951), 481.
12. W.Karush, *Minima of Functions of Several Variables with Inequalities as Side Constraints*, M.Sc. Dissertation (1939), Dept. of Mathematics, Univ. of Chicago, Chicago (USA).
13. D.P.Bertsekas, *Nonlinear Programming* (2nd edition: 1999), Athena Scientific (Belmont, USA).

Appendix-B: Applications of a Sparse Representation method and of the SparSpec algorithm to the analysis of magnetic fluctuations in JET tokamak plasmas.

The overall accuracy of the *SparSpec* code for the analysis of post-pulse and real-time data is evaluated using simulations performed on a synthetic dataset $S_{IN}(v_n)$:

$S_{IN}(v_n) = \left[\sum_{k=-L_{MAX}}^{k=L_{MAX}} A_k \exp(jl_k v_n + j\delta_k) + \sigma_{SIG} \times (r_{1k} + jr_{2k}) \right] + \sigma_{MEAS}(v_n) \times (r_{3n} + jr_{4n}). \quad (B1)$	
---	--

Here $v_n \in [0, 2\pi]$ is the position of each sensor, and the input signal $S_{IN}(v_n)$ is constructed as an arbitrary superposition of different components at the integer mode numbers l_k , $k \in [-L_{MAX}, L_{MAX}]$, where L_{MAX} is the highest mode number in the spectrum. Each l_k spectrum component has a fixed or randomized amplitude A_k and relative phase δ_k . The quantities $\sigma_{SIG} \in [0, 1]$ and $\sigma_{MEAS}(v_n) \in [0, 1]$ represent the standard deviation in the background noise on each spectral component and on the measurement itself at each sensor, respectively, and are known a priori (i.e. they have a fixed and unique value for each simulation) as they can in principle be measured directly on the system when installed. The quantities $\{r_{1k}, r_{2k}, r_{3n}, r_{4n}\}$ are random numbers chosen from a uniform distribution in the interval $[0.0 \rightarrow 1.0]$; note that the random seed used for $\{r_{1k}, r_{2k}\}$ can be different from the one used for $\{r_{3n}, r_{4n}\}$. With this approach, the noise has independent and un-correlated complex components satisfying the circularity property. In general, σ_{SIG} and σ_{MEAS} can be different and, more importantly, σ_{MEAS} can have different values for different sensors. Intuitively, σ_{SIG} can be associated to background noise from the plasma, for instance due to un-coherent turbulence; conversely, σ_{MEAS} is associated with “engineering” errors, such as tolerances on the position and alignment of the sensors, calibration errors, and various effects such as cross-talk, drifts, offset, signal pick-up and bit-noise in the cabling and electronics. These accuracy tests are performed using Matlab R14 on a 2GHz laptop with 1024MB of RAM.

The “confidence level” in the *SparSpec* calculations is then defined by comparing the output results for $\{A_k, l_k\}$ to their input value:

$\text{confidence level} = \exp \left(- \frac{(OutputData - InputData)^2}{(TotalVarianceInputData)^2} \right).$	(B2)
--	------

In eq.(B2) the total variance on the input data is taken as σ_{TOT} (P being the total number of sensors):

$TotalVarianceInputData = \sigma_{TOT}$ $\sigma_{TOT} = \sqrt{\frac{1}{2L_{MAX}} \sum_{k=-L_{MAX}}^{k=L_{MAX}} \sigma_{SIG}^2 r_{1k} + ir_{2k} ^2 + \frac{1}{P-1} \sum_{n=1}^P \sigma_{MEAS}^2 (\nu_n) r_{3n} + ir_{4n} ^2} . \quad (B3)$	
---	--

This definition of a confidence level is a very stringent criterion to measure whether the scientific requirements for the accuracy of the *SparSpec* code are satisfied. A normal distribution of the output data centred on an expectation value provided by the input data, with the variance taken as the total variance on such data, i.e. precisely as the ones given in eqs.(B2, B3), meets the accuracy requirements provided we achieve a confidence level in excess of $e^{-1/4}=0.7788$. This value for the confidence level is obtained when the absolute difference between output and input data is less than half the total variance on the input data.

We have translated this theoretical definition of a confidence level for the analysis of simulated data into the required accuracy for the analysis of actual data by matching the nominal tolerances for the ITER measurement requirements for high-frequency instabilities [22-24], specifically with respect to the accuracy of the toroidal mode number determination. This is very important as, for instance, a nominal 10% or a ± 1 tolerance on the toroidal mode number determination leads to very different requirements for the system depending on which “specific” mode number this tolerance is applied to. In terms of physics interpretation and for real-time plasma protection and control applications, wrongly interpreting the $n=1$ mode as an $n=0$ or an $n=2$ mode (i.e. $n=n\pm 1$) clearly does not have the same implications as wrongly interpreting the $n=10$ mode as an $n=11$ or an $n=9$ mode, i.e. still having a ± 1 error on n . Hence, the confidence level defined in eq.(B2) corresponds to the ability of obtaining the requested quantity with an accuracy given by the ITER measurement requirements for high-frequency instabilities.

Hence, we have decided to consider that the acceptable error is ± 0 on the toroidal mode number and $\pm 15\%$ on the measured mode amplitude $|\delta B_{MEAS}|$ for low- $|n| \leq 5$ modes of importance for plasma protection and control and for real-time measurements. Examples of these modes are the precursor for sawteeth, Edge Localized Modes and disruptions, and radially extended MHD instabilities such as global Alfvén Eigenmodes, Neoclassical Tearing Modes, Resistive Wall Modes and Alfvén Cascades. Conversely, a mode number measurement error ranging from ± 1 to ± 3 is deemed to be acceptable for MHD instabilities which are only of “physics” interest, for instance core localized Alfvén Eigenmodes with $|n|=6 \rightarrow 20$, for which the amplitude $|\delta B_{MEAS}|$ only needs to be measured within $\pm 30\%$. The required post-pulse measurement accuracies on the mode amplitude and toroidal mode number and can then be summarized as follows in TableB1:

toroidal mode number	acceptable error on toroidal mode number	acceptable error on mode amplitude
$ n \leq 5$	error=0 → CORRECT error≠0 → WRONG	error≤15% → CORRECT error≥15% → WRONG
$ n > 5$	$ \text{error} \leq \min(n /10, 3)$ → CORRECT $ \text{error} \geq \min(n /10, 3)$ → WRONG	error≤30% → CORRECT error≥30% → WRONG

TableB1: the measurement requirements on the toroidal mode number and mode amplitude which are used to define *correct* and *wrong* mode detection with the *SparSpec* algorithm.

FigureB1 shows the confidence level in achieving the ITER measurement requirements for the evaluation of the toroidal mode number and mode amplitude when applying the *SparSpec* algorithm to a synthetic dataset defined as in eq.(B1) using the high-frequency magnetic sensors available in JET in 1997, 2013 and 2016 (as for the spectral window data shown in figA4). More technical details on this method can be found in [25, 26]. The result shown in figB1 represent an overall summary of in excess of 50'000 simulation runs, using a frequency-degenerate input mode spectrum that consists of up to 10 (toroidal) modes with the same mode frequency, with randomized relative amplitudes in the range $A_k = 0.05 \rightarrow 1.00$ and relative phase in the range $\delta_k = 0 \rightarrow 1.95 * \pi$, and input toroidal mode numbers in the range $|n| \leq 30$. The two main *SparSpec* run-time analysis parameters λ_{NORM} and f_{MAX} , where f_{MAX} is the size of the dictionary, were scanned in the range $0.05 \leq \lambda_{NORM} \leq 0.95$ and $60 \leq |f_{MAX}| \leq 200$, respectively. For the 1997 and 2013 setups with 11 and 7 magnetic sensors, respectively, we note that the confidence level in the *SparSpec* calculations is very high, exceeding the nominal threshold value $=0.7788=e^{-1/4}$, up to at least $|n|=15$, and only drops significantly below this threshold for higher mode numbers $|n|>25$. These very high- $|n|$ modes are of no concern for real-time control applications in JET, and are also of relatively minor interest for JET physics but for incoherent turbulence studies. Conversely, the confidence level is very poor already for $|n|>5$ for the current configuration of 2016 with only 4 surviving magnetic sensors. Clearly, this configuration is no longer sufficient for the analysis of magnetic instabilities in JET.

It is also important to note that, strictly speaking, the sparsest solution minimizes the LS criterion penalized with the L0 (pseudo) norm, i.e. the number of non-zero components in the solution. However, minimizing such a criterion requires an exploration of all possible combinations of modes in the input dataset, similarly to the SVD technique proposed in [7], which is very demanding in terms of CPU-time consumption. Such an exploration is in fact avoided in *SparSpec* by considering the L1-norm, i.e. the sum of the absolute values of the mode amplitudes, instead of the L0 (pseudo) norm penalization. In practice, minimizing this L1-norm penalized LS fitting criterion is much easier than

minimizing the original one based on the L_0 (pseudo) norm, and many computationally efficient algorithms have been developed, some of which can be made compatible with a real-time system using a 1kHz clock-time. However, minimizing a penalization criterion based on the L_0 pseudo-norm does not necessarily lead to the same solution as minimizing a penalization criterion based on the L_1 -norm, i.e. sufficient conditions for the equivalence between the L_0 (pseudo) norm and L_1 -norm need to be satisfied. Much theoretical work has been performed to determine the conditions of equivalence between the L_0 (pseudo) norm and the L_1 -norm penalization criteria (see for instance [41, 43, 44]).

For example, it can be shown that if the signal can be represented with $\|\bar{x}\| < (1+1/\mu)/2$ components, with $\mu = \max_{k \neq l} \left(\left| \overline{w_k^H w_l} \right| \right)$, where \mathbf{w}^H is the Hermitian transposition of \mathbf{w} , then minimizing both penalization criteria will lead to the selection of the same atoms. Hence, the accuracy of our calculations (both real-time and post-pulse) is also guaranteed by the comparison between a model input spectrum (with/out background noise) and the output spectrum as calculated by *SparSpec* using the actual geometry of magnetic sensors.

Figure4 and fig5 in the main text of this paper shows an overview of the measurement of the damping rate, mode frequency and mode amplitude for TAEs with different toroidal mode numbers for various JET discharges. These results were obtained using the real-time and the post-pulse implementation of the *SparSpec* algorithm. A few points need to be noted for these results.

	<i>SparSpec-RT</i>	<i>SparSpec-PP</i>
calibration	fixed value at 200kHz	full frequency-dependence
input data	up to 8 complex-valued signals from magnetic measurements only one normalization signal for the antenna drive	up to 16 complex-valued signals from magnetic and turbulence measurements selection between many normalization signals for the antenna drive
output data	many amplitude and phase pairs, one pair for each selected mode number	many amplitude and phase pairs, one pair for each selected mode number
algorithm	least-square renormalization of output amplitude not implemented	least-square renormalization of output amplitude implemented
CPU limit	850μsec @1GHz	un-limited (user choice of hardware)
RAM limit	512MB	un-limited (user choice of hardware)
maximum n 	n ≤ f _{MAX} =15: mode number determined in real-time with relative error Δn/n ≤0.1	n ≤ f _{MAX} =30: mode number determined post-pulse with relative error Δn/n ~0.1

<p>TableB1: comparison and differences between the real-time and post-pulse implementation of the <i>SparSpec</i> algorithm.</p>

The main differences between the real-time and post-pulse implementation of the *SparSpec* algorithm are described in details in [27] and can be summarized as follows in TableB1, where we use the labels “RT” and “PP” to indicate the real-time and post-pulse implementations of this algorithm. These differences are essentially due to the CPU and RAM limitations of the real-time analysis, which is performed in JET by the AELM system using a commercial real-time off-the-shelf Emerson Network Power VMEbus 5500 card with a 1GHz PowerPC and 512MB RAM, executing software running under the Wind River VxWorks operating system (i.e. the same software used in the NASA’s Mars rovers, Spirit and Opportunity). The CPU time limit of $<850\mu\text{sec}$ is related to the AELM being a *hard* real-time embedded system: the results of missing a deadline are classed as a failure. For JET operation this results in missing data or a wasted experiment, but avoids the worst case, which would be damage to the AEAD or to the JET machine itself. This is contrary to a *soft* real-time embedded system, which would simply produce a reduced quality of service, such as a Graphical User Interface suffering sluggish mouse response. Post-pulse processing does not suffer of these limitations in computing power, hence allows for a much greater flexibility and scope in the analysis, but at the expense of a much longer computational time.

The typical uncertainty on the measurement of the mode frequency is within 50Hz, due to the accuracy of the digital synchronous detection used in the AEAD system. For the accuracy on the determination of the mode numbers one has to consider the possible statistical and systematic errors due to the algorithm used to extract such data. Typically, the toroidal mode number can be determined exactly (i.e. $n=n\pm 0$) up to $|n|\sim 10$, and it is then subject to at least a ± 1 or $\sim 10\%$ error for higher- $|n|$ modes. The amplitude of such $|n|<10$ modes is then known to within a factor ~ 2 , and the damping rate is subject to an uncertainty of the order of 15% for the typical cases that we have considered in this analysis. Many different toroidal mode numbers are present in the frequency-degenerate spectrum of stable TAEs, and these modes have amplitude and damping rate that can easily vary by up to a factor 10 at any given time point. Hence, an accurate real-time mode number discrimination is clearly an essential ingredient for obtaining damping rate and (at least relative) amplitude data that could meaningfully be used for plasma control purpose.

Appendix-C: Optimization of the design of the ITER HF magnetic diagnostic system.

The 2 arrays for toroidal mode number detection are made with $2 \times 18 = 36$ equi-spaced sensors each (indicated by red dots in fig9a in the main text) and are positioned at two different heights with respect to the geometrical centre of the machine, using the corners of each equatorial port on the low field side of the vacuum vessel wall. Hence, these two arrays have by construction a $20\text{deg}/18$ -fold periodicity, giving an intrinsic (pseudo-) Nyquist number $|n|=18$ on each of the two periodic sub-assemblies, whereas the ITER measurement requirements specify accurate detection of modes up to $|n|=30$. (We use here the wording pseudo-Nyquist as in fact $W(|n|=18)=0.2$ and not $=1$ as for a true Nyquist value: this occurs because the two equi-spaced sub-assemblies with 18 sensors each are spatially separated and do not correspond to a single equi-spaced assembly with 36 sensors, which would indeed have $W(|n|=18)=1$ exactly). This system design has another weakness, namely the absence of sensors for n-number detection on the high field side of the ITER vacuum vessel. This does not allow distinction between ballooning and anti-ballooning instabilities, nor can these arrays be used to diagnose MHD modes during start-up plasmas limited on the high-field side wall.

The poloidal mode number detection system is built around 18 un-evenly spaced sensors located on six (out of the nine) machine sectors, covering the entire poloidal cross-section but for the divertor region. The addition of the in-vessel active ELM coil assemblies to the ITER design would reduce the six m-number arrays to 16 sensors each, as the position of two of the sensors clashes with that of the ELM assemblies. This poloidal mode number detection system uses a large number of sensors, but not optimized: the array redundancy is significant, but the measurement system essentially suffers from a limited number of sensors in each one array.

The toroidal and poloidal mode number detection systems can be improved in the ITER original design layout, with initial provisions already being made for this, by adding a high-resolution mini-array on the horizontal (toroidal analysis) and vertical (poloidal analysis) edges of some of the equatorial ports on the low-field side, as shown in fig9b in the main text of this paper. For toroidal mode number analysis, the addition of such high-resolution arrays will in principle remove the $n=18$ toroidal Nyquist value by adding un-evenly spaced sensors to the two baseline periodic sub-assemblies. For poloidal mode number analysis, only one high-resolution array can be added, and this will increase the number of sensors that can be used for such measurements.

Finally, note also that the ensemble of the m-numbers measurement arrays in the six machine sectors give rise to 14×6 -sensors arrays for n-number detection (two of the sensors in the poloidal measurement arrays are in fact located at the corners of the equatorial ports, hence are common with the toroidal measurement arrays). These arrays can in principle be used to detect, and possibly

remove, low- n modes up to $|n| \leq 3$ from the measured spectrum, for instance for real-time control and plasma protection applications.

We have performed the baseline analysis and optimization of the ITER HF magnetic diagnostic system using a new approach based on the SRS method. The *SparSpec* algorithm has been applied to a model dataset of input modes for various implementations of the ITER HF magnetic sensor geometry for $n(m)$ -number detection. The ITER measurement requirements and the expected measurements' errors and tolerances, as highlighted in TableC1, are explicitly considered in this algorithm to define the *correct* and the *wrong* detection of the modes.

toroidal mode number	acceptable error on toroidal mode number	acceptable error on mode amplitude
$ n \leq 5$	error=0 → CORRECT error≠0 → WRONG	error≤15% → CORRECT error≥15% → WRONG
$ n > 5$	$ \text{error} \leq \min(n /10, 3)$ → CORRECT $ \text{error} \geq \min(n /10, 3)$ → WRONG	error≤30% → CORRECT error≥30% → WRONG

TableC1: the measurement requirements on the toroidal mode number and mode amplitude which are used to define *correct* and *wrong* mode detection with the *SparSpec* algorithm.

As the ITER vacuum vessel is still undergoing design changes, a system optimization that takes fully into account in-vessel engineering constraints is not yet possible. Hence, in addition to the physics constraints for the measurement requirements, a *cost function* has been included in the optimization algorithm to reflect the currently foreseen procurement and installation costs for the sensors. This cost function is constructed as follows:

- (1) each individual sensors costs 7→10 cost-units end-to-end, i.e. from the initial R&D, to the detailed design and manufacturing, and from installation to the final data acquisition;
- (2) each high-resolution sensor in any of the equatorial ports bears an additional installation cost of 1→2 cost-units due to the different needs for mechanical fixing, requiring further R&D work and additional mechanical interfaces with the vessel structure;
- (3) each poloidal sensor located in the regions $60 < \theta(\text{deg}) < 120$ and $270 < \theta(\text{deg}) < 315$ bears an additional installation cost of 1→2 cost-units, due to more difficult cabling access;
- (4) each high-field side poloidal sensor located in the region $120 < \theta(\text{deg}) < 220$ bears an additional installation cost of 2→3 cost-units, again due an even more difficult cabling access;
- (5) each high-field side poloidal sensor located in the divertor region $220 < \theta(\text{deg}) < 270$ bears an additional installation cost of 4→7 cost-units, again due to an even more difficult in-vessel cabling access and to need for improved RF screening of image and eddy currents;
- (6) finally, if we have more than 8 toroidal sensors (including high-resolution ones) in any one of

the 9 machine sectors, the cost increases by $1 \rightarrow 2$ cost-units for each additional group of 8 sensors due to need of installing one further cabling loom in that sector.

The ratio between the confidence level in the measurement performance, and the costs necessary to achieve this performance, as defined above, can then give an additional indication of the overall system performance, one where we have integrated physics and budgetary requirements: the highest ratio defines the cheapest (financially) way to obtain a satisfactory measurement performance.

The baseline analysis and system optimization is performed by scanning various parameters for the input mode spectrum: number of components with their relative (truly) randomized amplitude, phase and mode number, and standard deviation σ in the background (white) Gaussian noise. The different elements of the cost function described above are then added for each particular sensors' arrangement, and the overall average is taken as the measure of the cost-normalized measurement performance. Finally, various run-time analysis parameters are in principle required for the *SparSpec* calculations, as described in details in [25-27]. However, only the value of λ_{NORM} and the size (f_{MAX}) of the *SparSpec* dictionary are actually of paramount importance for the simulations reported here: using these previous results, we set $\lambda_{\text{NORM}}=0.85$ and f_{MAX} to be five times larger than the maximum physical mode number present in the input spectrum.

Figure1, D.Testa et al., PaperPPCF2016.

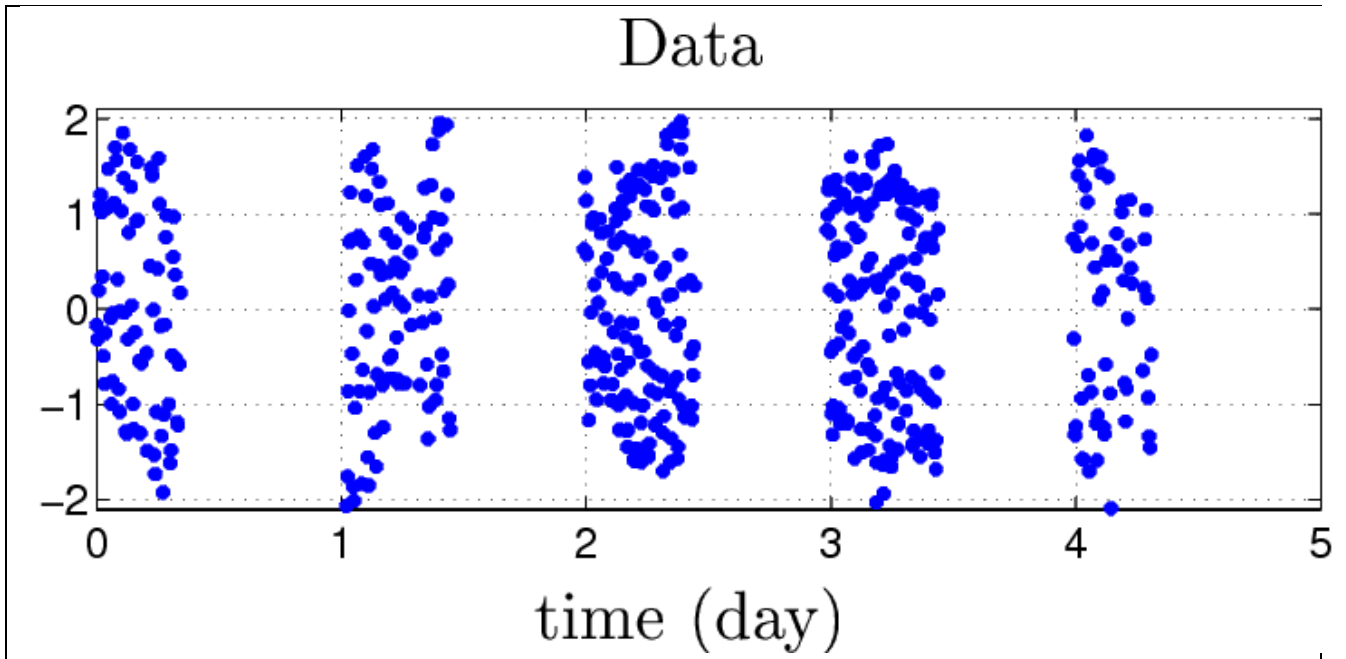


Figure1 (reproduced from [27], the original measurements have been reported in [59, 60]). Observation for the radial velocity curve of the Herbig Ae star HD 104237. These data correspond to five observing nights of high resolution spectroscopy at the South African Astronomical Observatory during April 1999. The irregular data sampling due to the day/night alternation is very clear, leaving large gaps in the temporal coverage of the measurements.

Figure2, D.Testa et al., PaperPPCF2016.

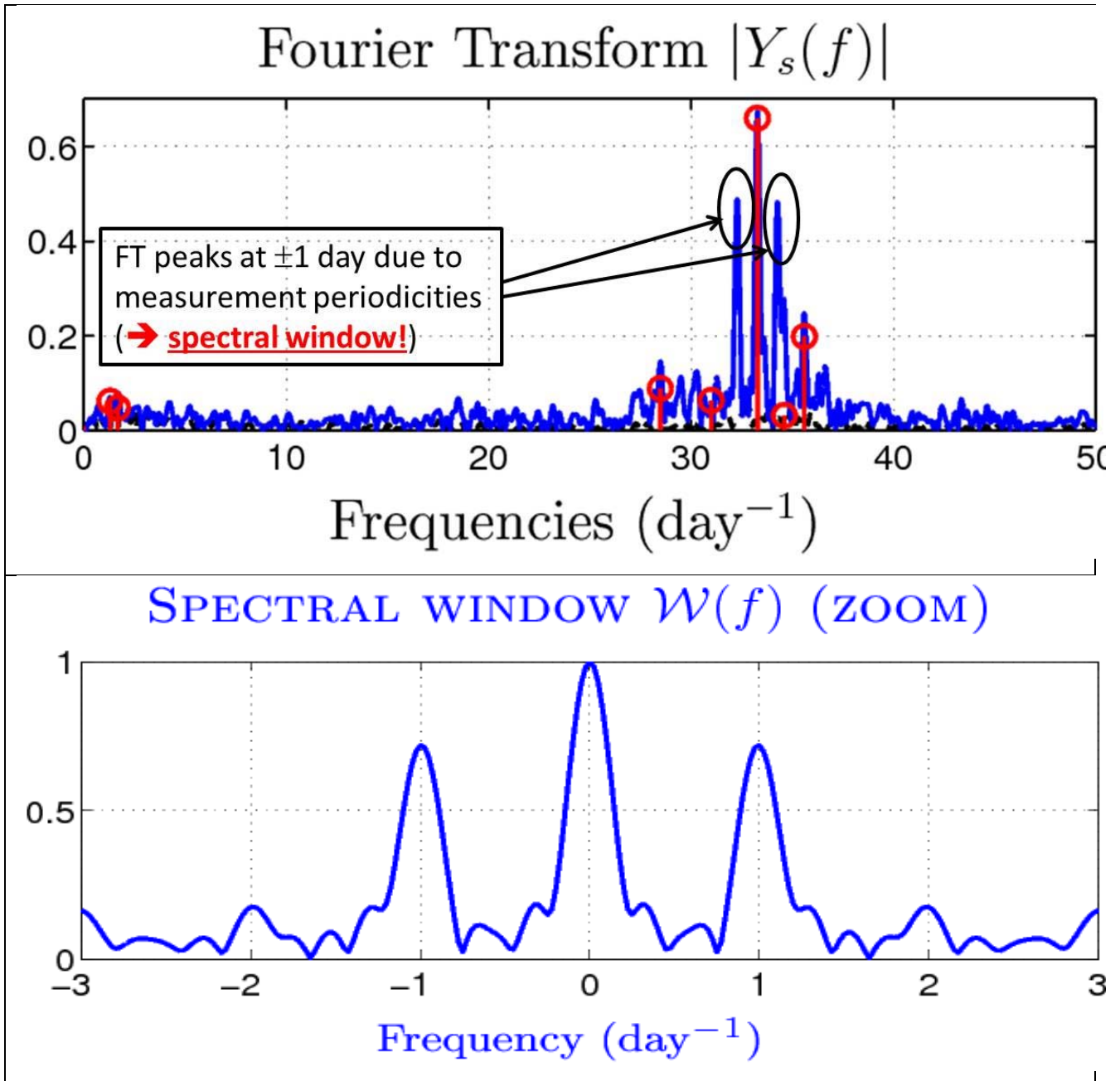


Figure2 (reproduced from [27], the original analysis has been performed in [16, 59, 60]). Top frame: Fourier Transform of the data (blue line) presented in fig1 and *SparSpec* detection results. Various peaks have been detected (indicated by the red vertical lines terminating in a red circle), the lower frequency ones being related to various orbital movements residuals. The black dotted line corresponds to the FT of the estimation residuals. Bottom frame: the (zoomed) spectral window for the measurements presented in fig1: there are very clear ± 1 secondary lobes corresponding to the one-day periodicity in the lack of measurements. The sidebands peaks at ± 1 day are therefore removed from the FT data in the detection results shown in the top frame.

Figure3, D.Testa et al., PaperPPCF2016.

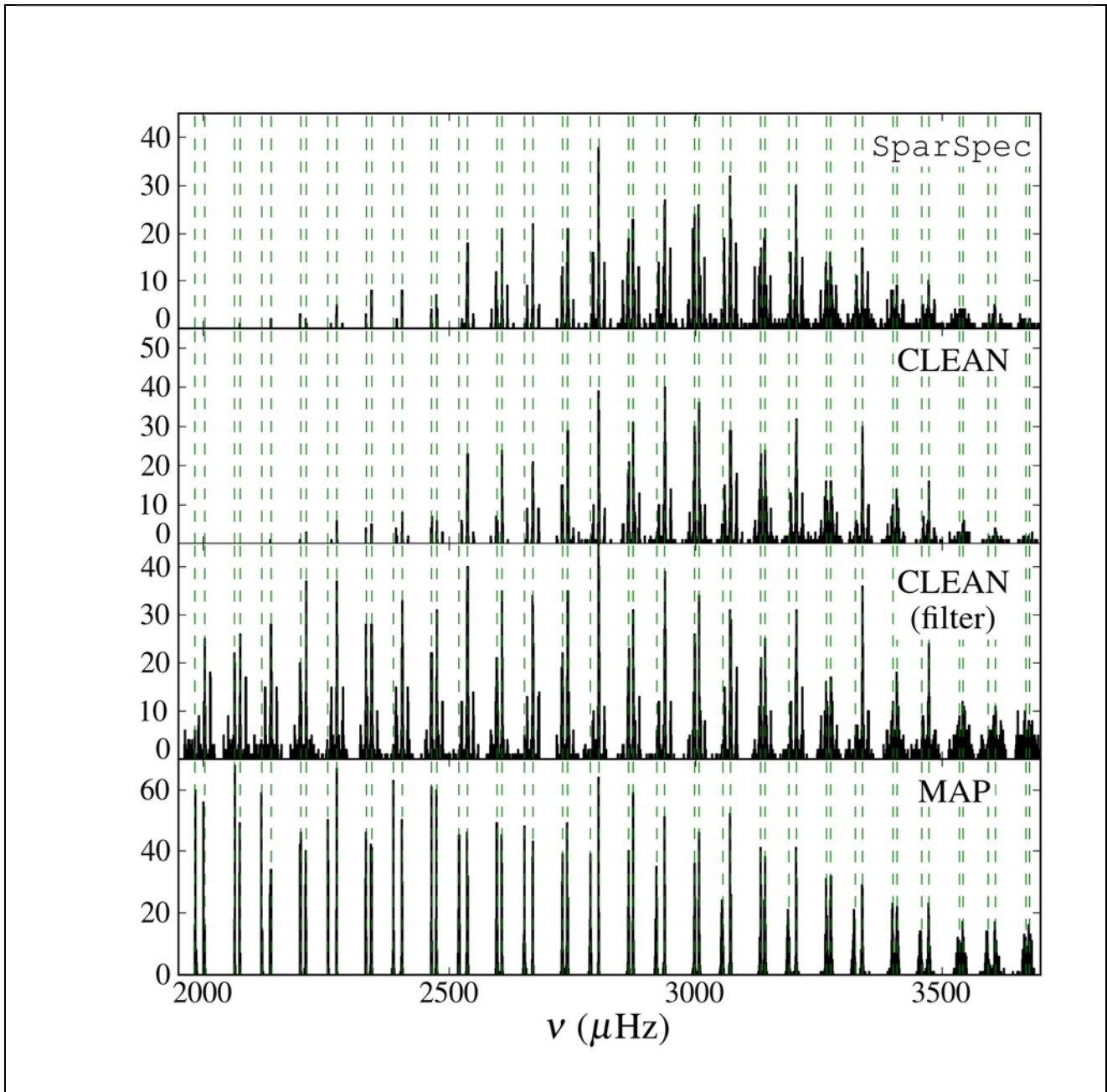


Figure3 (reproduced with the Author's permission from [30]). An histogram showing the results of the *SparSpec* and CLEAN (with and without prior filtering) frequency analyses of 100 irregularly sampled time series constructed so as to match the photometric observations using the input frequencies marked as vertical green lines, in comparison with the results obtained from the MAP method. Note that only in the frequency range between $\sim 2800\mu\text{Hz}$ and $\sim 3400\mu\text{Hz}$ the four methods give similar results, and therefore only 21 of the 52 pulsation modes originally detected with MAP are deemed to be sufficiently reliable to be employed for further astero-seismology analyses.

Figure4, D.Testa et al., PaperPPCF2016.

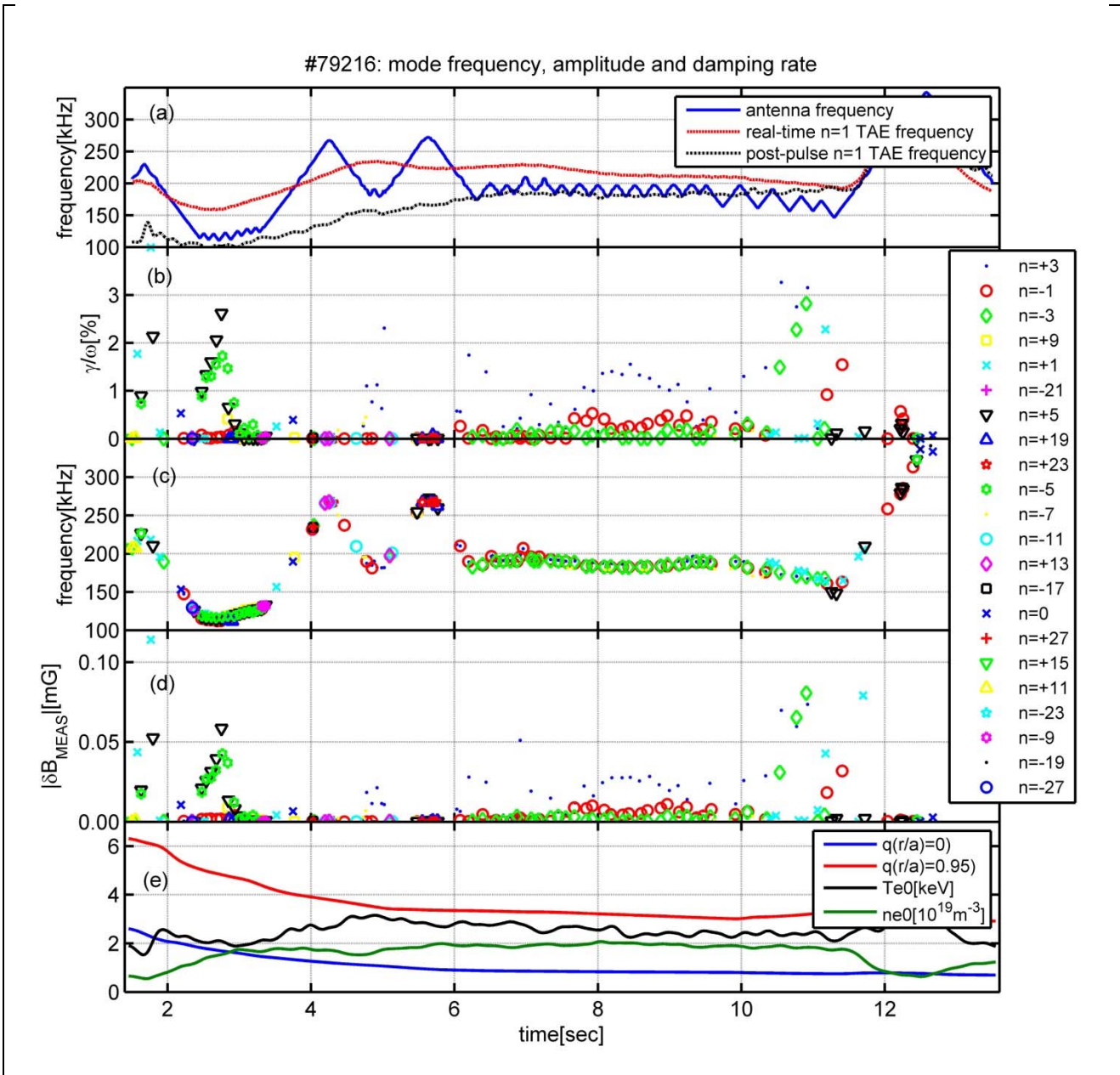


Figure4. Measurement of the mode damping rate (frame-b), mode frequency (frame-c), and mode amplitude (frame-d) for TAEs with different toroidal mode numbers for the He4 discharge #79216. These results were obtained using the post-pulse implementations of the *SparSpec* algorithm. The main plasma parameters are shown in frame-e: $q(r/a)$ is the value of the q -profile at different normalized radial positions r/a (where a is the plasma minor radius), T_{e0} and n_{e0} are the electron temperature and electron density on the magnetic axis (at $r/a=0$). Frame-a show the antenna driving frequency, and the value of the central frequency of the $n=1$ TAE gap as computed using the real-time and the post-pulse data.

Figure5, D.Testa et al., PaperPPCF2016.

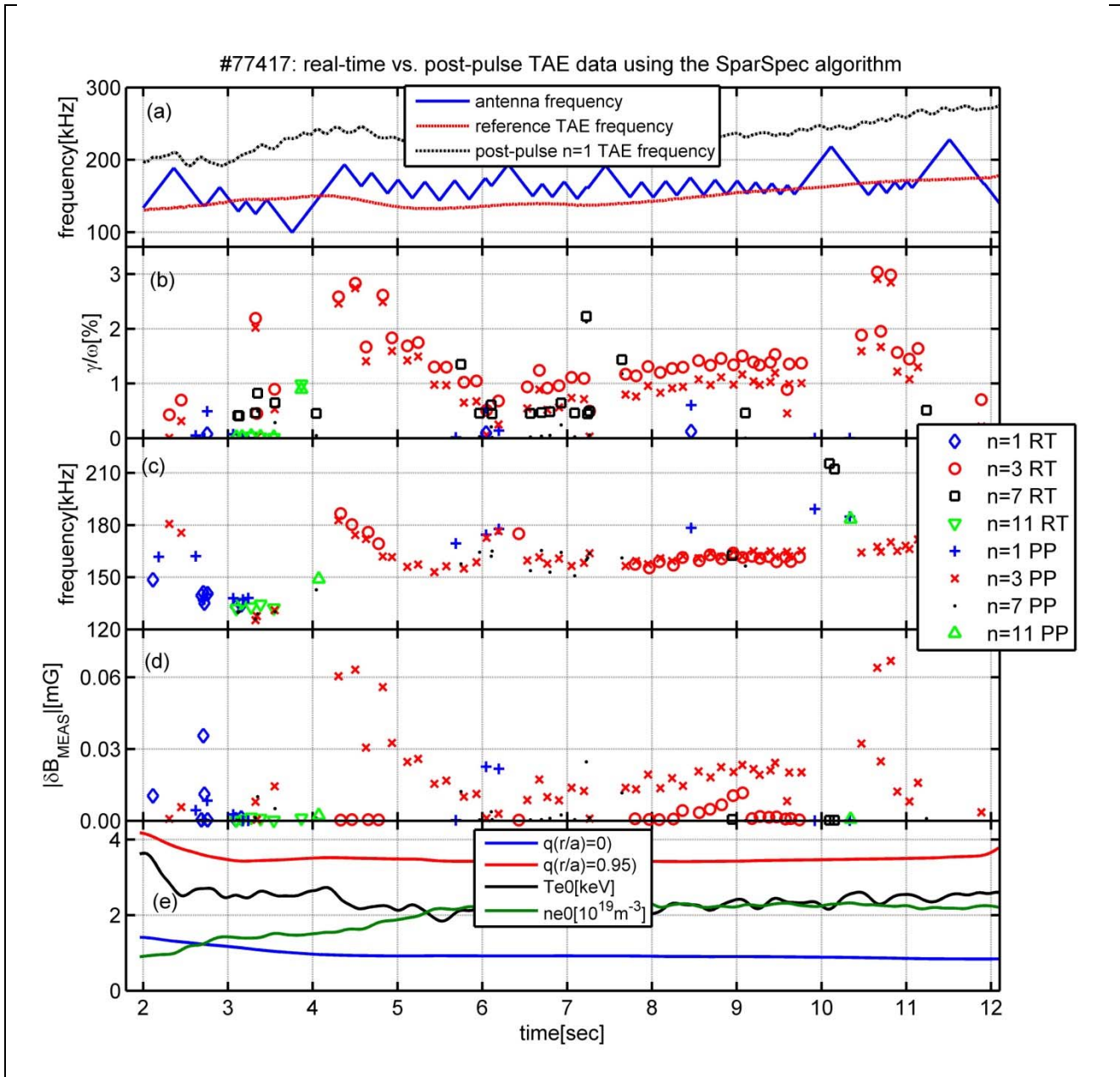


Figure5. Comparison between the real-time and post-pulse measurement of the mode damping rate (frame-b), mode frequency (frame-c), and mode amplitude (frame-d) for TAEs with different toroidal mode numbers for the D2 discharge #77417. The post-pulse analysis was performed using $\lambda_{NORM}=0.65$ and $|f_{MAX}|=150$, whereas for the real-time analysis we set $\lambda_{NORM}=0.85$ and $|f_{MAX}|=20$ to satisfy the CPU and RAM limits. The antenna driving frequency, the value of the central frequency of the n=1 TAE gap as computed using the real-time and the post-pulse data and the main plasma parameters are shown in frame-a and frame-e, respectively, using the same format of fig4.

Figure6a, D.Testa et al., PaperPPCF2016.

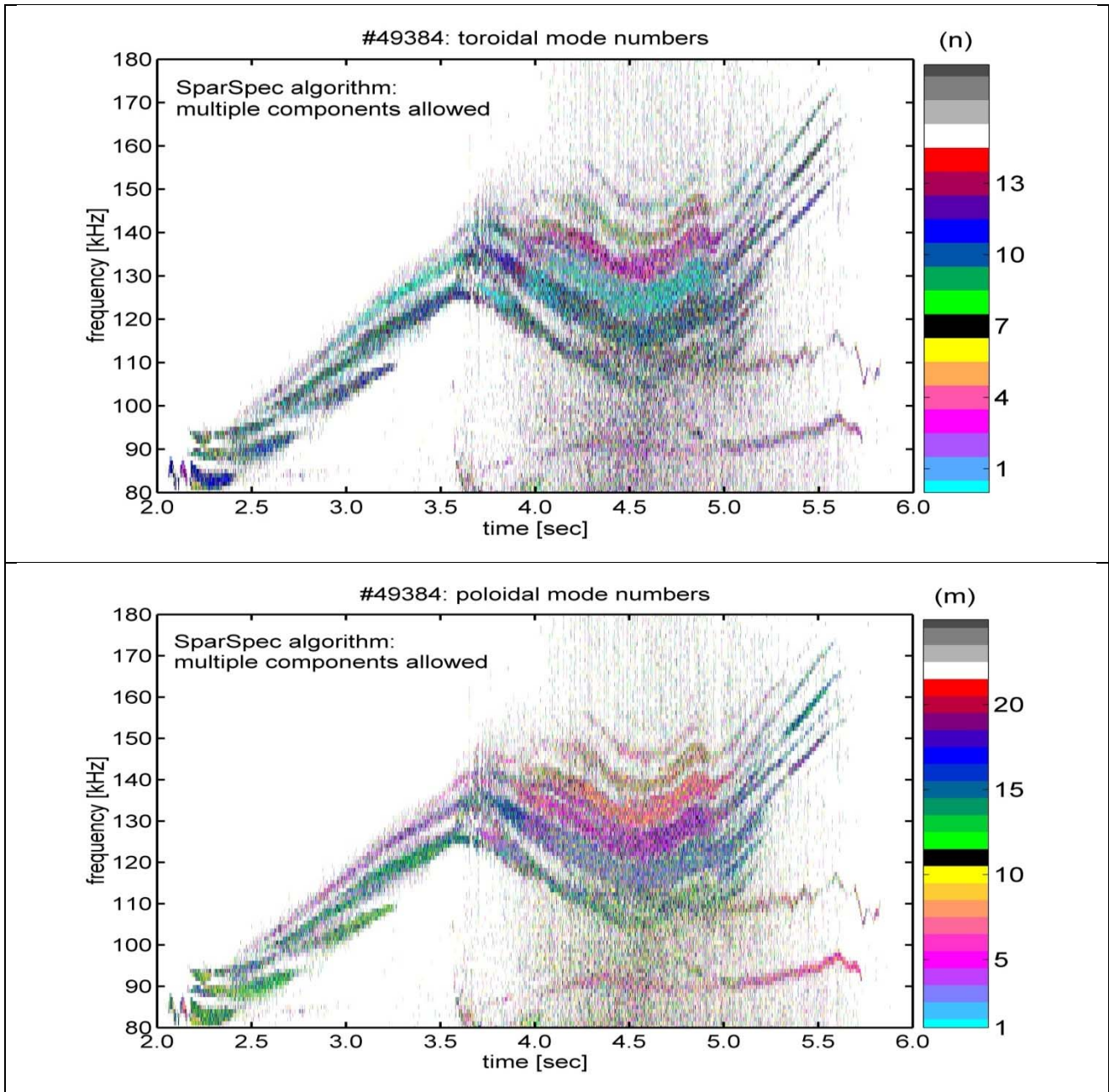


Figure6a. Measurement of the toroidal (top frame) and poloidal (bottom frame) mode numbers for unstable TAEs driven by an energetic ion population. The toroidal mode numbers are measured using magnetic probes located on the low-field-side wall, whereas for the poloidal mode number analysis we combine low- and high-field side probes and use the θ^* correction [58] to the probes' position to map the probe geometrical location on the vessel wall onto the corresponding one on the (not-purely circular) resonant flux surface given by r_{RES} . Here the analysis is performed using the post-pulse implementation of the *SparSpec* algorithm, allowing for multiple components at any given time and frequency point, setting $\lambda_{NORM}=0.65$ and $|f_{MAX}|=150$ for the toroidal mode number analysis, and $\lambda_{NORM}=0.35$ and $|f_{MAX}|=300$ for the poloidal mode number analysis, respectively.

Figure6b, D.Testa et al., PaperPPCF2016.

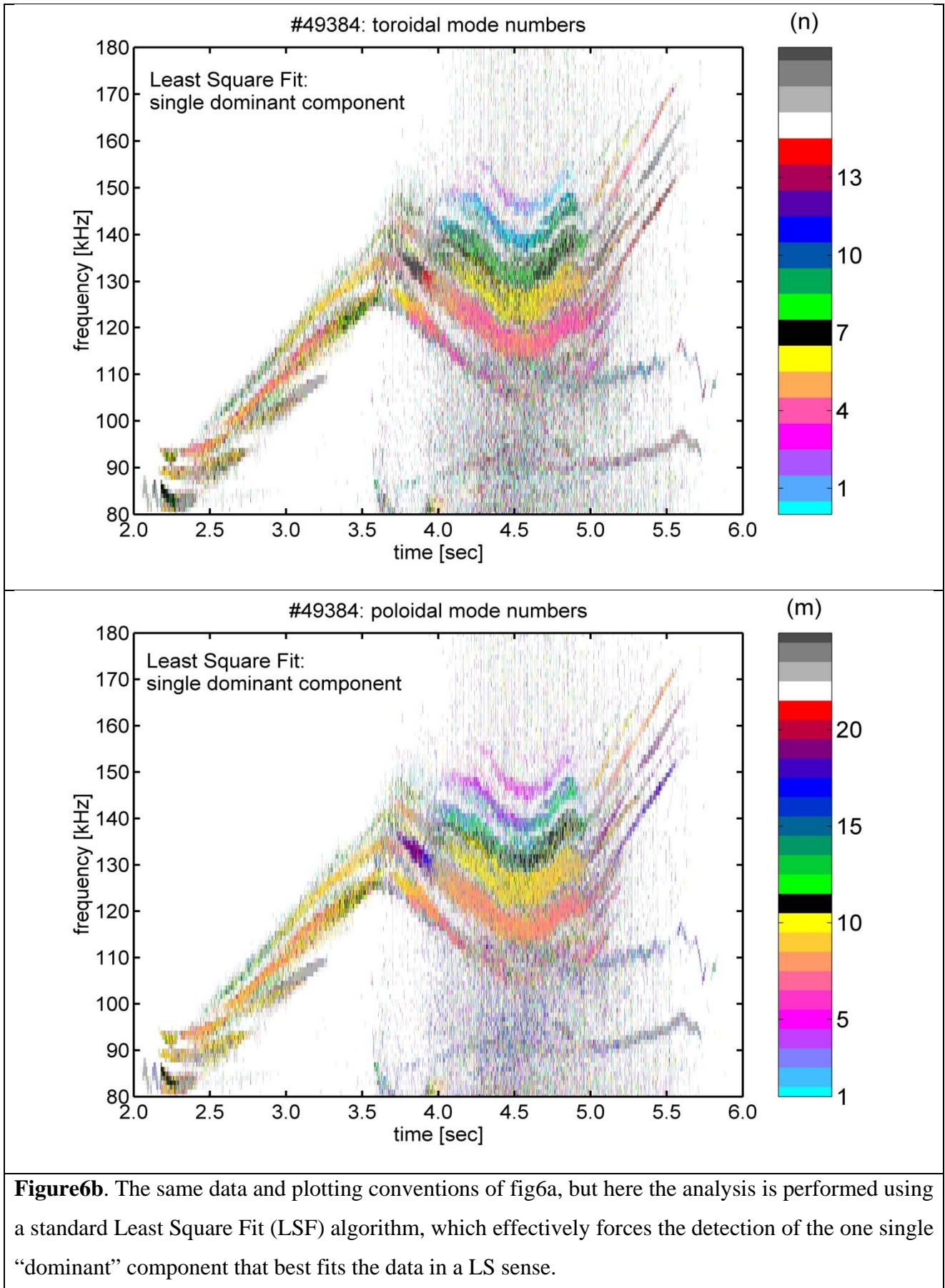


Figure6b. The same data and plotting conventions of fig6a, but here the analysis is performed using a standard Least Square Fit (LSF) algorithm, which effectively forces the detection of the one single “dominant” component that best fits the data in a LS sense.

Figure6c, D.Testa et al., PaperPPCF2016.

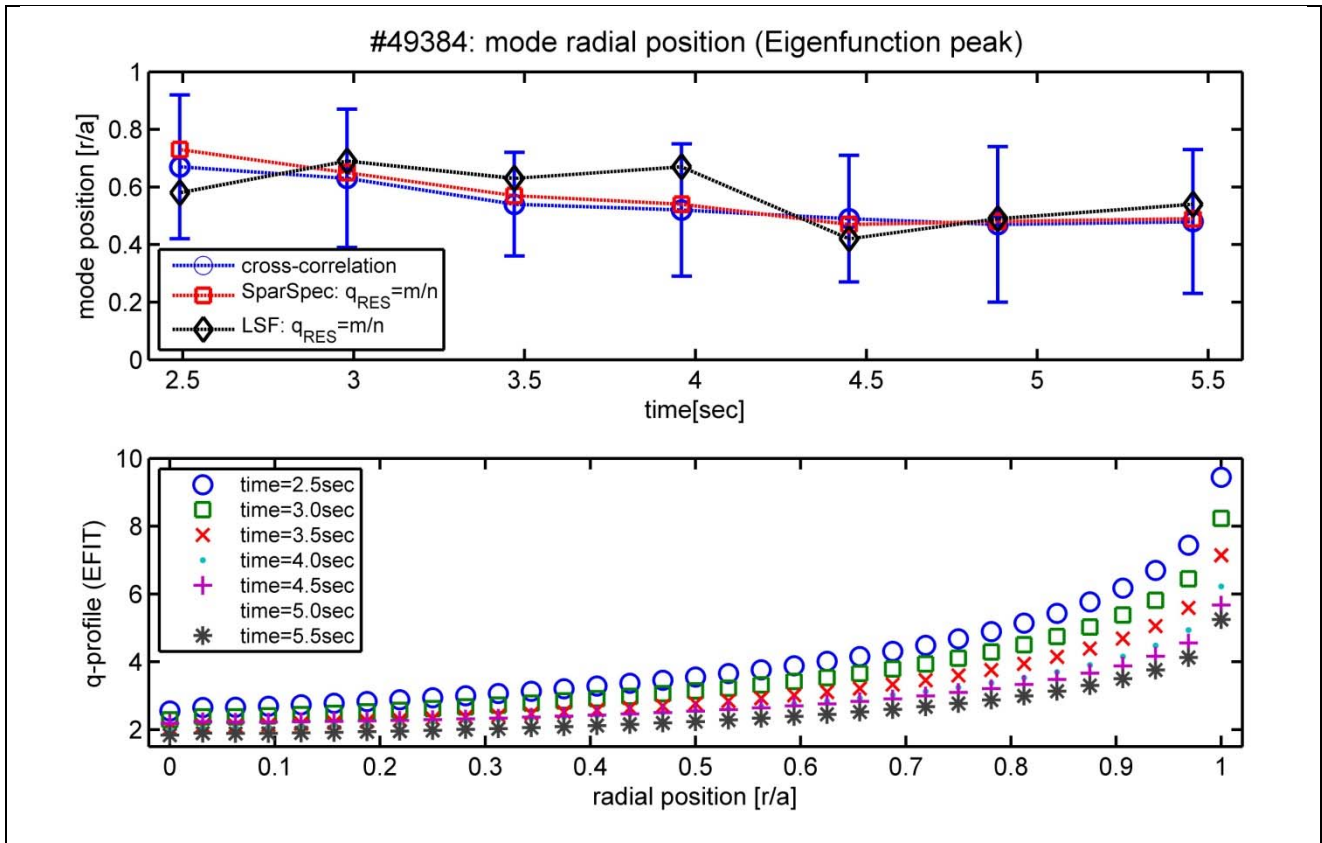


Figure6c. Comparison of the mode radial position, defined as the peak in the radial Eigenfunction, obtained through cross-correlation analysis and using $q_{RES}(t_{RES})=(2m+1)/2n$, with the toroidal and poloidal mode numbers provided by the *SparSpec* and LSF algorithms. The vertical “error bar” for the cross-correlation data indicates the estimated width of the radial mode Eigenfunction. We also show the EFIT reconstruction of the q-profile. The three estimates of r_{RES} are broadly consistent with each other, but we note that as the current profile relaxes and the q-value drops across the entire plasma cross-section, the cross-correlation analysis shows that the mode radial location progressively and smoothly shifts inwards towards the magnetic axis, and this trend is only captured correctly when using the *SparSpec* data to obtain r_{RES} .

Figure7, D.Testa et al., PaperPPCF2016.

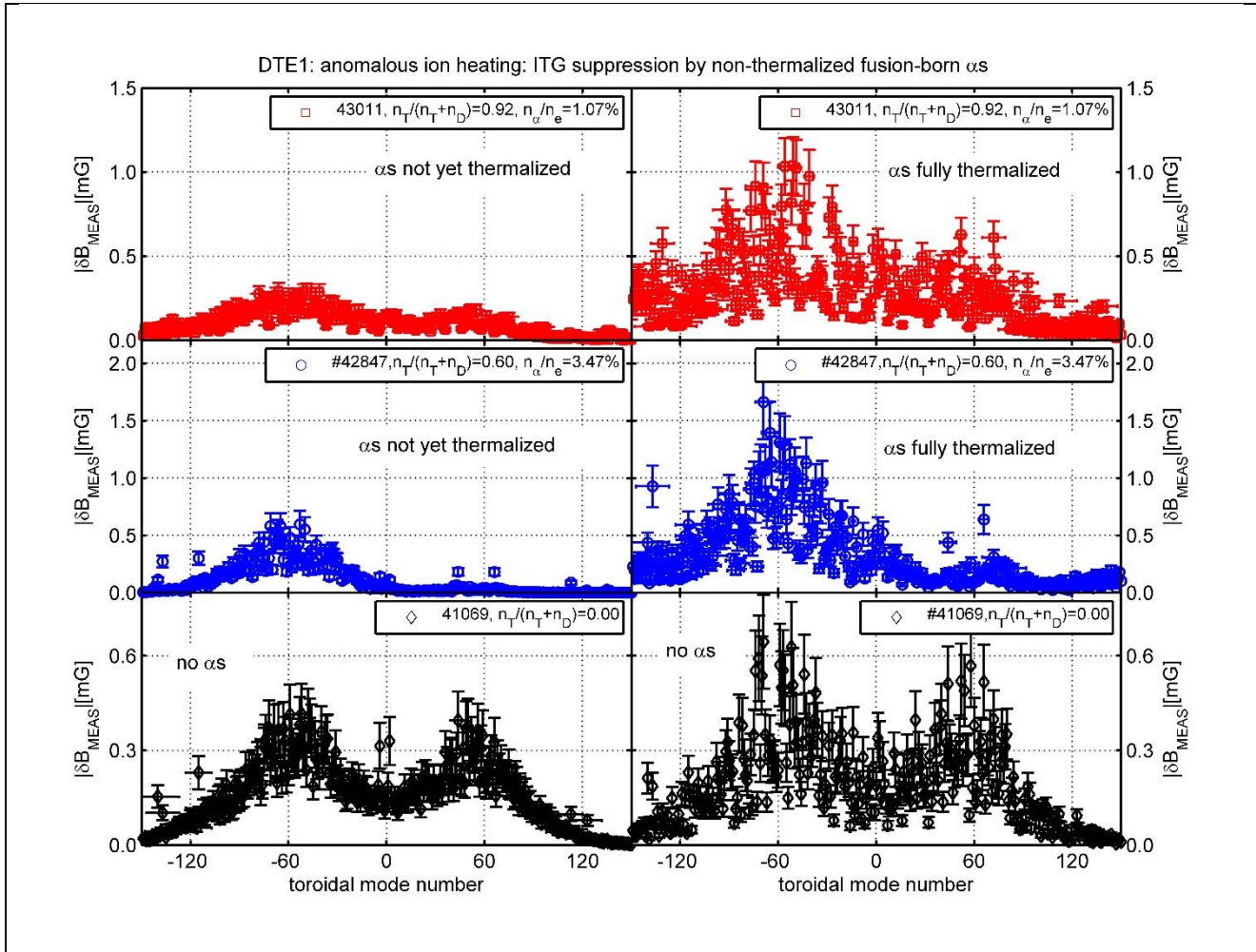


Figure7. Summary of the turbulence spectral data for the alpha heating experiment during the JET DTE1 campaign. The data are shown as function of the toroidal mode number for three discharges with different densities of α particles, and separately for two phases in the discharge: before (time=13.15sec) and after (time=14.20sec) full thermalization of the α s. For the discharge #41069 (bottom frame), which does not have α s, the data have been analysed at the same two time points used for the discharges #42847 and #43011, which had α s. These results were obtained using the post-pulse implementation of the *SparSpec* algorithm, with $\lambda_{\text{NORM}}=0.15$ and $|f_{\text{MAX}}|=500$.

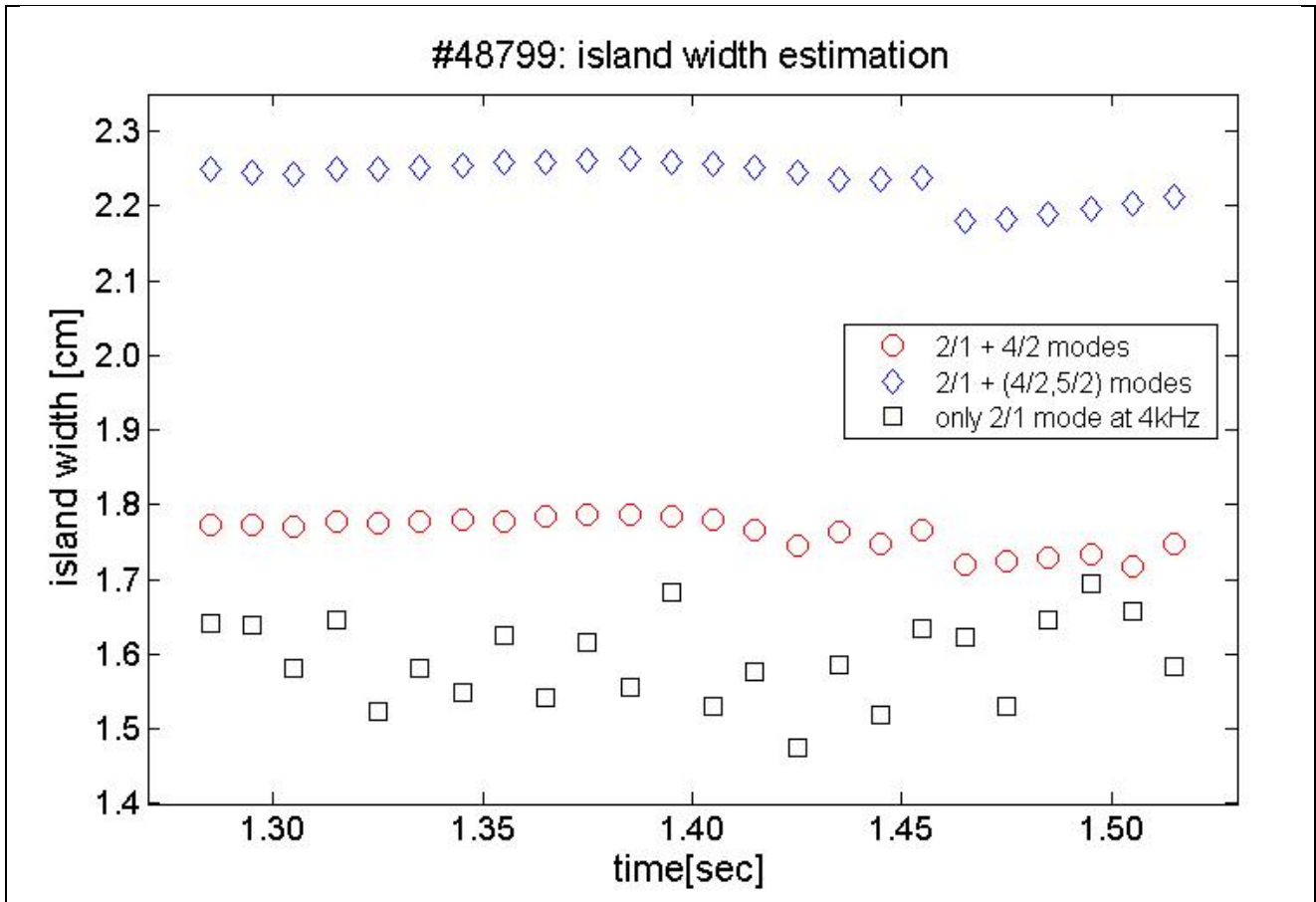
Figure8, D.Testa et al., PaperPPCF2016.

Figure8. Three different estimation of the island width for the TCV discharge #48799 using the magnetic measurements made with Mirnov coils mounted on the vacuum vessel wall. The first estimate (symbol = (blue) squares) is obtained when considering only a pure $m/n=2/1$ mode at around 4kHz, the second (symbol = (black) diamonds) summing up this mode and a pure $m/n=4/2$ mode at around 8kHz, and the third (symbol = (red) circles) including the contribution of the sideband $m/n=5/2$ component at 8kHz obtained through the *SparSpec* analysis. We note that by considering only the $m/n=2/1$ mode at 4kHz the estimate of the island width oscillate significantly over the time window of interest, from ~ 1.47 cm to ~ 1.70 cm. By also considering the contribution of the same helicity $m/n=4/2$ mode at 8kHz, not only the island width estimate increases to ~ 1.8 cm, but it is also much more stable over time. Finally, when also including the sideband mode at 8kHz with different helicity $m/n=5/2$, we obtain a larger estimate for the island width, increasing from ~ 1.8 cm to ~ 2.2 cm over the time window of interest.

Figure9a, D.Testa et al., PaperPPCF2016.

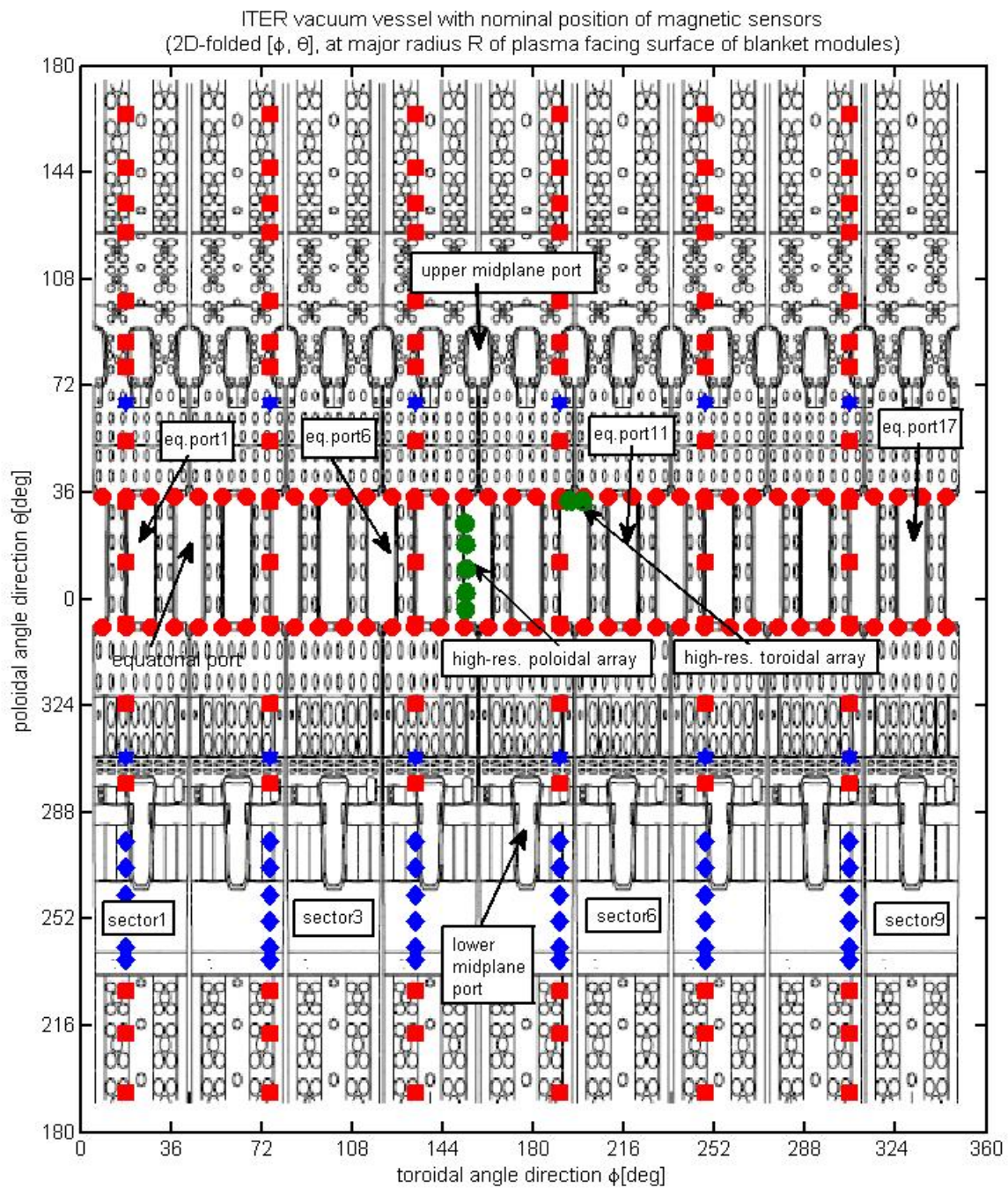


Figure9a. The proposed layout for the ITER HF magnetic diagnostic system as in 2009, for toroidal (filled red circles) and poloidal (filled red square) mode number; the filled blue lozenges indicate the HF magnetic sensors in the divertor region, blacked-out in this analysis as the presence of the divertor and shaping coils induces parasitic and eddy currents in the surrounding metal structures that prevent reliable measurements of HF instabilities using magnetic pick-up sensors; the filled blue stars indicate the sensors that have been removed because their position clashes with that of the newly designed active coils for ELM control.

Figure9b, D.Testa et al., PaperPPCF2016.

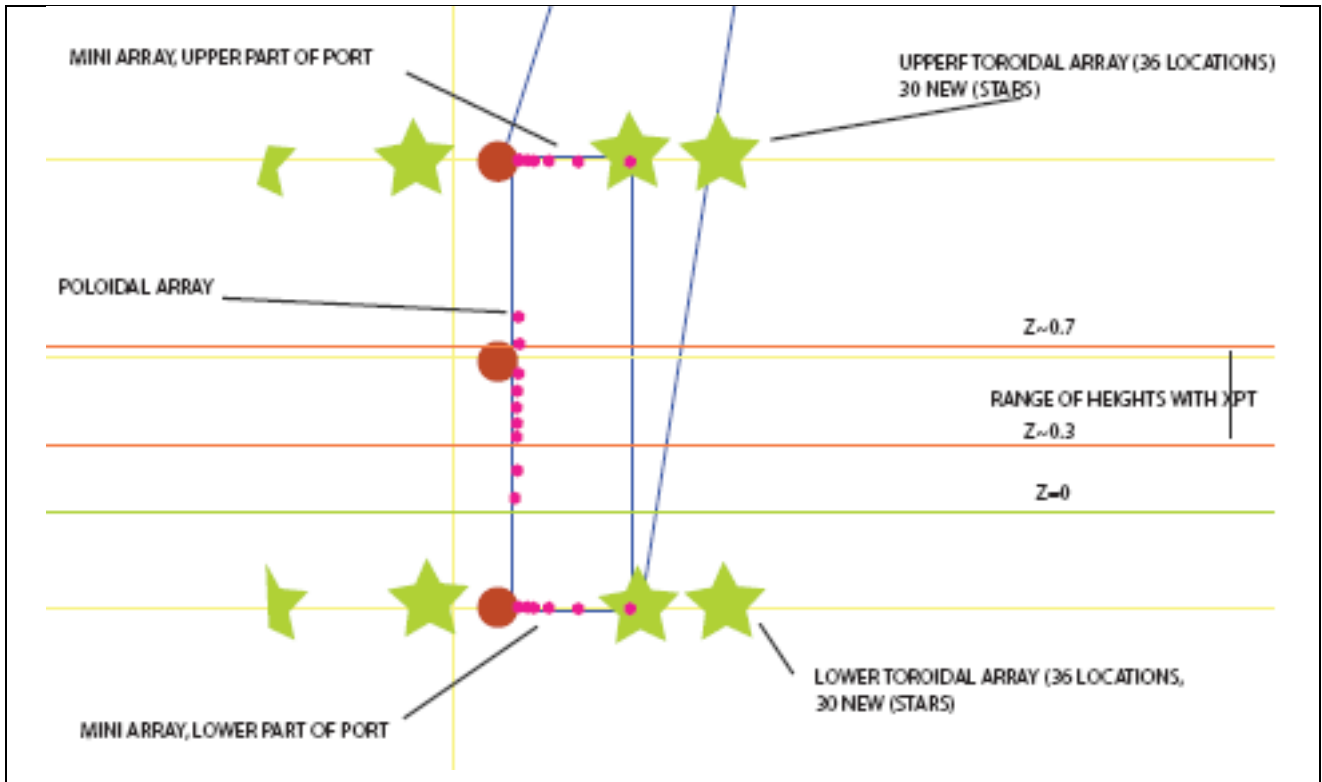


Figure9b. The proposed implementation for the high-resolution toroidal and poloidal arrays (indicated by the small red dots) to be located on the horizontal and vertical edges of some of the equatorial ports. The poloidal mode number measurement arrays, indicated by the big red dots, are located in the sectors [#1, #3, #4, #6, #7, #9], whereas the toroidal mode number measurement arrays, indicated by the big green stars, are only located on the low-field side at the two Z-heights of the horizontal edges of the equatorial ports.

Figure10a, D.Testa et al., PaperPPCF2016.

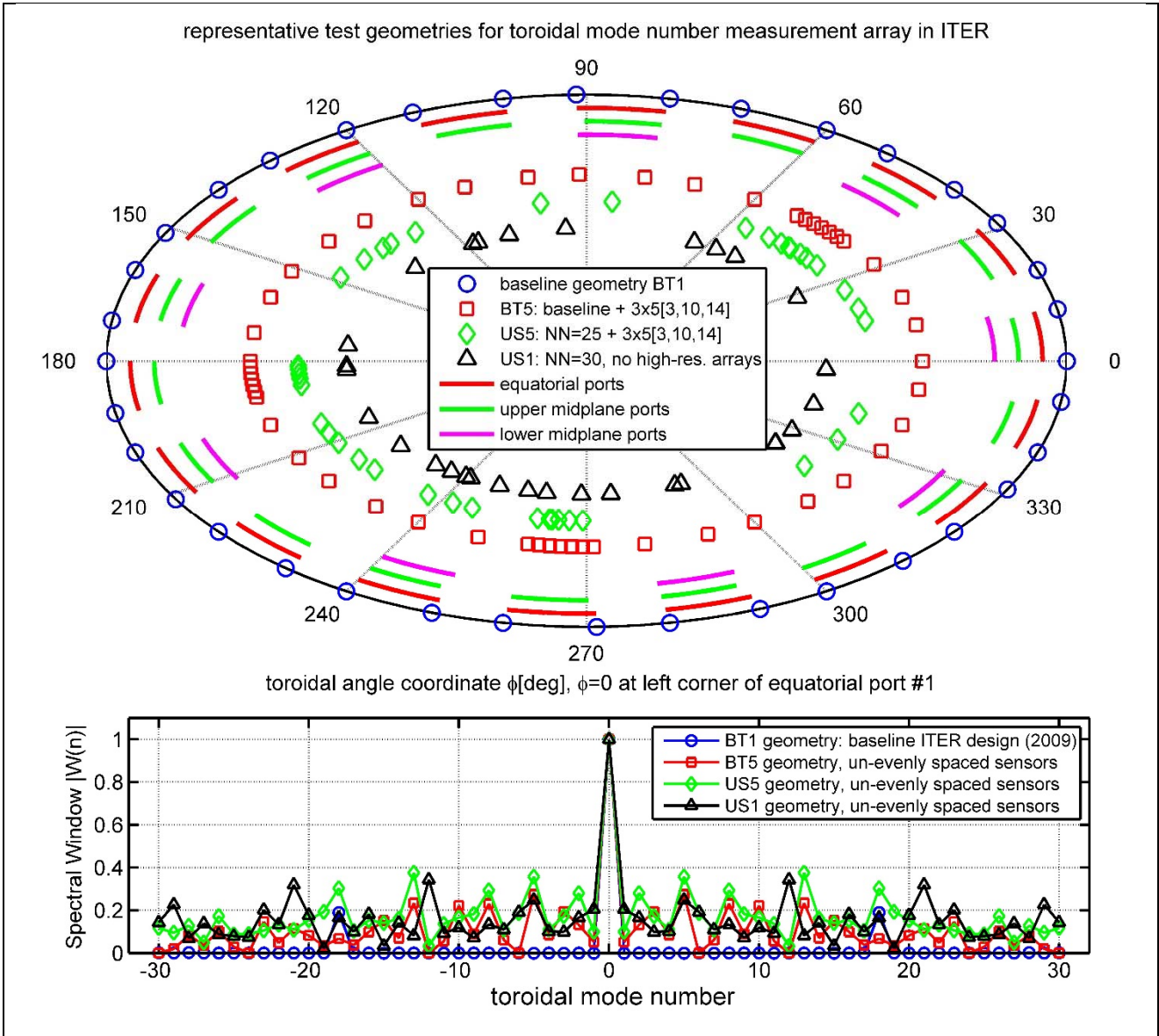


Figure10a. Some representative examples of the test geometries used for the performance analysis for the ITER toroidal mode number measurement array. The baseline geometry BT1 represents the nominal ITER design as in 2009, with $NN=2 \times 18$ equi-spaced sensors. The BT5 geometry adds to this baseline design 3 arrays of 5 high-resolution sensors each in the equatorial ports [#3, #10, #14]. The US5 geometry replaces the $NN=2 \times 18$ equi-spaced (baseline) sensors of the BT5 geometry with $NN=25$ un-evenly spaced sensors. Finally, the US1 geometry uses $NN=30$ un-evenly spaced sensors without high-resolution arrays. Note that only the BT1 geometry with equi-spaced sensors has $W(v)=0$ exactly for all mode numbers except $n=0$ and the pseudo-Nyquist value $|n|=18$.

Figure10b, D.Testa et al., PaperPPCF2016.

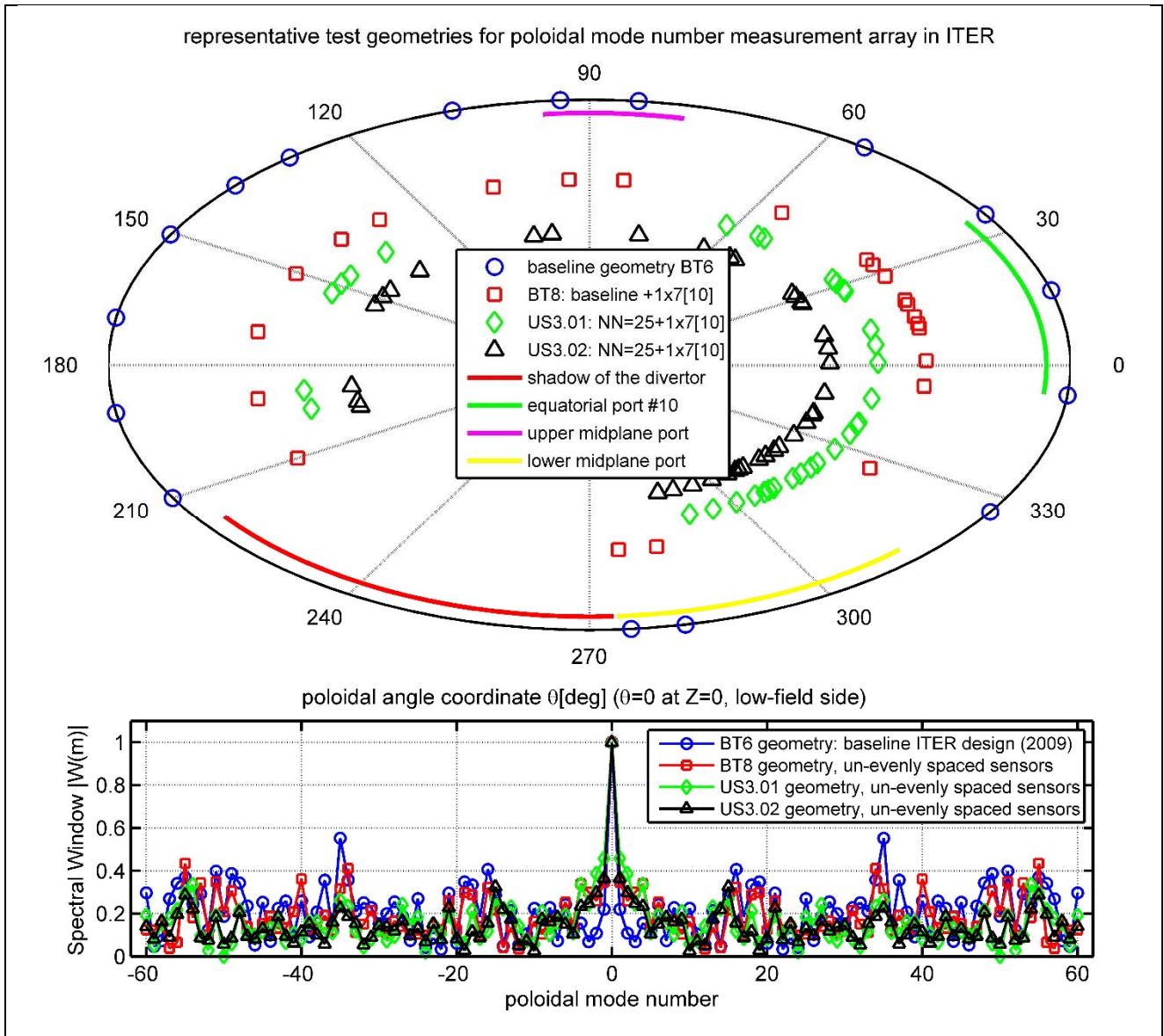


Figure10b. Some representative examples of the test geometries used for the performance analysis for the ITER poloidal mode number measurement array. The baseline geometry BT6 represents the nominal ITER design as in 2009, with NN=16 un-evenly spaced sensors. The BT8 geometry adds to this baseline design one array of 7 high-resolution sensors in the equatorial port [#10]. The US3.01 and US3.02 geometries replace the NN=16 baseline sensors of the BT6 geometry with NN=25 un-evenly spaced sensors, but located at different test positions. As the sensors are not equi-spaced, no geometry has $W(v)=0$ exactly for any mode numbers.

Figure11, D.Testa et al., PaperPPCF2016.

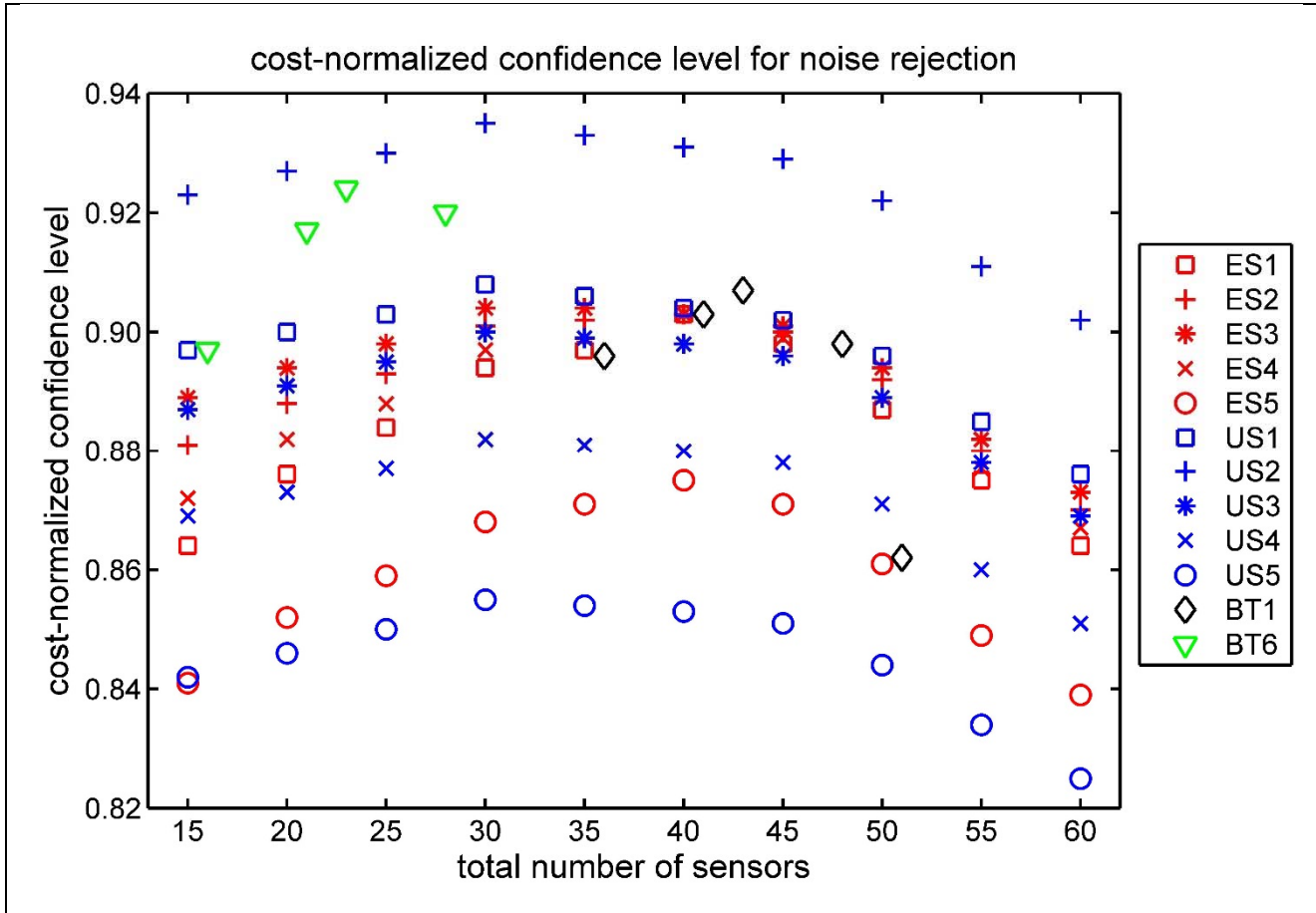


Figure11. Cost-normalized confidence level for noise rejection for evenly (ES#) and un-evenly (US#) spaced sensors, comparing the data with the baseline toroidal (BT1) and poloidal (BT6) assemblies. The data are plotted as a function of the total number of sensors. We note a characteristic bell shape in the cost-normalized confidence level, which is obtained because above a certain number of sensors, the cost increases much more rapidly than the confidence level for noise rejection, which remains almost constant. For this analysis we used $\lambda_{\text{NORM}}=0.05$ (i.e. so as to detect “all” modes) and set f_{MAX} to be five times larger than the maximum physical mode number present in the input spectrum, i.e. $|f_{\text{MAX}}|=150$ and $|f_{\text{MAX}}|=300$ for toroidal and poloidal mode number analysis, respectively.

Figure12, D.Testa et al., PaperPPCF2016.

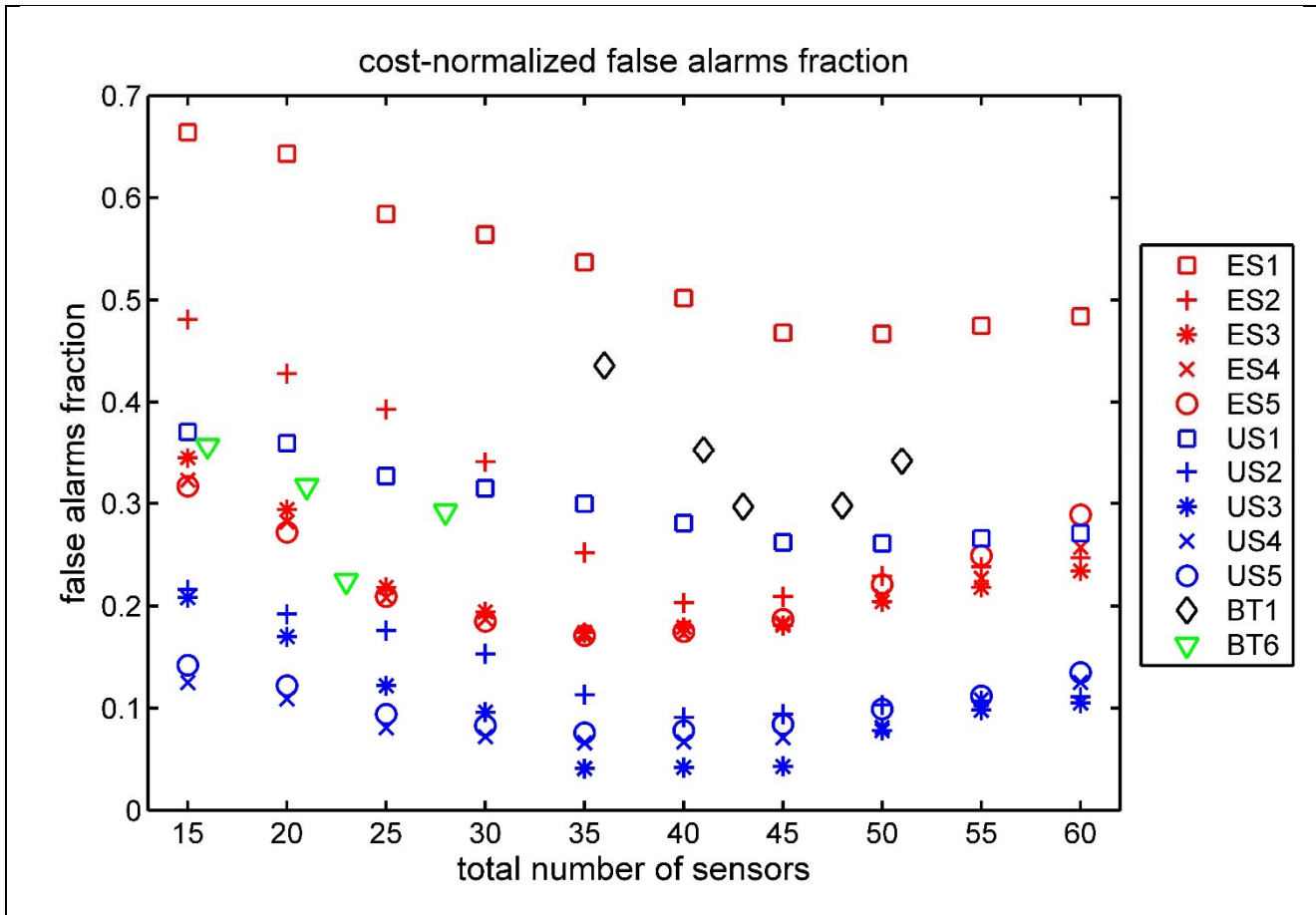


Figure12. Cost-normalized false alarm fraction for evenly and un-evenly spaced geometries, and also comparing the results with those for the baseline toroidal and poloidal assemblies. The labelling format for the different geometries is the same as in fig11. We note again the characteristic (in this case inverted up-down) bell shape in the cost-normalized fraction of false alarms, which is obtained because above a certain number of sensors, the cost increases much more rapidly than the false alarm rejection, which remains almost constant. For this analysis we used $\lambda_{\text{NORM}}=0.85$ and set f_{MAX} to be five times larger than the maximum physical mode number present in the input spectrum, i.e. $|f_{\text{MAX}}|=150$ and $|f_{\text{MAX}}|=300$ for toroidal and poloidal mode number analysis, respectively.

Figure13, D.Testa et al., PaperPPCF2016.

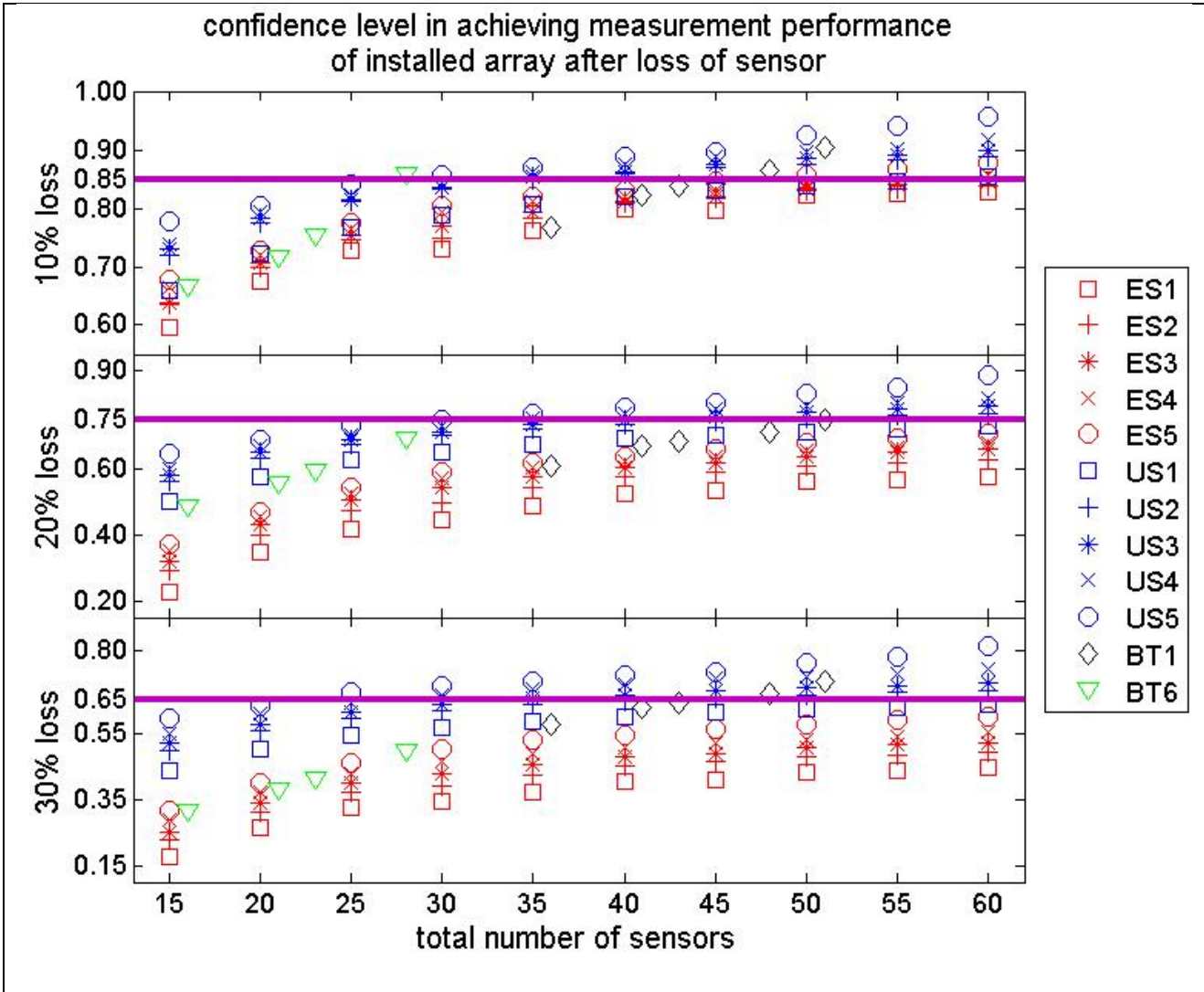


Figure13. The confidence level in achieving the same measurement performance of the installed assembly after a [10%, 20%, 30%] loss of sensors. The labelling format for the different geometries is the same as in fig11. As in fig12, for this test we used $\lambda_{\text{NORM}}=0.85$ and $|f_{\text{MAX}}|=150$ and $|f_{\text{MAX}}|=300$ for toroidal and poloidal mode number analysis, respectively. The threshold values $=0.85$, $=0.75$ and $=0.65$ associated to a nominal [10%, 20%, 30%] loss of sensors, respectively, for acceptance of this test are explicitly shown by the horizontal (magenta) lines to guide the eye. For toroidal mode number analysis, the nominal ITER geometry satisfies the requirements for the resilience in the measurement performance when adding at least 1x12 high-resolution sensors, for a total of 48 sensors. Conversely, an assembly with 25 un-evenly spaced sensors in total, comprising 3x5 high-resolution arrays, satisfies these requirements. For poloidal mode number analysis, only an assembly with 30 to 35 un-evenly spaced sensors in total, comprising one array of 12 high-resolution sensors, satisfies the requirements for resilience in the measurement performance.

Figure14, D.Testa et al., PaperPPCF2016.

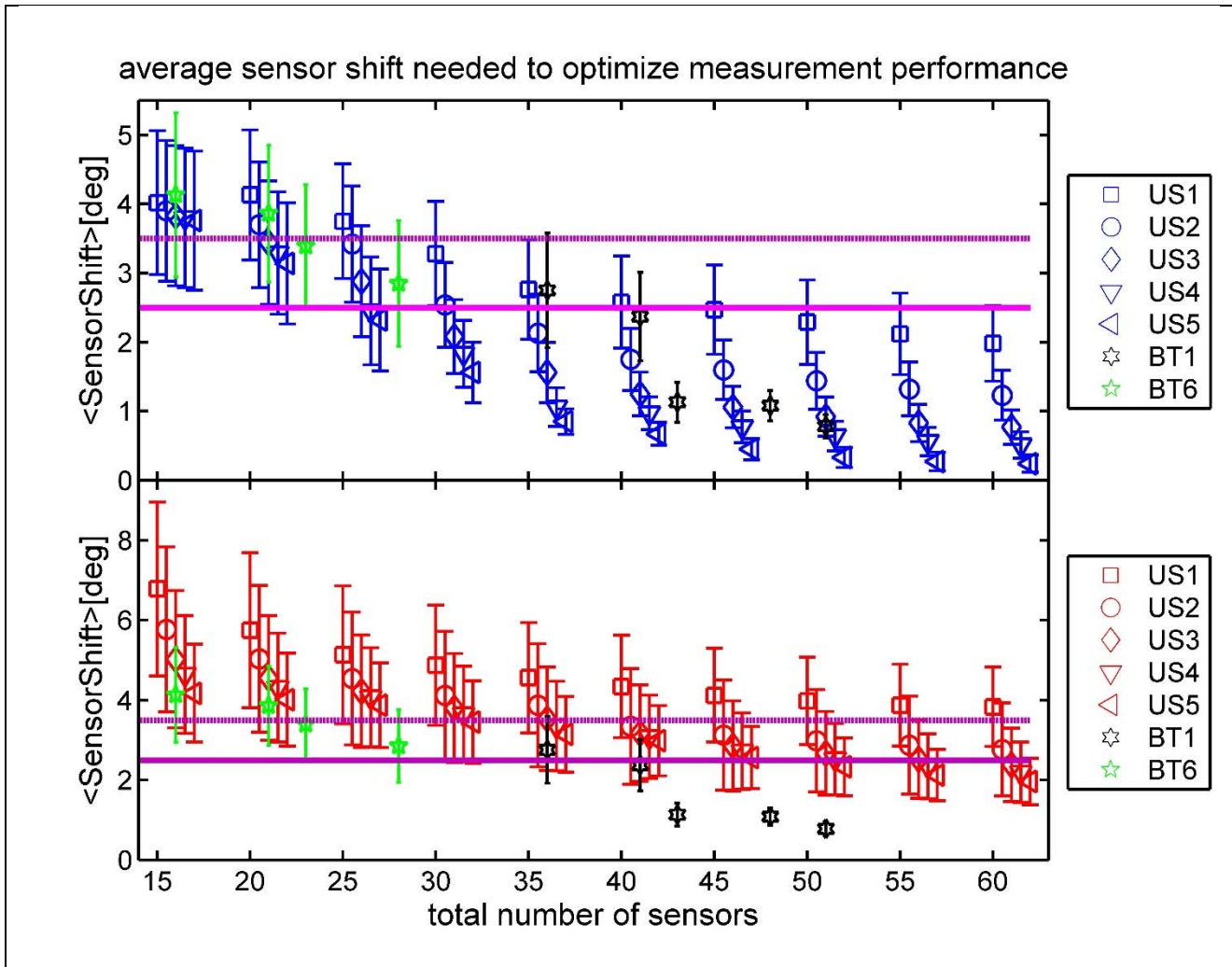


Figure14. The average positional shift required to optimize the measurement performance given an initial sensors' arrangement. The horizontal axis is slightly displaced for the different geometries to improve the clarity of the graphical presentation, and the vertical error-bar indicates the standard deviation on the average value of the required sensor shift. The labelling format for the different geometries is the same as in fig11. As in fig12, for this test we used $\lambda_{\text{NORM}}=0.85$ and $|f_{\text{MAX}}|=150$ and $|f_{\text{MAX}}|=300$ for toroidal and poloidal mode number analysis, respectively. The threshold values $=2.5\text{deg}$ and $=3.5\text{deg}$ for acceptance of this test are explicitly shown by the horizontal (magenta) lines to guide the eye. We find that the nominal ITER geometry for toroidal mode number analysis satisfies the optimization requirements only when adding at least 1x7 high-resolution sensors, i.e. for a total of 43 sensors, whereas an assembly with 25 un-evenly spaced sensors in total, comprising 3x5 high-resolution arrays, satisfies these requirements. The nominal ITER geometry for poloidal mode number analysis does not satisfy the optimization requirements, whereas an assembly with 25 un-evenly spaced sensors in total, comprising 12 high-resolution sensors, satisfies them.

Figure15a, D.Testa et al., PaperPPCF2016.

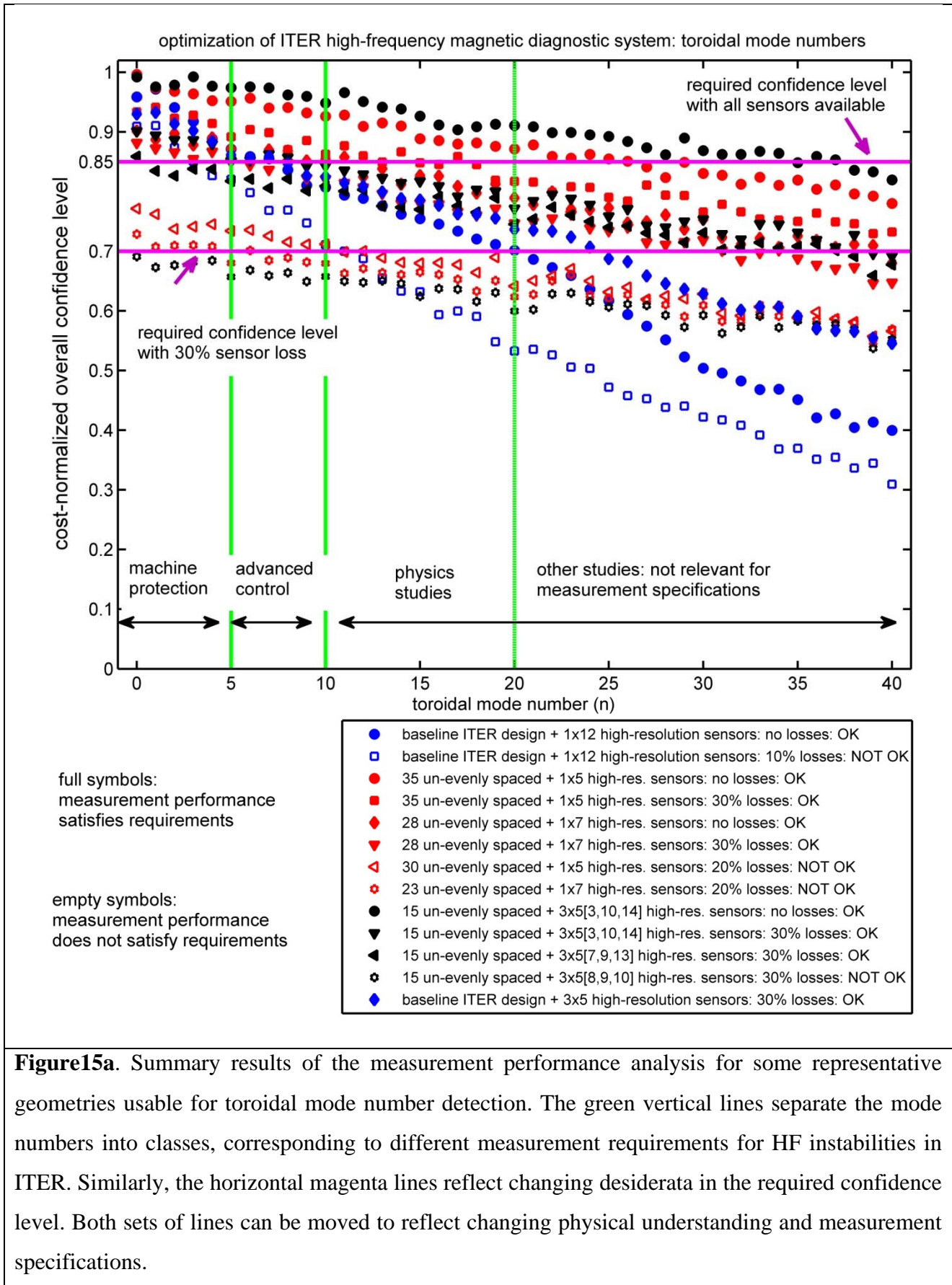


Figure15b, D.Testa et al., PaperPPCF2016.

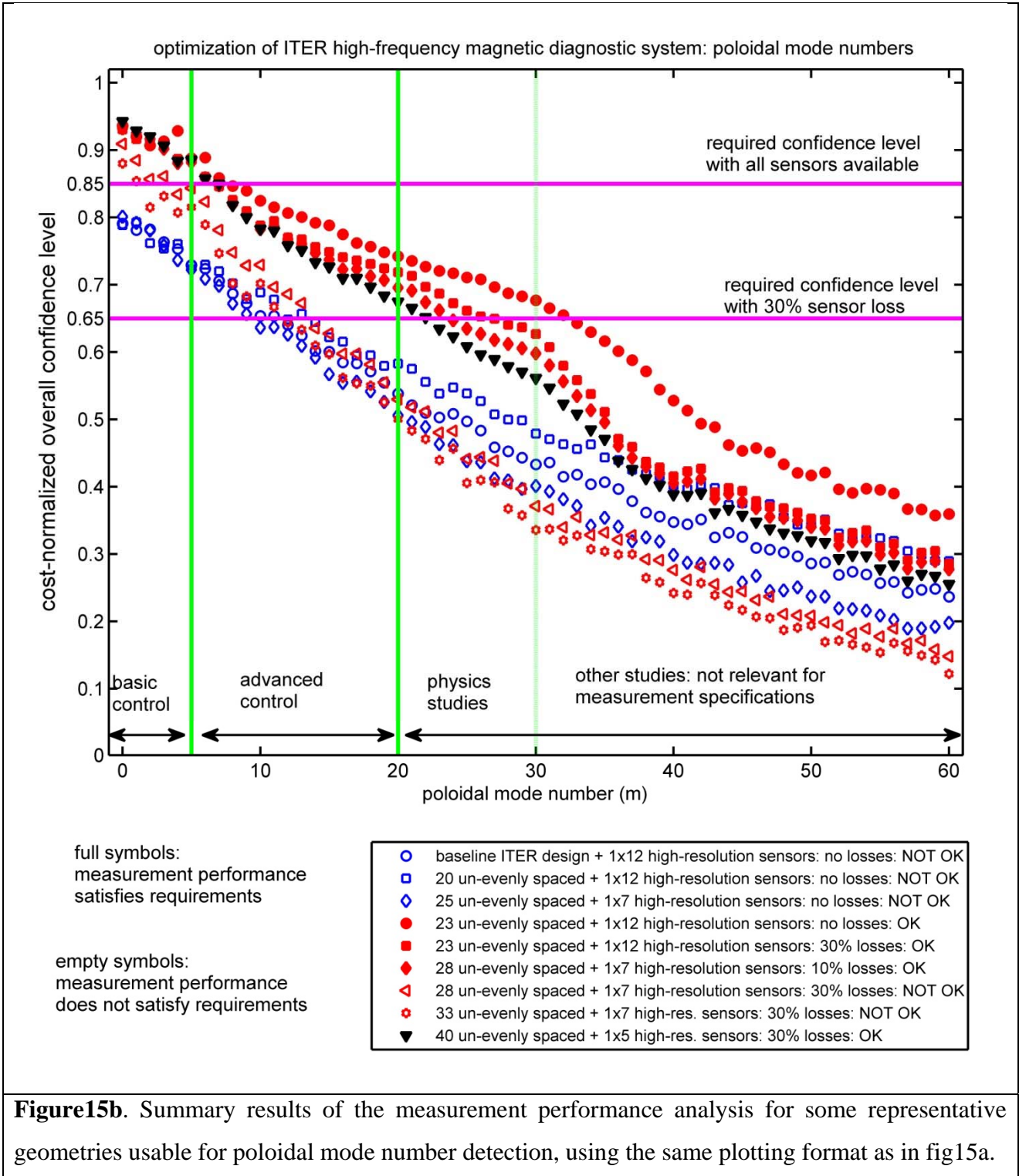


Figure16, D.Testa et al., PaperPPCF2016.

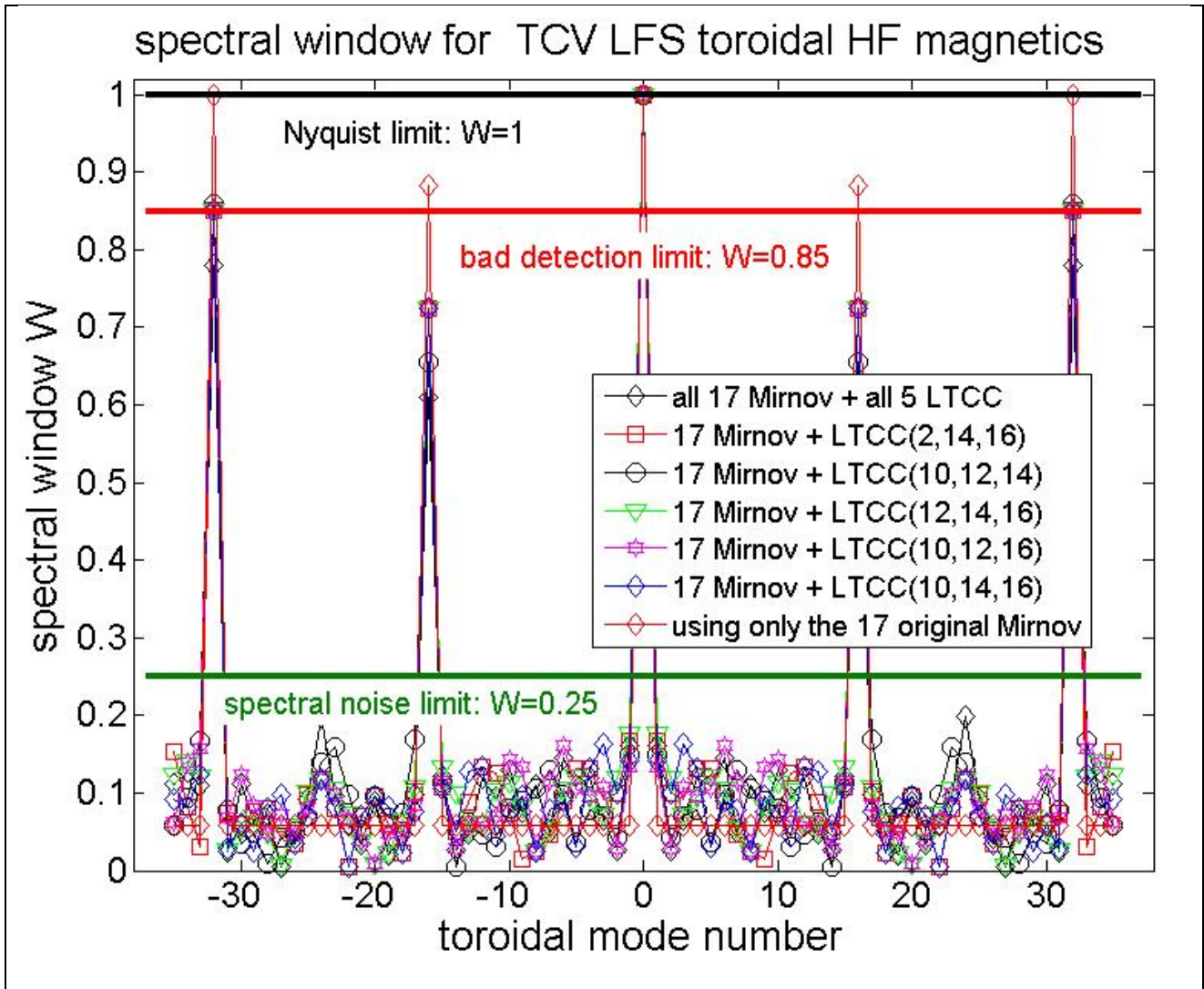


Figure16. Expected measurement performance for the HF magnetic diagnostic system on TCV, toroidal LFS array. With the addition of only three out of five LTCC-3D sensors, the best measurement performance is obtained by installing them in the sectors [#2, #14, #16], as we have $W(|n|=16)=0.72$ and $W(|n|=32)\sim 0.84$, marginally below the bad detection limit. All other selection of three sectors out of five give rise to $W(|n|=32)>0.87$, marginally above the bad detection limit. For comparison purposes, the measurement performance of the TCV toroidal-LFS array with the addition of the five new LTCC-3D sensors has $W(|n|=16)=0.61$ and $W(|n|=32)\sim 0.79$. With only the original 17 Mirnov-1D sensors, $W(|n|=16)\sim 0.88$, already above the bad detection limit.

Figure17, D.Testa et al., PaperPPCF2016.



Figure17. On the left, an example of a ziro with its content: a canopo, the black cinerary urn inside the ziro, and the objects found accompanying the canopo: a bronze daggar, two bronze rings, probably used to close the garments, and a ceramic dish. On the right, a decorated canopo that was found during the excavation of the proto-Villanovian necropolis of Poggio della Pozza.

Figure18, D.Testa et al., PaperPPCF2016.

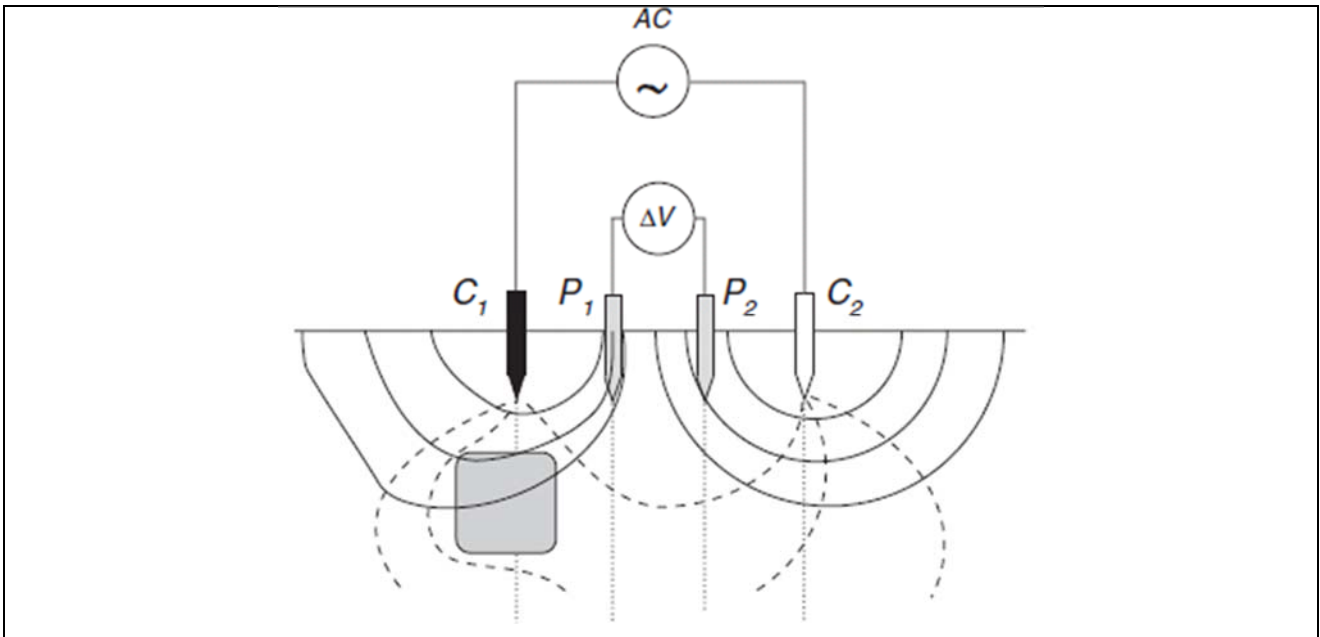


Figure18. A schematic drawing illustrating the principle of electrical prospecting. Assuming an otherwise uniform ground resistance, the effect of a large object buried under the AC current-driving electrode C_1 is shown on the equi-potential contours measured with the potential electrodes P_1 and P_2 . The underground object distorts the current flow patterns, which significantly modifies the potential measurements made with P_1 , but only little those made with P_2 . By moving electrodes, a map of the perturbation can be obtained, which can be modelled as the superposition of point sources, due to the current-carrying electrodes and the localized underground objects.

[image credits [97]: Neil Linford, English Heritage, *Rep. Prog. Phys.* **69** (2006), 2205]

Figure19, D.Testa et al., PaperPPCF2016.

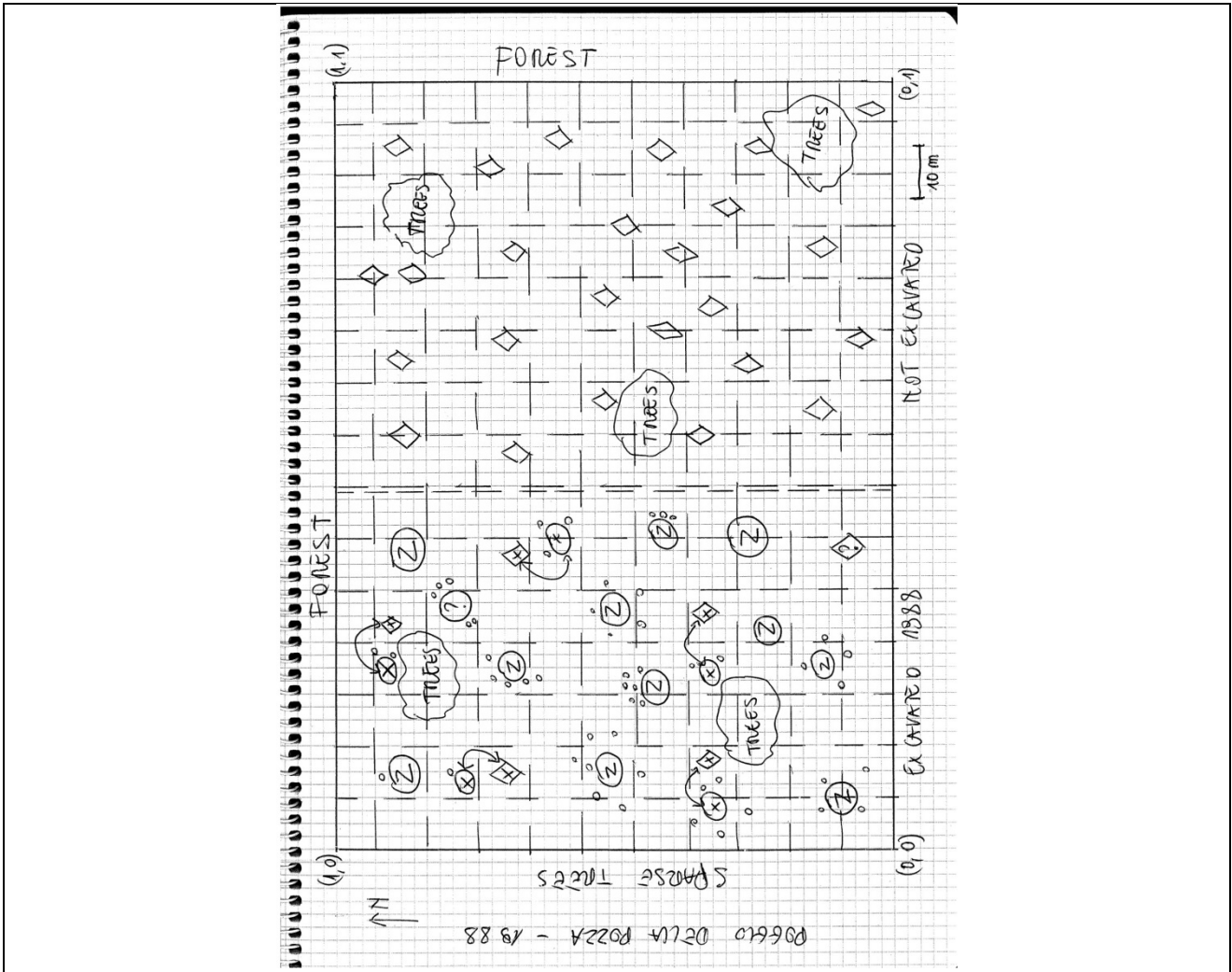
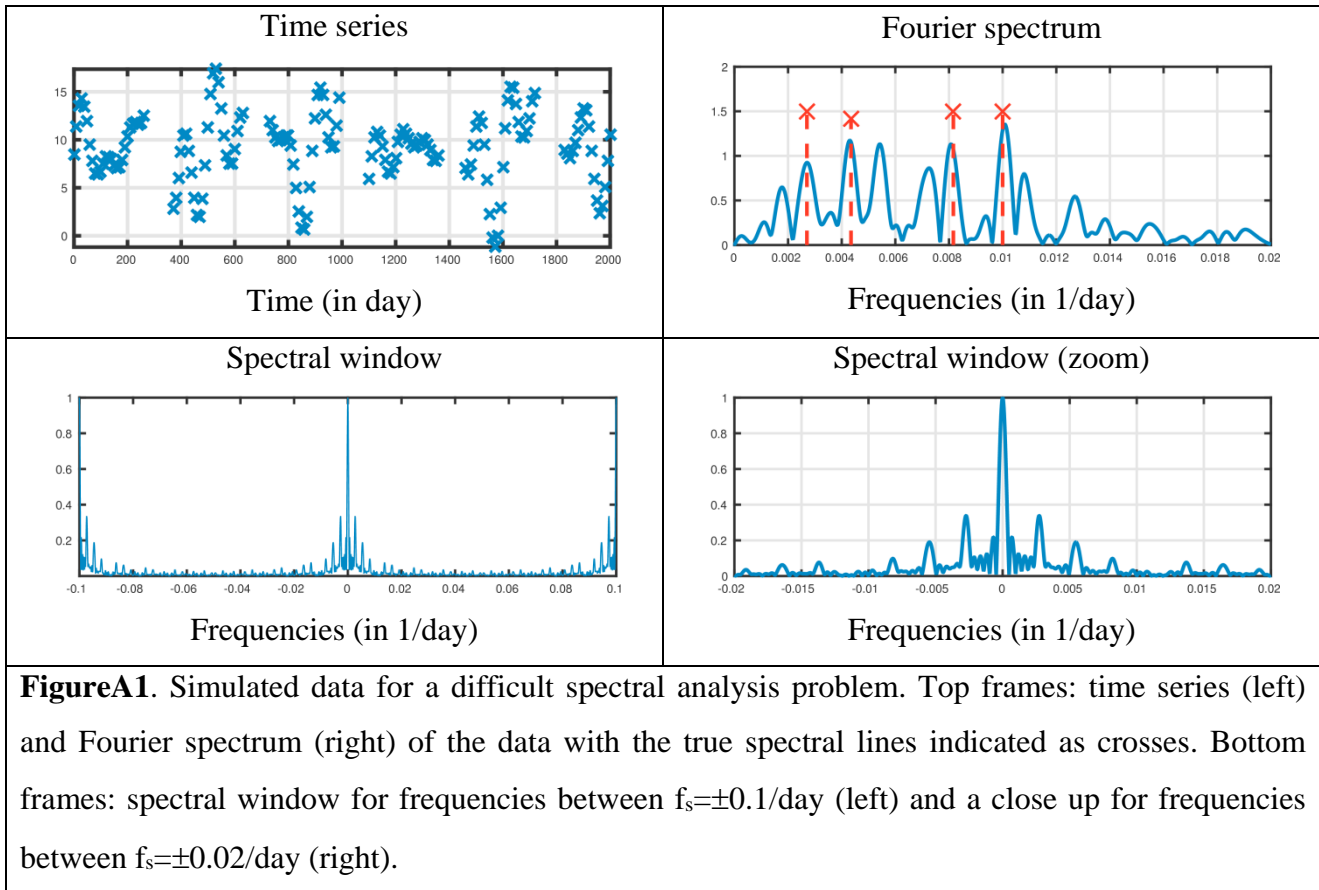
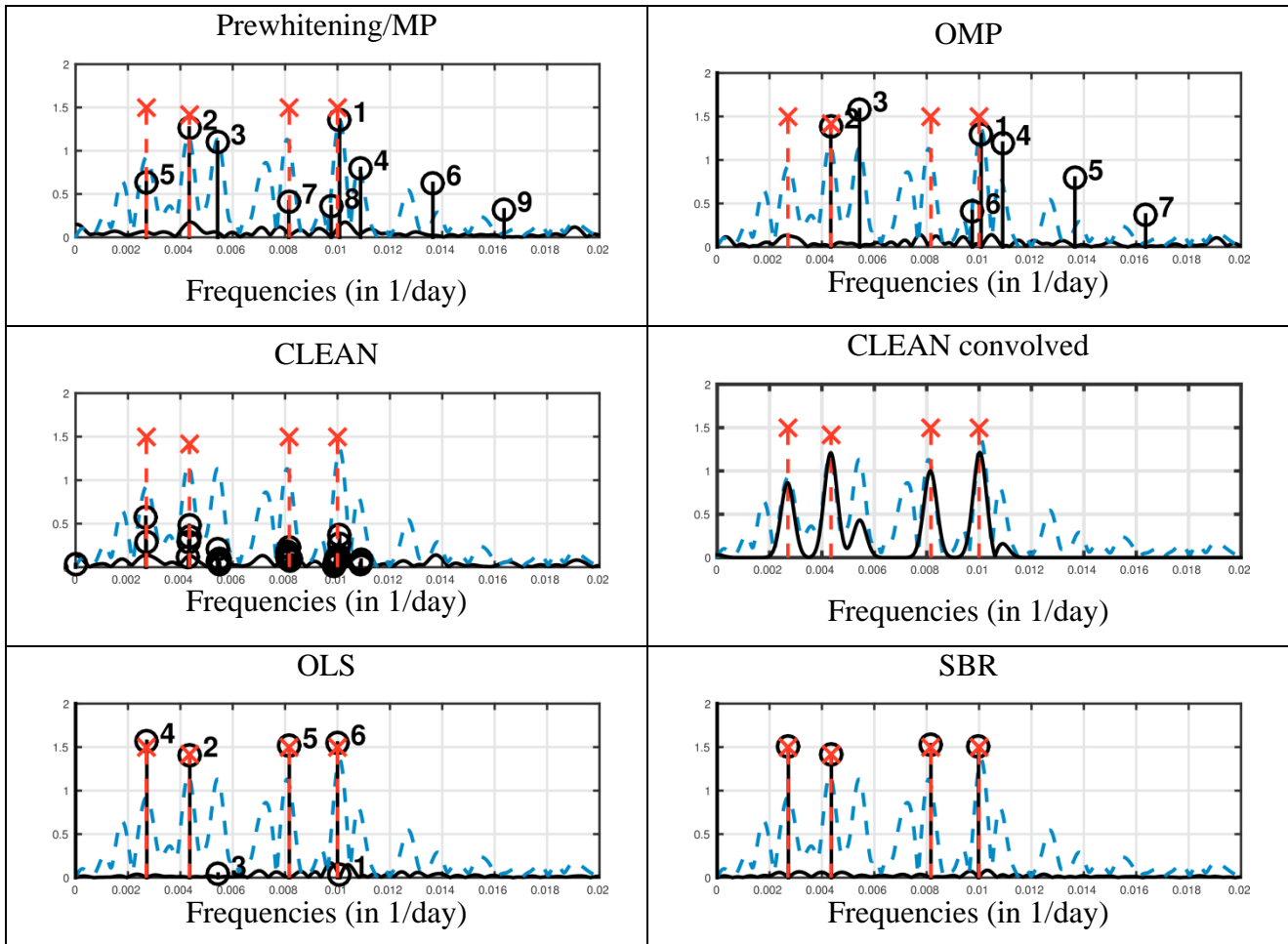


Figure19. A modified copy of original field drawings illustrating the ground situation of the proto-Villanovian necropolis of Poggio della Pozza as of the 1988 excavations. The site is limited by a dense forest on three sides and by sparse trees leading to an abrupt downhill slope on the fourth side. Only a fraction of the site was excavated up to 1988 and in this area a number of ziri were found, indicated by larger and smaller surrounding circles. The lozenges indicate the position of a ziro as determined from the analysis of the electrical survey performed in 1988. In the excavated area, when the estimated position of a ziro matches within 3m the position of one actually found, the circle is inscribed with the letter “Z”, and no lozenge is drawn, otherwise the circles and the lozenges are inscribed with an “X” if the two are not too far apart (within 10m), or with a question mark if no correspondence is found.

FigureA1, D.Testa et al., PaperPPCF2016.

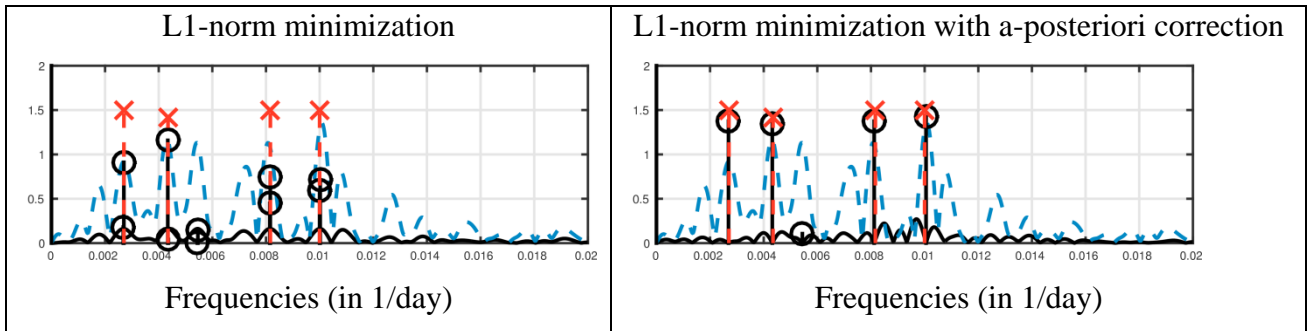


FigureA2, D.Testa et al., PaperPPCF2016.



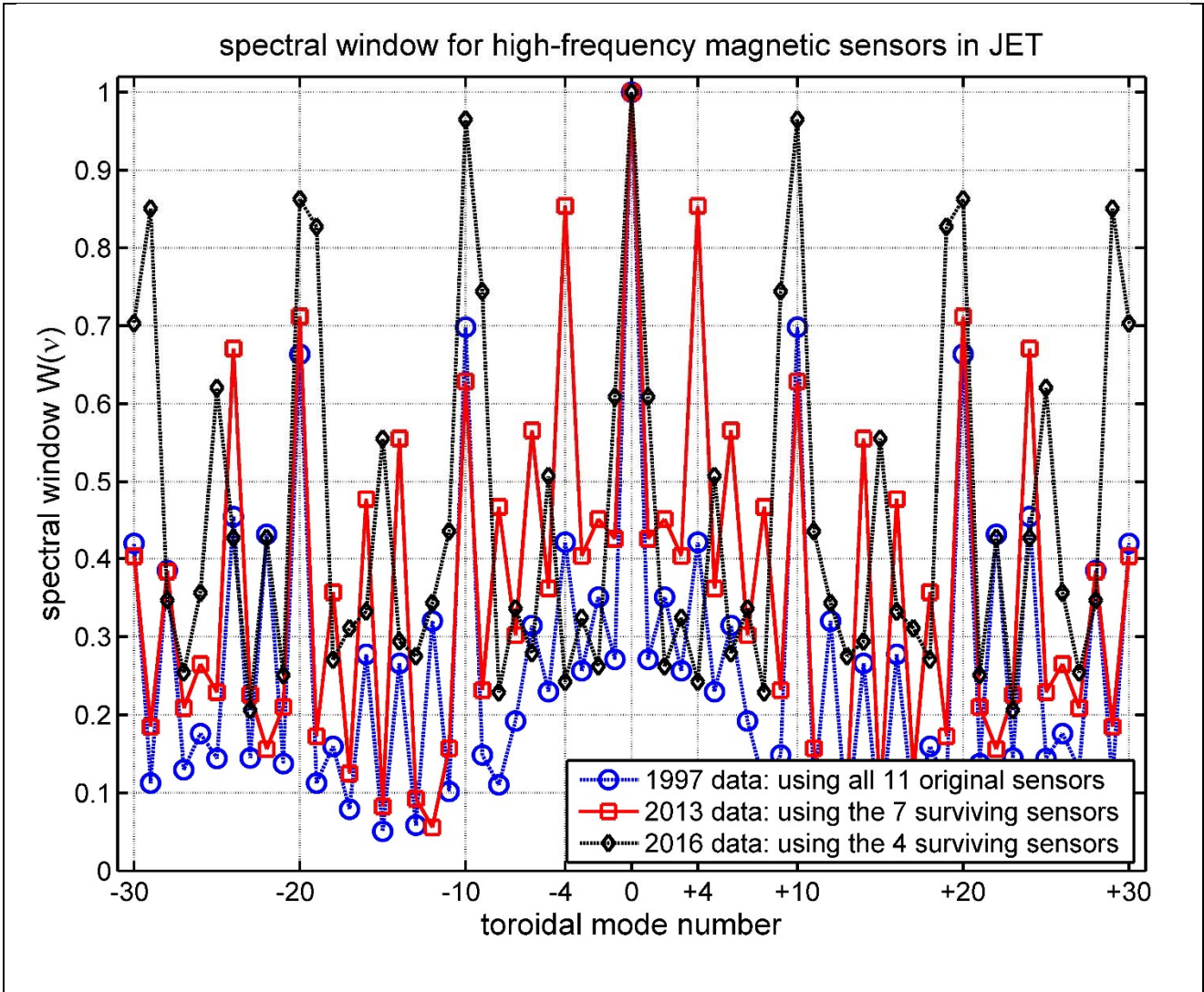
FigureA2. Results of greedy algorithms for the frequencies estimation from the data of figA1: Fourier spectrum of the data (--), true spectral lines (x), estimated lines (o) and estimation residual (-). For the MP, OMP and OLS algorithms, the order of the frequencies detection is indicated. In the case of *Clean*, the detected frequencies are represented as well as the final spectrum obtained after convolution with a Gaussian kernel (-.).

FigureA3, D.Testa et al., PaperPPCF2016.



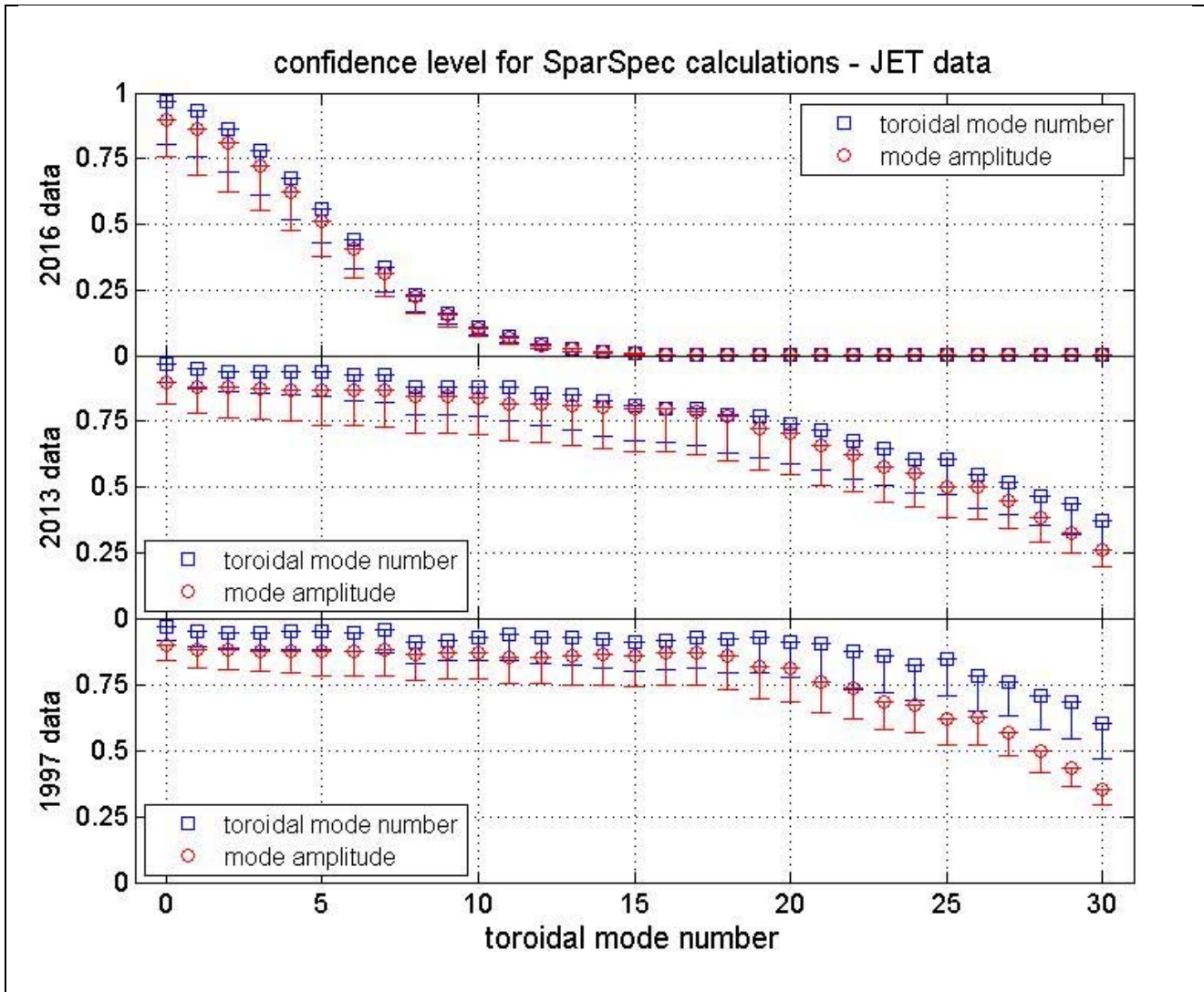
FigureA3. Results of convex relaxation for the frequencies estimation from the data of figA1: Fourier spectrum of the data (--), theoretical lines (x), estimated lines (o) and estimation residual (-). Results given with the L_1 -norm penalized least squares criterion (left frame) and after barycentric frequencies correction and least squares amplitude re-estimation (right frame).

FigureA4, D.Testa et al., PaperPPCF2016.



FigureA4 (partially reproduced and adapted from [27]). The spectral windows $W(v)$ for the original (1997) and complete set of 11 high-frequency magnetic sensors of JET usable for toroidal mode number analysis, and for the 7 and 4 surviving sensors between them that could be used in 2013 and currently (2016). Note that the original secondary lobe at $n=\pm 10$ with $W(v)=0.70$ was supplemented already in 2013 by an even higher secondary lobe at $n=\pm 4$ with $W(v)=0.87$, which is much more difficult to deal with. Moreover, the failure of three additional sensors between 2013 and the end of 2015 means that now there is an even more problematic secondary lobe at $n=\pm 1$ with $W(v)=0.62$.

FigureB1, D.Testa et al., PaperPPCF2016.



FigureB1. The confidence level in achieving the ITER measurement requirements for the evaluation of the toroidal mode number and mode amplitude when applying the *SparSpec* algorithm to a synthetic dataset defined as in eq.(B1) using the high-frequency magnetic sensors available in JET in 1997 (initial configuration, bottom frame), 2013 (middle frame) and 2016 (top frame). The simulation results are symmetric with respect to the toroidal mode number within the simulation accuracy, hence only the positive- n half spectrum is plotted. In excess of 50'000 simulation runs for each setup were performed to produce this graph, using a frequency-degenerate input mode spectrum that consists of up to 10 modes, with randomized relative amplitudes $A_k=0.05 \rightarrow 1.00$, relative phase $\delta_k=0 \rightarrow 1.95*\pi$ and input toroidal mode numbers up to $|n| \leq 30$. The two main *SparSpec* run-time analysis parameters λ_{NORM} and f_{MAX} were scanned in the range $0.05 \leq \lambda_{NORM} \leq 0.95$ and $60 \leq |f_{MAX}| \leq 200$, respectively.