

# Extraction of coherent bursts from turbulent edge plasma in magnetic fusion devices using orthogonal wavelets

Marie Farge,<sup>1</sup> Kai Schneider,<sup>2</sup> and Pascal Devynck<sup>3</sup>

<sup>1</sup>*LMD-IPSL-CNRS, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France*

<sup>2</sup>*LMSNM-CNRS & CMI, Université de Provence,*

*39 rue F. Joliot-Curie, 13453 Marseille Cedex 13, France*

<sup>3</sup>*Association EURATOM-CEA, CEA/DSM/DRFC,*

*Centre de Cadarache, 13108 Saint Paul Lez Durance, France*

(Dated: January 3, 2006)

A new method to extract coherent bursts from turbulent signals is presented. It is based on the wavelet representation which keeps track of both time and scale and thus preserves the temporal structure of the analyzed signal. This is in contrast to the Fourier representation which scrambles the temporal information among the phases of all coefficients. It is shown that, using an orthogonal wavelet basis, a turbulent signal can be decomposed into coherent and incoherent components which are orthogonal and whose properties can thus be studied independently. Diagnostics designed from the wavelet coefficients are introduced to compare the statistical properties of the original signal with its coherent and incoherent contributions. The wavelet-based extraction method is applied to the saturation current fluctuations measuring the plasma density fluctuations at the edge of the tokamak Tore Supra, Cadarache, France. This procedure disentangles coherent bursts from incoherent background fluctuations. One finds that the coherent contribution contains most of the density variance, is intermittent and correlated with non-Gaussian statistics. In contrast, the incoherent contribution is much weaker, non-intermittent, noise-like and almost decorrelated with quasi-Gaussian statistics. We conjecture that the coherent bursts are responsible for turbulent transport, while the remaining incoherent fluctuations only contribute to turbulent diffusion.

## I. INTRODUCTION

### A. Coherent bursts

The radial transport at the edge of tokamaks is known to be dominated by turbulent processes. Understanding them is important, as they will determine the confinement properties of the overall plasma in the bulk region and the energy density that must be handled by the limiter or divertor components in the shadowed region of the plasma where the magnetic field lines are opened. The turbulent transport of plasma density has been extensively studied at the edge of plasma by means of Langmuir probes [6, 13, 28], particles beams [24, 25] and more recently 2D visible imaging [32, 35]. All these diagnostics observe a turbulent transport of the plasma density in the scrape-off layer (SOL) that can be described as a superposition of convective events, which are responsible for the transport of matter over long radial distances at a fraction (of the order of 10%) of the ion sound speed [1, 7], and of background turbulence.

The convective events are detected as coherent bursts of plasma density, but with a signature different from the one expected for turbulent eddies, since they exhibit a probability distribution function (PDF) which is skewed. Typically, it is found that these convective events account for a small fraction of the time and substantial proportion of the turbulence intensity [2], which underlines their importance in the turbulent transport. There are many efforts to analyse these bursts independently from the background turbulence. For this purpose different extraction methods have been developed, which are based

on signal clipping (*e.g.*, [2]), correlation with given templates or conditional averaging. These methods require strong hypotheses on the signal, which has to be statistically steady, and also on the bursts in order to choose the appropriate threshold value. Actually the clipping method presents two drawbacks. Firstly, the duration of the bursts and their turbulent intensity strongly varies depending on the threshold value (*e.g.*, from 4% to 20% of the total time and between 20% and 50% of the total turbulent intensity [2]), which unfortunately cannot be estimated *a priori*. Secondly, the clipping method does not preserve the regularity [9] of the signal, since the threshold introduces discontinuities which affect the Fourier spectrum and hence yields an erroneous scaling. Although these methods give some information about the dynamics [2, 8], other methods requiring less hypotheses to extract the bursts are needed.

Since 1988 we have proposed to use the wavelet representation to analyze [14, 15] and extract [18, 19, 22] coherent structures out of turbulent flow fields, as the wavelet representation does not require any hypothesis on the statistical stationarity and homogeneity of the process under study. In this paper we demonstrate the advantages of wavelets to separate coherent bursts from turbulent fluctuations in edge plasma. We present a wavelet-based extraction algorithm, which does not even require any parameter, such as threshold value, to be adjusted. We then apply it to study the plasma density fluctuations measured in the SOL of the tokamak Tore-Supra, Cadarache, France [10].

## B. Wavelet representation

Since turbulent signals are highly fluctuating, one studies them statistically, using classical diagnostics such as correlation functions, spectra or structure functions. Unfortunately those diagnostics lose the temporal structure of the signal, since they are computed with time integrals and the Fourier modes used as basis functions are not localized in time.

The wavelet transform is more appropriate than the Fourier transform to analyze and represent non stationary, non homogeneous and intermittent signals, such as those encountered in turbulence. It uses analyzing functions which are generated by translation and dilation of a so-called 'mother wavelet', well localized (*i.e.*, have a finite support) in both physical and spectral space. In contrast, the Fourier transform uses trigonometric functions, which are non local (having an infinite support) in physical space but well localized in spectral space, and the analyzing functions are generated by modulation rather than dilation. The localization of the basis functions and the invariance group of the transform constitute the main differences between wavelet and Fourier representations. For a general presentation of the different types of wavelet transforms and their applications to turbulence, we refer the reader to several review articles [16, 17, 20].

Trigonometric functions used by the Fourier transform oscillate for all times and the temporal information of the transformed signal is scrambled among the phases of all Fourier coefficients. In contrast, the wavelet coefficients preserve the temporal properties of the signal. Thus, when a wavelet coefficient is filtered out, the effect on the reconstructed signal remains local in time and does not affect the overall signal, as the Fourier transform does. This property allows to study the behaviour of a limited portion of the signal directly from its wavelet coefficients.

If a turbulent signal is stationary, non intermittent and supposed to be made up of a superposition of waves, not having any nonlinear behaviour such as chirps, solitons or shocks, only in this case one can define without ambiguity the associated frequencies. However, if a turbulent signal is supposed to be a superposition of elementary structures localized in space and time, and nonlinearly interacting (*e.g.*, vortices, shocklets), the wavelet representation should be preferred, because it preserves the locality of information in both space and scale. Actually, these two different transforms translate into mathematical language two different interpretations of turbulent signals [16].

In the context of plasma physics the continuous wavelet transform has already been used to analyze signals measured in magnetic fusion devices, see *e.g.*, [12, 27]. In this paper we propose to use the orthogonal wavelet transform instead, since it has been proved to be optimal for denoising signals corrupted with additive Gaussian white noise [11]. A generalisation to correlated noise is straightforward, and a similar method has been developed [30]

to treat non Gaussian noises, *i.e.*,  $\chi^2$  distribution. To improve the choice of the threshold we have proposed a recursive algorithm [3], that we have applied to extract coherent structures out of incompressible turbulent flows [18]. In the present paper we demonstrate its use to study turbulence in edge plasmas of magnetic fusion devices, such as tokamaks or stellarators.

## C. Content of the paper

The paper is organized as follows. First, we present the wavelet-based extraction method. We then explain the recursive algorithm and validate it on an academic signal. We finally apply it to a saturation current signal measured in the SOL of the tokamak Tore Supra, Cadarache. We thus show that the coherent bursts can be efficiently extracted. We also present new statistical diagnostics based on the wavelet representation, that we use to compare the original signal with its coherent and incoherent components. Finally, some conclusions are drawn and perspectives for future work are given.

## II. EXTRACTION OF COHERENT BURSTS

### A. Principle

We propose a new method to extract coherent structures from turbulent flows, as encountered in fluids (*e.g.*, vortices, shocklets) or plasmas (*e.g.*, bursts), in order to study their role in transport and mixing.

As already mentioned, we first replace the Fourier representation by the wavelet representation, which keeps track of both time and scale, instead of frequency only. The second improvement consists in changing our viewpoint about coherent structures. Since there is not yet an universal definition of coherent structures in turbulent flows, we prefer starting from a minimal but more consensual statement about them, that everyone hopefully could agree with: '*coherent structures are not noise*'. Using this apophatic method we propose the following definition: '*coherent structures correspond to what remains after denoising*'.

For the noise we use the mathematical definition stating that a noise cannot be compressed in any functional basis. Another way to say this is to observe that the shortest description of a noise is the noise itself. Notice that plasma physicists typically call 'noise' what is actually 'experimental noise', measured when there is no plasma. Their definition includes what we define as 'noise', plus possibly some organized features (*e.g.*, parasite waves) that we do not consider as 'noise' according to the mathematical definition above.

This new way of thinking about coherent structures presents the advantage of being able to process 'incomplete fields'. What does this mean? A typical example

of incompleteness is encountered in the experimental setting, where typically one measures the time evolution of a three dimensional field using a probe located at one point, thus obtaining a one dimensional cut of a four dimensional space-time field. Notice that incompleteness is different from discretization, *i.e.*, sampling, that one should also take into account. If the algorithm used to extract coherent structures requires templates of typical structures, it becomes intractable when the measured field is incomplete, because, in order to define the template, one should then consider how the probe sees all possible motions and distortion of the coherent structures passing by in order to define the templates. Since our algorithm requires a model of the noise, but not of the coherent structures themselves (no templates are needed), it treats any field, complete or incomplete, the same way.

Considering our definition of coherent structures, turbulent signals are split into two contributions: coherent bursts, corresponding to that part of the signal which can be compressed in a wavelet basis, plus incoherent noise, corresponding to that part of the signal which cannot be compressed, neither in wavelets nor in any other basis. We will then check *a posteriori* that the incoherent contribution is spread, and therefore does not compress, in both Fourier and grid point basis. Since we use the orthogonal wavelet representation, both coherent and incoherent components are orthogonal and therefore the  $L^2$ -norm, *i.e.*, energy, is the sum of coherent and incoherent contributions.

Assuming that coherent structures are what remains after denoising, we need a model, not for the structures, but for the noise. As a first guess, we choose the simplest model and suppose the noise to be additive, Gaussian and white, *i.e.*, uncorrelated. Having this model in mind, we then rely on Donoho's and Johnstones theorem [11] to compute the value used to threshold the wavelet coefficients. Since the threshold value depends on the variance of the noise, which in the case of turbulence is not *a priori* known, we propose a recursive method to estimate it from the variance of the weakest wavelet coefficients, *i.e.*, those whose modulus is below the threshold value.

After applying our algorithm to a turbulent signal, we then check *a posteriori* that the incoherent component is indeed noise-like, spread in physical space, quasi-Gaussian and quasi-uncorrelated (*i.e.*, also spread in Fourier space), which thus confirms the hypotheses we have chosen for the noise.

## B. Orthogonal wavelet representation

The construction of orthogonal wavelet bases and the associated fast numerical algorithm are based on the mathematical concept of multiresolution analysis, which considers approximations at different scales. A function or a signal (sampled function) can thus be decomposed into a set of embedded coarser and coarser approximations. The originality of the wavelet representation is to

encode the differences between successive finer approximations, instead of the approximations themselves. The amount of information needed to go from a coarse approximation to a finer approximation is then described using orthogonal wavelets. A function or a signal is thus represented by its coarsest approximation, encoded by the scaling coefficients, plus the differences between the successive finer approximations, encoded by the wavelet coefficients.

We consider a signal  $S(t)$  of duration  $T$  sampled on  $N = 2^J$  equidistant instants  $t_i = iT/N$ , with  $i = 0, \dots, N - 1$ . We project it onto an orthogonal wavelet basis [16, 26] to represent it at different instants  $t_i$  and different time scales  $\tau = 2^{-j}$ , with  $j = 0, \dots, J - 1$ .

The signal is thus developed into an orthogonal wavelet series,

$$S(t) = \bar{S}_{00}\phi_{00}(t) + \sum_{(j,i) \in \Lambda_J} \tilde{S}_{ji} \psi_{ji}(t) \quad (1)$$

where  $\phi_{00}$  is the scaling function and  $\psi_{ji}$  the corresponding wavelets,  $i$  is the index for the instant  $t$  and  $j$  the index for the time scale  $\tau$ . To simplify the notation, we introduce  $\Lambda_J$ , which indexes all wavelets constituting the basis, defined as

$$\Lambda_J = \{(j, i), j = 0, \dots, J - 1, i = 0, \dots, 2^j - 1\}. \quad (2)$$

Due to orthogonality of the basis functions, the coefficients are computed using the  $L^2$  inner-product, denoted by  $\langle f, \psi \rangle = \int_{-\infty}^{\infty} f(t)\psi(t)dt$ . The scaling coefficients are  $\bar{S}_{00} = \langle S, \phi_{00} \rangle$  and the wavelet coefficients are  $\tilde{S}_{ji} = \langle S, \psi_{ji} \rangle$ . The scaling coefficients encode the approximation of the function  $S$  at the largest scale  $\tau_0 = 2^0 = 1$ , which corresponds to the mean value, while the wavelet coefficients encode the differences between approximations at two different time scales which correspond to the details added to get a finer time resolution. In this paper we use the Coifman 12 wavelet, which generates all functions of the wavelet basis from a set of two discrete filters, a low-pass and a band-pass filter, each of length 12 [26]. The scaling function  $\phi(t)$ , defined by the low-pass filter, and the corresponding wavelet  $\psi(t)$ , defined by the band-pass filter, together with the modulus of their Fourier transforms  $|\hat{\phi}(\omega)|$  and  $|\hat{\psi}(\omega)|$ , are shown in Fig. 1. The Fourier transform we use is defined by

$$\hat{\phi}(\omega) = \int_{-\infty}^{\infty} \phi(t) e^{-i2\pi\omega t} dt \quad , \quad (3)$$

with  $\iota = \sqrt{-1}$  and where  $\omega$  denotes the frequency.

## C. Wavelet denoising

As explained above, we define the coherent bursts to be what remain after denoising the turbulent signal  $S(t)$ . We then propose a wavelet-based method to split the signal  $S(t)$  into two orthogonal components: the coherent

signal  $S^C(t)$ , which retains the coherent bursts, and the incoherent signal  $S^I(t)$ , which corresponds to the turbulent fluctuations assumed to be noise-like. For this we first project  $S(t)$  onto an orthogonal wavelet basis and we compute a threshold value  $\epsilon$ . We then separate the wavelet coefficients  $\tilde{S}_{ij}$  into two classes: those whose modulus is larger than the threshold value  $\epsilon$  correspond to the coherent coefficients  $\tilde{S}_{ij}^C$ , while the remaining coefficients correspond to the incoherent coefficients  $\tilde{S}_{ij}^I$ . Finally, the coherent component is reconstructed in physical space using the inverse wavelet transform to get  $S^C(t)$ , while the incoherent components is easily obtained as  $S^I(t) = S(t) - S^C(t)$ . It could also be obtained by applying the inverse wavelet transform to  $\tilde{S}_{ij}^I$ .

We choose the simplest model for the noise to be eliminated, we suppose it to be additive, Gaussian and white. If we know *a priori* the noise's variance  $\sigma^2$ , the optimal threshold value is given by

$$\epsilon_D = (2\sigma^2 \ln N)^{1/2}. \quad (4)$$

Indeed, Donoho and Johnstone [11] have proven that such a wavelet thresholding is optimal to denoise signals in presence of additive Gaussian white noise, because it minimizes the maximal  $L^2$ -error (between the denoised signal and the noise-free signal) for functions with inhomogeneous regularity, such as intermittent signals. However, to compute the threshold  $\epsilon_D$  the variance of the noise has to be known.

In [3, 18] we have proposed a recursive algorithm to estimate the variance of the noise when it is not known *a priori*, which is the case for most practical applications, in particular for coherent burst extraction. The recursive algorithm is based on the observation that, given a threshold  $\epsilon_n$ , the variance of the noise estimated using Parseval's theorem

$$\sigma_n^2 = \frac{1}{N} \sum_{(j,i) \in \Lambda^J, |\tilde{S}_{ji}| < \epsilon_n} |\tilde{S}_{ji}|^2 \quad (5)$$

yields a new variance  $\sigma_{n+1}^2$  and hence a threshold  $\epsilon_{n+1}$  closer to the optimal threshold  $\epsilon_D$  than  $\epsilon_n$ . In [3] we studied the mathematical properties of this algorithm and proved its convergence for signals having sufficiently sparse representation in wavelet space, such as intermittent signals.

#### D. Algorithm

The recursive extraction algorithm can be summarized as follows:

##### Initialization

- given the signal  $S(t)$  of duration  $T$ , sampled on an equidistant grid  $t_i = iT/N$  for  $i = 0, N-1$ , with  $N = 2^J$ ,

- set  $n = 0$  and perform a wavelet decomposition, *i.e.*, apply the Fast Wavelet Transform [26] to  $S$  to obtain the wavelet coefficients  $\tilde{S}_{ji}$  for  $(j, i) \in \Lambda_J$ ,
- compute the variance  $\sigma_0^2$  of  $S$  as a rough estimate of the variance of the incoherent signal  $S^I$  and compute the corresponding threshold  $\epsilon_0 = (2 \ln N \sigma_0^2)^{1/2}$ , where  $\sigma_0^2 = \frac{1}{N} \sum_{(j,i) \in \Lambda^J} |\tilde{S}_{ji}|^2$ ,
- set the number of coefficients considered as noise to  $N_I = N$ , *i.e.*, to the total number of wavelet coefficients.

##### Main loop

Repeat

- set  $N_I^{old} = N_I$  and count the number of wavelet coefficients smaller than  $\epsilon_n$ , which yields a new value for  $N_I$ ,
- compute the new variance  $\sigma_{n+1}^2$  from the wavelet coefficients smaller than  $\epsilon_n$ , *i.e.*,  $\sigma_{n+1}^2 = \frac{1}{N} \sum_{(j,i) \in \Lambda^J} |\tilde{S}_{ji}^I|^2$ , where

$$\tilde{S}_{ji}^I = \begin{cases} \tilde{S}_{ji} & \text{for } |\tilde{S}_{ji}| \leq \epsilon_n \\ 0 & \text{else,} \end{cases} \quad (6)$$

and the new threshold  $\epsilon_{n+1} = (2 \ln N \sigma_{n+1}^2)^{1/2}$ ,

- set  $n = n + 1$

until  $(N_I == N_I^{old})$ .

##### Final step

- reconstruct the coherent signal  $S^C$  from the coefficients  $\tilde{S}_{ji}^C$  using the inverse Fast Wavelet Transform, where

$$\tilde{S}_{ji}^C = \begin{cases} \tilde{S}_{ji} & \text{for } |\tilde{S}_{ji}| > \epsilon_n \\ 0 & \text{else} \end{cases} \quad (7)$$

- finally, compute pointwise the incoherent signal  $S^I(t_i) = S(t_i) - S^C(t_i)$  for  $i = 0, \dots, N-1$ .

##### End

Note that the decomposition yields  $S(t) = S^C(t) + S^I(t)$  and orthogonality implies that energy is split into  $\sigma^2 = \sigma_C^2 + \sigma_I^2$ , since  $\langle S^C, S^I \rangle = 0$ .

The Fast Wavelet Transform (FWT), proposed by Mallat [26], requires  $(2mN)$  multiplications for its computation, where  $m$  is the length of the discrete filter defining the orthogonal wavelet used. Hence, the extraction algorithm we propose is computed in  $(2nmN)$  operations, with a number of iterations  $n$  very small, typically less than  $\log_2 N$ . Recall that the operation count for the Fast Fourier Transform (FFT) is proportional to  $N \log_2 N$  operations.

This algorithm defines a sequence of estimated thresholds  $(\epsilon_n)_{n \in \mathbb{N}}$  and the corresponding sequence of estimated variances  $(\sigma_n^2)_{n \in \mathbb{N}}$ . The convergence of these sequences within a finite number of iterations has been demonstrated in [3] applying a fixed point type argument to the iteration function

$$\mathcal{I}_{S,N}(\epsilon_{n+1}) = \left( \frac{2 \ln N}{N} \sum_{(j,i) \in \Lambda^J} |\tilde{S}_{ji}^I(\epsilon_n)|^2 \right)^{1/2}. \quad (8)$$

The algorithm thus stops after  $n$  iterations when  $\mathcal{I}_{S,N}(\epsilon_n) = \epsilon_{n+1}$ .

Furthermore, we have shown that the convergence rate of the recursive algorithm depends on the signal to noise ratio ( $SNR = 10 \log_{10}(\sigma^2/\sigma_I^2)$ ), since the smaller the SNR, *i.e.*, the stronger the noise, the faster the convergence. Moreover, if the algorithm is applied to a Gaussian white noise only, it converges in one iteration and removes the noise (in statistical mean). If it is applied to a signal without noise, the signal is fully preserved. Finally, we have proven that the algorithm is idempotent, *i.e.*, if we apply it several times, the noise is eliminated the first time, and the coherent signal is no more modified in the subsequent applications, as it would have been the case for a Gaussian filter. As a consequence, this algorithm yields a nonlinear projector [3].

### E. Application to an academic test signal

To illustrate the properties of the recursive algorithm we apply it to a one-dimensional noisy test signal  $S$  (Fig. 2, middle). This signal has been constructed by superposing a Gaussian white noise  $W$ , with zero mean and variance  $\sigma_W^2 = 1$ , to a function  $F$ , normalized such that its variance yields 10, which corresponds to a signal to noise ratio  $SNR = 10 \log_{10}(\sigma_F^2/\sigma_W^2) = 20$  dB (Fig. 2, top). The function  $F$  is a piecewise polynomial function which presents several discontinuities, either in the function or in its derivatives. The number of samples is  $N = 2^{13} = 8192$ .

We apply the recursive extraction algorithm to the test signal  $S(t)$  and obtain after  $n = 5$  iterations the coherent part  $S^C(t)$  and the incoherent noise  $S^I(t)$  (cf. Fig. 2, bottom). We observe that  $S^C(t)$  yields a denoised version of the test signal  $S(t)$  which is very close to  $F(t)$ , while the incoherent part  $S^I(t)$  is homogeneous and noise like with flatness  $\mathcal{F} = 3.03$ , which corresponds to quasi-Gaussianity. Note that the flatness  $\mathcal{F}$  is defined as the ratio of the centered fourth order moment divided by the square of the variance, and  $\mathcal{F} = 3$  for a Gaussian process. Fig. 2 (bottom, left) shows that the coherent signal retains all discontinuities and spikes present in the original function  $F(t)$ , without smoothing them as it would have been the case with standard denoising methods, *e.g.*, with low pass Fourier filtering. Nevertheless, we observe slight overshoots in the vicinity of the discontinuities, although they remain much more local than the

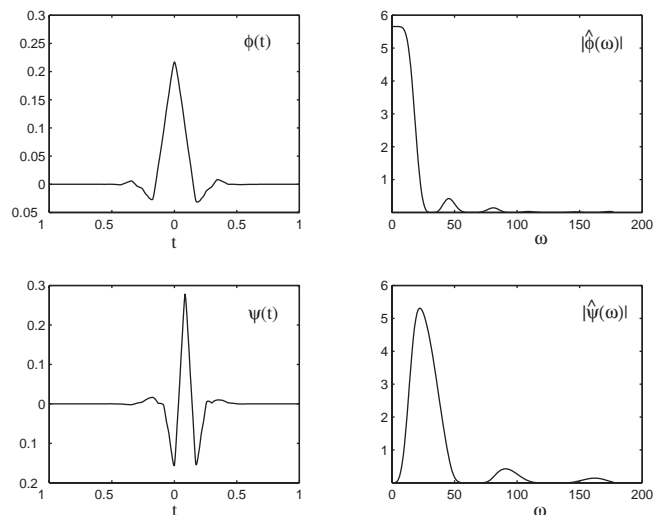


FIG. 1: Coifman 12 wavelet. Top: scaling function  $\phi(t)$  and the modulus of its Fourier transform  $|\hat{\phi}(\omega)|$ . Bottom: wavelet  $\psi(t)$  and the modulus of its Fourier transform  $|\hat{\psi}(\omega)|$ .

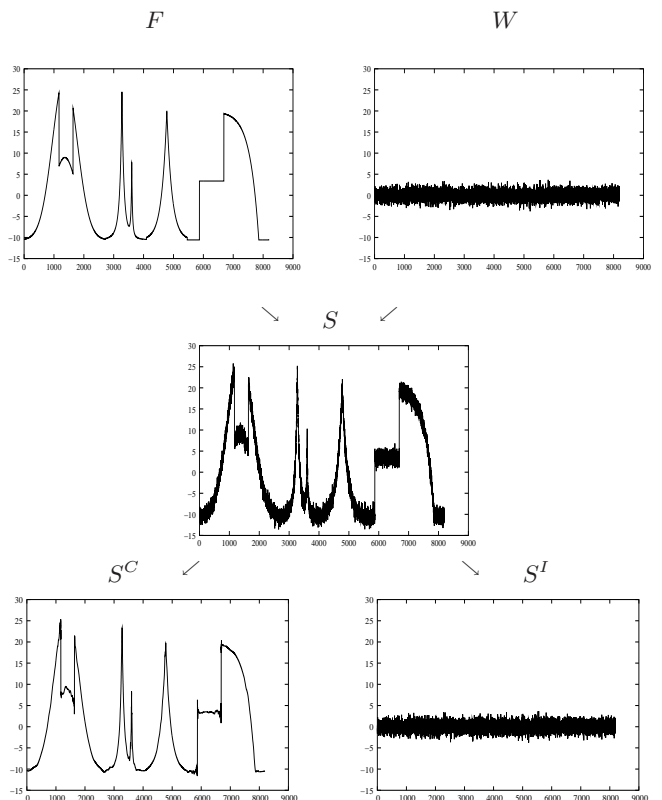


FIG. 2: Construction (top) of a 1D noisy signal  $S = F + W$  (middle), and results obtained by the recursive algorithm (bottom), which gives  $S = S^C + S^I$ .

classical Gibbs phenomena, and could easily be removed using the translation invariant wavelet transform [26].

### III. APPLICATION TO TURBULENT EDGE PLASMA

#### A. Density fluctuations

We have measured the time evolution of the ion saturation current during  $8\text{ ms}$  in the SOL of the tokamak Tore Supra in Cadarache (France). This signal, denoted  $S(t)$ , gives an approximation of the density fluctuations.

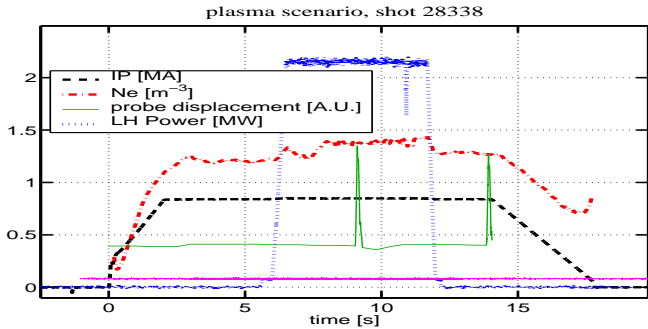


FIG. 3: Plasma scenario of the shot 28338 from the tokamak Tore Supra, Cadarache. The duration of the shot is  $18\text{ s}$ . The plasma density fluctuations are measured by a fast reciprocating Langmuir probe. When the probe is  $2.8\text{ cm}$  away from the LCFS in the SOL, the signal is acquired during time windows of  $8\text{ ms}$ .

The measure was taken according to the following plasma scenario: the shot 28338 lasted  $18\text{ s}$  and the signal has been recorded in the middle of the plasma current plateau. The large radius was  $R = 2.33\text{ m}$ , the small radius  $a = 0.77\text{ m}$ , the mean plasma density  $\bar{n}_i = 1.37 \cdot 10^{19}\text{ m}^{-3}$ , the plasma current  $I_p = 0.84\text{ MA}$  and the edge safety factor  $q = 6.71$ . Moreover,  $2.1\text{ MW}$  of lower hybrid waves were applied to the plasma.

The ion saturation current fluctuations were measured by a fast reciprocating Langmuir probe. The total duration of the probe motion into the plasma was  $300\text{ ms}$  and, when the probe reached  $2.8\text{ cm}$  away from the last closed flux surface (LCFS). The signal has been recorded at  $1\text{ MHz}$  during  $8\text{ ms}$  (Fig. 3), which gives  $N = 2^{13} = 8192$  samples. A high-pass filter at frequency  $0.1\text{ kHz}$  and a low-pass filter at frequency  $500\text{ kHz}$  have been applied to eliminate both low frequencies and aliasing.

#### B. Extraction of coherent bursts

We use the wavelet extraction algorithm to split the signal  $S(t)$  (Fig. 4, top) into two orthogonal components, the coherent bursts  $S^C(t)$  (Fig. 4, middle) and the incoherent turbulent fluctuations  $S^I(t)$  (Fig. 4, bottom). The

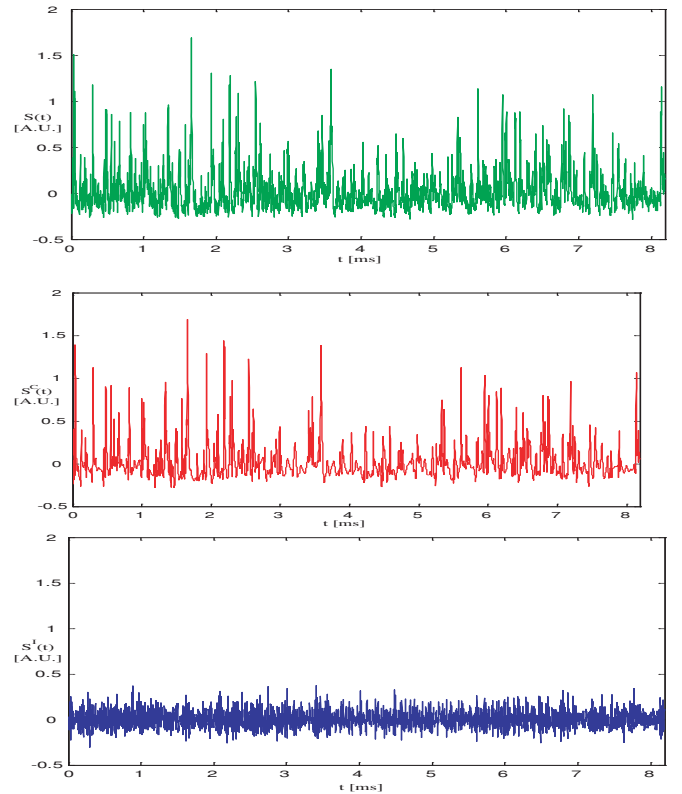


FIG. 4: Signal  $S(t)$  of duration  $8.192\text{ ms}$ , corresponding to saturation current fluctuations measured at  $1\text{ MHz}$  in the SOL of the tokamak Tore Supra, Cadarache. Top: total signal  $S$ . Middle: coherent part  $S^C$ . Bottom: incoherent part  $S^I$ .

optimal threshold value has been obtained after  $n = 12$  iterations of the algorithm (Figure 5).

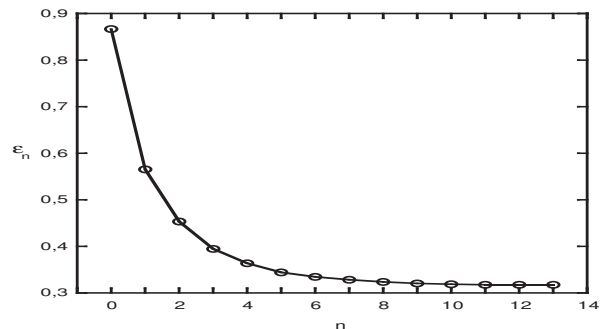


FIG. 5: Threshold value  $\epsilon_n$  versus iteration number  $n$ .

As results, we observe that the coherent signal  $S^C(t)$ , made of  $5.8\%N$  wavelet coefficients, retains  $86.6\%$  of the total variance and the extrema are preserved (Table I). In contrast, the incoherent contribution  $S^I(t)$ , is made of  $94.2\%N$  wavelet coefficients but contributes to only  $3.4\%$  of the total variance (Table I), which corresponds to a signal to noise ratio  $SNR = 10 \log_{10}(\sigma^2/\sigma_I^2) = 8.72\text{ dB}$ .

The decomposition shows that the bursty and coherent part of the signal dominates over the turbulent fluctua-

tions of the background, this more strongly than what has been found with previous methods based on clipping [2].

TABLE I: Statistical properties of the signal  $S(t)$  from the tokamak Tore Supra, Cadarache, for the signal and its coherent and incoherent components using the Coifman 12 orthogonal wavelet.

Signal	total $S$	coherent $S^C$	incoherent $S^I$
# of coefficients	8192	479	7713
% of coefficients	100 %	5.8 %	94.2 %
min value	-0.284	-0.282	-0.307
max value	1.689	1.686	0.374
mean value	0.019	0.019	$< 10^{-11}$
Variance $\sigma^2$	0.0417	0.0361	0.0056
% of variance	100 %	86.6 %	3.4 %
Skewness	2.564	2.842	0.383
Flatness	12.001	14.224	4.026

Fig. 6 shows the Probability Distribution Functions (PDFs) in log–lin coordinates for the total, coherent and incoherent contributions, estimated using histograms with 50 bins with integrals normalized to one. The PDFs of the total signal and the coherent part are skewed and present the same behaviour: positive values have exponential tails with  $p(S) \propto \exp(-5/2S)$ , while negative values yield a Gaussian behaviour (Fig. 6). In contrast, the PDF of the incoherent component is almost symmetric, with skewness 0.38, instead of 2.56 and 2.84 for the total and coherent part, respectively. It has a quasi-Gaussian shape with flatness 4.03, instead of 12.00 and 14.22, respectively (Fig. 6).

### C. Fourier spectrum and modified periodogram

To get more information on the spectral distribution of the density variance for the different components, we consider the Fourier spectrum

$$E(\omega) = \frac{1}{2} \left| \widehat{S}(\omega) \right|^2, \quad (9)$$

where  $\widehat{S}(\omega)$  denotes the Fourier transform as defined in equation (3). As estimator for the spectrum we take the periodogram, which is a discrete version of equation (9), although it is known to be a non consistent estimator due

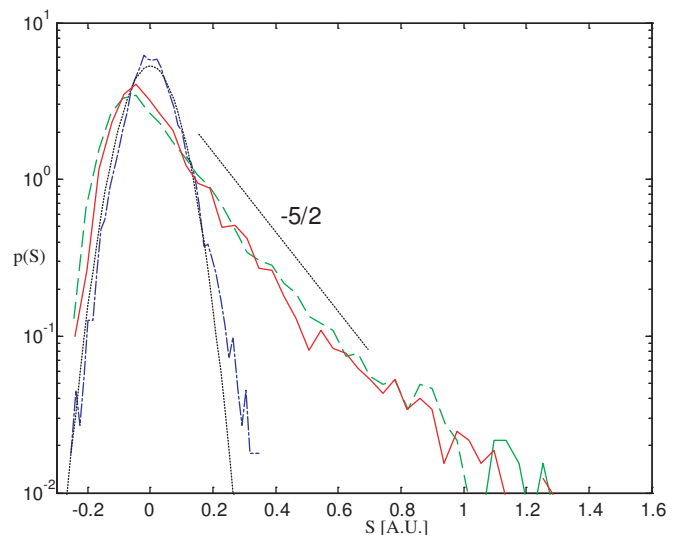


FIG. 6: Probability density function  $p(S)$ , estimated using histograms with 50 bins. PDF of the total signal  $S$  (green dashed line), of the coherent component  $S^C$  (red solid line) and of the incoherent component  $S^I$  (blue dotted-dashed line), together with a Gaussian fit with variance  $\sigma_I^2$  (black dotted line).

to the presence of oscillations [29]. To obtain a consistent estimator we also compute the modified periodogram, by first tapering the data with a raised cosine window (affecting 40 data points at each boundary), and then convolving the periodogram with a Gaussian window (with standard deviation of 40 data points). Figure 7 shows the periodogram and the modified periodogram for  $S$ ,  $S^C$  and  $S^I$ , which confirms that the latter yields a stabilized estimator of the spectrum, with no more oscillations.

### D. Wavelet spectrum

The wavelet decomposition, given in equation (1), yields the distribution of the variance of the signal scale per scale, which is called scalogram [16]. It is defined as

$$\widetilde{E}_j = \frac{1}{2} \sum_{i=0}^{2^j-1} \left( \widetilde{S}_{ji} \right)^2. \quad (10)$$

Parseval's theorem implies that  $E = \sum_{j \geq 0} \widetilde{E}_j$ . Using the relation  $\omega_j = \frac{\omega_\psi}{2^j}$  between the scale index  $j$  and the frequency  $\omega$ , the wavelet spectrum can be defined as  $\widetilde{E}(\omega_j) = \widetilde{E}_j / \omega_\psi$ , with  $\omega_\psi$  being the centroid frequency of the mother wavelet whose value is  $\omega_\psi = 1.3$  for the Coifman 12 wavelet used here. It corresponds to a smoothed version of the Fourier spectrum (9), the smoothing kernel being the square of the Fourier transform of the wavelet, since

$$\widetilde{E}(\omega) = \frac{1}{\omega_\psi} \int_0^{+\infty} E(\omega') \left| \widehat{\psi} \left( \frac{\omega_\psi \omega'}{\omega} \right) \right|^2 d\omega'. \quad (11)$$

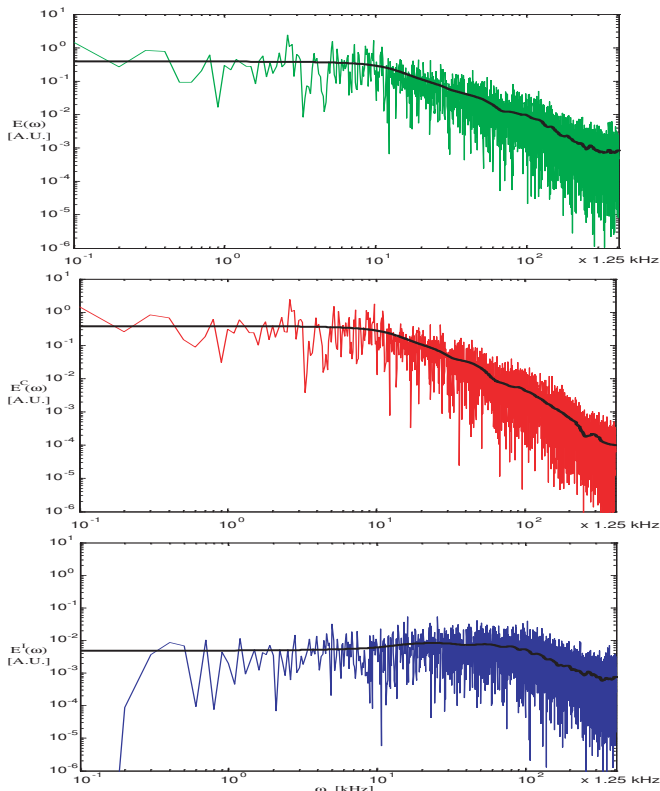


FIG. 7: Fourier spectrum  $E(\omega)$ . Top: spectrum of the total signal  $S(t)$ . Middle: spectrum of the coherent component  $S^C(t)$ . Bottom: spectrum of the incoherent component  $S^I(t)$ . Note that the periodogram is plotted in green, red and blue for the total, coherent and incoherent signal, respectively. Superimposed are the modified periodograms (black thick line).

Note that, as the frequency increases, *i.e.*, when one goes to small scale, the smoothing interval becomes larger which explains why the wavelet spectrum is a well-conditioned statistical estimator. The advantage of the wavelet spectrum in comparison to the modified periodogram is that the smoothing window is automatically adjusted by the wavelet representation, since wavelets correspond to filters with constant relative bandwidth  $\frac{\Delta\omega}{\omega}$  [16].

In Fig. 8 the wavelet spectra, together with the modified periodograms, are displayed.

We observe that the signal and its coherent component present a similar scaling in  $\omega^{-5/3}$ , which characterizes correlation since the spectral slope is negative. As proposed in [2], this may be interpreted as an inverse energy cascade, similar to what is encountered in two-dimensional fluid turbulence. In contrast, the incoherent component has a different scaling with a flat spectrum up to frequency  $\omega = 120 \text{ kHz}$ , corresponding to decorrelation. For higher frequencies we observe a  $\omega^{-1}$  scaling, which may be due to experimental noise, which presents the same scaling at high frequencies, although its amplitude remains smaller than the incoherent fluctuations. Fig. 8 also shows that the wavelet spectrum almost coin-

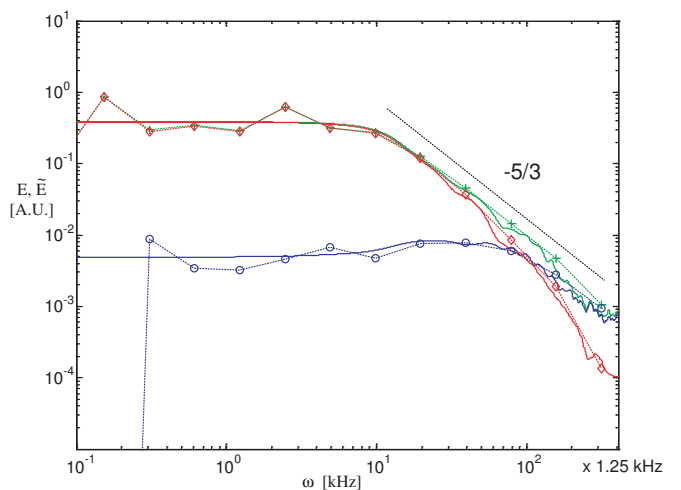


FIG. 8: Wavelet spectra  $\tilde{E}(\omega_j)$  (lines with symbols) and modified periodograms  $E(\omega)$  (lines) of the total signal  $S$  (green and +), of the coherent signal  $S^C$  (red and  $\diamond$ ) and of the incoherent signal  $S^I$  (blue and  $\circ$ ).

cides with the modified periodogram, and that the higher the frequency the better the stabilization thus obtained.

Note that the scalogram and the wavelet spectrum are optimal to characterize scaling laws, as long as the analyzing wavelet has at least  $M$  vanishing moments, with  $M > \frac{\beta-1}{2}$ , to detect power laws in  $\omega^{-\beta}$ , see *e.g.*, [20, 31].

## E. Intermittency

Intermittency characterizes the fact that the time support of the fluctuations decreases with the scale [5, 23]. It therefore quantifies how bursty a signal is. Townsend [34] has proposed to define the ‘intermittency factor’ as the ratio between the time support of active and quiescent regions. The problem is that such diagnostic depends on the choice of the threshold below which the variation is considered to be inactive [33]. As already mentioned above, one of the drawbacks of such a clipping method is that the active bursts, and therefore the corresponding intermittency factor, depend on the choice of the threshold. This can be avoided by using the wavelet representation.

Biskamp stated in [5] that ‘*the spottiness of the dissipative eddies is a special feature of what is now believed to be a general property of fully developed turbulence that with decreasing scale turbulent fluctuations become less and less space-filling, i.e., are concentrated in regions of smaller and smaller volume but increasingly complicated shape. This phenomenon is called intermittency, which is a central topic in actual turbulence research*’. Frisch explained in [23] that intermittency can be quantified by computing the variation of the flatness when scale decreases: if flatness remains constant the signal is non-intermittent, if it increases when scale decreases it is in-



termittent. We use the same definition of intermittency and compute the scale dependent flatness from the higher order moments of the wavelet coefficients  $\tilde{S}_{ji}$ , as introduced in [20, 31]. By summing up the  $p$ -th power of the wavelet coefficients over all positions  $i$ , one obtains the  $p$ -th order moments

$$\tilde{\mathcal{M}}_j^p = \frac{1}{2^j} \sum_{i=0}^{2^j-1} \left( \tilde{S}_{ji} \right)^p. \quad (12)$$

The scale dependent flatness is then defined as

$$\tilde{\mathcal{F}}_j = \frac{\tilde{\mathcal{M}}_j^4}{\left( \tilde{\mathcal{M}}_j^2 \right)^2}. \quad (13)$$

The relation between scale and frequency allows to express the flatness as function of the frequency  $\omega_j$ , similarly to the wavelet spectrum. Note that Gaussian white noise, which is by definition non-intermittent, would yield a flatness equal to three for all frequencies.

To characterize the intermittency of the signal and its different contributions we plot in Fig. 9 the flatness  $\tilde{\mathcal{F}}_j$  versus the frequency  $\omega_j$ . We observe that the flatness of the coherent contribution increases faster for high frequencies than that of the total signal. This proves that the coherent contribution is more intermittent than the signal itself, which is obvious since it only retains the bursts. In contrast, the flatness of the incoherent contribution decreases to the value  $\tilde{\mathcal{F}}_j = 3$ , up to frequency  $\omega = 120 \text{ kHz}$ , which gives evidence for its non-intermittent behaviour. The wavelet based flatness cor-

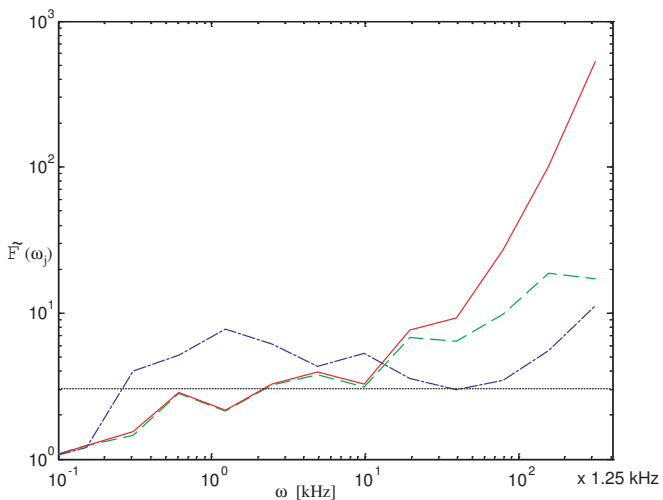


FIG. 9: Flatness of the band-pass filtered signal  $\tilde{\mathcal{F}}$  versus frequency  $\omega_j$  for the total signal  $S$  (green dashed line), of the coherent signal  $S^C$  (red solid line) and of the incoherent signal  $S^I$  (blue dotted-dashed line). The horizontal dotted line  $\tilde{\mathcal{F}}(\omega_j) = 3$  corresponds to the flatness of a Gaussian process.

responds to the flatness of the band-pass filtered signal, as typically used in the fluid turbulence community [23].

Note that the signal reconstructed from its wavelet coefficients at given scale  $j$  corresponds to the band-pass filtered signal around the frequency  $\omega_j = \omega_\psi/2^j$ .

For comparison we also show in Fig. 10 the flatness of the low-pass filtered signal, for dyadically increasing cut-off frequencies  $\omega_C = \omega_\psi/2^{J_C}$ . Therefore, we reconstruct the signal in physical space on  $N$  grid points using only wavelet coefficients up to a given scale  $J_C$ , corresponding to the filter cut-off. The wavelet coefficients for scales  $j \geq J_C$  are set to zero and the low-pass filtered signal is computed by the inverse wavelet transform using eq. (1).

Similarly to Fig. 9, we observe in Fig. 10 that the flatness of the total and coherent signal increases with frequency for  $\omega > 3 \text{ kHz}$ . Considering the signal filtered at  $20 \text{ kHz}$  we observe that its flatness is just above 7, however the signal contains only large bursts, since all smaller details have been filtered out. This shows that the signal is already intermittent at large scales. For the small scales, *i.e.*, for  $\omega \geq 20 \text{ kHz}$ , the flatness of the total and the coherent signal is above 10. This shows that adding small details to the large scale bursts increases the flatness, and hence the signal's intermittency as quantified by its flatness.

The flatness  $\mathcal{F}_<$  of the low-pass filtered signal, considered for increasing cut-off frequencies, quantifies the intermittency of the signal reconstructed up to the corresponding cut-off scales, while the flatness  $\tilde{\mathcal{F}}$  of the band-pass filtered signal, considered for bands of increasing wavenumber, yields incremental information on the flatness of the signal scale by scale. The latter quantity can be compared with the energy spectrum which gives the energy distribution scale by scale, while the former gives some cumulative information, *i.e.*, information on the flatness of the lower frequency contributions of the signal is included in the flatness of the higher frequency contributions. Hence, both quantities do not yield the same values if the PDF of the signal varies with scale.

#### IV. CONCLUSION

We presented a wavelet-based recursive method to extract coherent bursts out of turbulent signals. The algorithm decomposes the signal into an orthogonal wavelet basis and reconstructs the coherent contribution from the wavelet coefficients whose modulus is larger than a given threshold. The threshold value is recursively determined without any adjustable parameter. Moreover, we have shown that this algorithm is fast since it has only linear complexity.

Compared to classical extraction methods, which are based, either on thresholding in physical space ('clipping'), or on conditional averaging, working in wavelet space presents the following advantages:

- there is no need to suppose the signal to be statistically stationary in time,
- the wavelet decomposition preserves the spectral

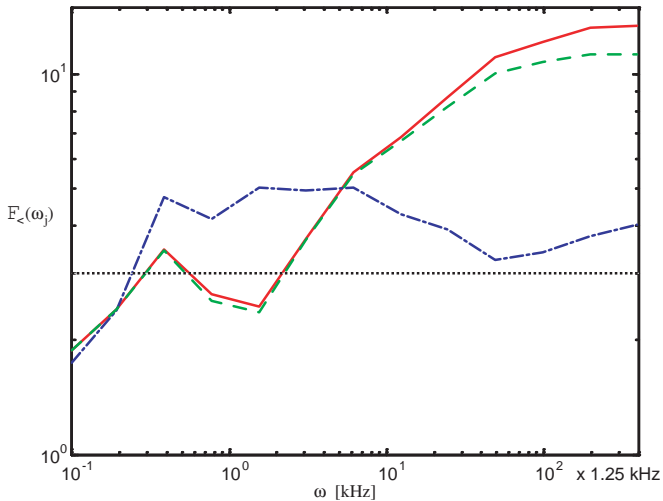


FIG. 10: Flatness of the low-pass filtered signal  $\mathcal{F}_<$  versus frequency  $\omega_j$  for the total signal  $S$  (green dashed line), of the coherent signal  $S^C$  (red solid line) and of the incoherent signal  $S^I$  (blue dotted-dashed line). The horizontal dotted line  $\mathcal{F}_>(\omega_j) = 3$  corresponds to the flatness of a Gaussian process.

properties of the signal, and thus respects its scaling as long as the analyzing wavelet is smooth enough (which depends on the number of vanishing moments for orthogonal wavelets),

- the wavelet-based extraction method does not require any prior about the shape or the intensity of the bursts to be extracted; the only prior is to assume the noise to be Gaussian and white.

We have applied this recursive wavelet algorithm to ion saturation current measured in the SOL of the tokamak Tore Supra. We have thus extracted the coherent bursts from an incoherent background noise. The former contains most of the density variance and are correlated with non-Gaussian statistics, while the latter is almost decorrelated and quasi-Gaussian. We have also observed that the non-Gaussianity of the PDF of the coherent component increases with the frequency, which confirms that the bursts are highly intermittent. In contrast, the incoherent component remains quasi-Gaussian up to high frequencies, which confirms the non intermittency of the background noise. By analogy with previous studies we have made in the context of two-dimensional fluid turbulence [4], we conjecture that the coherent bursts are due to organized structures produced by nonlinear interactions and responsible for turbulent transport. On the other hand, the incoherent background corresponds to the turbulent fluctuations which only contribute to turbulent diffusion. Moreover the variance of the incoherent fluctuations yields a good estimation of the turbulence level.

In [21] we applied this extraction method to both plasma velocity and density signals, measured at different poloidal positions, to study turbulent fluxes and

thus characterize the transport properties of the coherent bursts. These results will be subject of a forthcoming paper. We also have already extended this extraction method to treat two and three dimensional, scalar and vector, fields [18, 19, 22], and we plan to apply it to spatio-temporal signals and images of plasma density fluctuations, obtained by fast framing cameras, to improve the characterization of coherent bursts.

**Acknowledgements:** M.F. and K.S. acknowledge financial support from the contract CEA/EURATOM n°. V.3258.001. M.F. is grateful to Trinity College, Cambridge (U.K.), for its hospitality.

- 
- [1] G. Y. Antar, S. I. Krasheninnikov, P. Devynck, R. P. Doerner, E. M. Hollmann, J. A. Boedo, S. C. Luckhardt, R. W. Conn. *Phys. Rev. Letters*, **87**, 065001 (2001).
- [2] G.Y. Antar, P. Devynck, X. Garbet, S. Luckhart. *Phys. Plasmas*, **8**, 1612-1624 (2001).
- [3] A. Azzalini, M. Farge, K. Schneider. *Appl. Comput. Harm. Anal.*, **18**(2), 177–185 (2005).
- [4] C. Beta, K. Schneider, M. Farge. *Comm. in Nonlinear Science and Numerical Simulation*, **8**(3–4), 537-545 (2003).
- [5] D. Biskamp. *Nonlinear magnetohydrodynamics*. Cambridge University Press (1993).
- [6] J. Bleuel, M. Endler, H. Niedermeyer, M. Schubert, H. Thomsen and the W7- AS Team. *New Journal of Physics*, **4**, 38 (2002).
- [7] J. A. Boedo, D. Rudakov, R. Moyer, S. Krasheninnikov, D. Whyte, G. McKee, G. Tynan, M. Schaffer, P. Stangeby, P. West, S. Allen, T. Evans, R. Fonck, E. Hollmann, A. Leonard, A. Mahdavi, G. Porter, M. Tillack, G. Antar. *Phys. Plasmas*, **8**(11), 4826-4833 (2001).
- [8] J. A. Boedo, D. L. Rudakov, R. A. Moyer, G. R. McKee, R. J. Colchin, M. J. Schaffer, P. G. Stangeby, W. P. West, S. L. Allen, T. E. Evans, R. J. Fonck, E. M. Hollmann, S. Krasheninnikov, A. W. Leonard, W. Nevins, M. A. Mahdavi, G. D. Porter, G. R. Tynan, D. G. Whyte, X. Xu. *Phys. Plasmas*, **10**(5), 1670-1677 (2003).
- [9] The regularity of degree  $r$  of a function implies that the function and its derivatives up to degree  $r$  are continuous and therefore do not present any discontinuity.
- [10] P. Devynck, X. Garbet, Ph. Ghendrih, J. Gunn, C. Honor, B. Pegourie, G. Antar, A. Azeroual, P. Beyer, C. Boucher, V. Budaev, H. Capes, F. Gervais, P. Hennequin, T. Loarer, A. Quéméneur, A. Truc, J.C. Vallet. *Nuclear Fusion*, **42**, 697-707 (2002).
- [11] D. Donoho, I. Johnstone. *Biometrika*, **81**, 425-455 (1994).
- [12] V. Dose, G. Venus, H. Zohm. *Phys. Plasmas*, **4**(2), 323–328 (1997).
- [13] M. Endler, H. Niedermeyer, L. Giannone, E. Holzauer, A. Rudyj, G. Theimer, N. Tsois and the ASDEX Team. *Nucl. Fusion*, **35**, 1307 (1995).
- [14] M. Farge and G. Rabreau. *C. R. Acad. Paris.*, **307**, II, 1479–1486 (1988).
- [15] M. Farge and R. Sadourny. *J. Fluid Mech.*, **206**, 433–462 (1989).
- [16] M. Farge. *Ann. Rev. of Fluid Mech.*, **24**, 395–457 (1992).
- [17] M. Farge, N. Kevlahan, V. Perrier and E. Goirand. *IEEE Proc.*, **84**, 4, 639–669 (1996).
- [18] M. Farge, K. Schneider and N. Kevlahan. *Phys. Fluids*, **11**, 2187–2201 (1999).
- [19] M. Farge, G. Pellegrino and K. Schneider. *Phys. Rev. Lett.*, **87** (5), 054501 (2001).
- [20] M. Farge and K. Schneider. *New trends in turbulence. Les Houches 2000*, **74** (Eds. M. Lesieur, A. Yaglom, F. David), Springer, 449-503 (2002).
- [21] M. Farge, K. Schneider and P. Devynck. Extraction des événements cohérents et estimation des flux de particules dans les signaux de plasma de bord à l'aide des paquets d'ondelettes et des ondelettes. *Annual report CEA-Euratome*, Cadarache, V.3258.001 (2003).
- [22] M. Farge, K. Schneider, G. Pellegrino, A. A. Wray and R. S. Rogallo. *Phys. Fluids*, **10**, 2886–2896 (2003).
- [23] U. Frisch. *Turbulence*. Cambridge University Press (1995).
- [24] G. R. McKee, R. Ashley, R. Durst, R. Fonck, M. Jakubowski, K. Tritz, K. Burrell, C. Greenfield, J. Robinson. *Rev. Sci. Instrum.*, **70**, 913-916 (1999).
- [25] G. R. McKee, C. Fenzi, R. J. Fonck, M. Jakubowski. *Rev. Sci. Instrum.*, **74**, 2014-2019 (2003).
- [26] S. Mallat. *A wavelet tour of signal processing*. Academic Press (1998).
- [27] B. Ph. van Miligen, C. Hidalgo and E. Sanchez., *Phys. Rev. Lett.*, **74**, 3, 395 (1995).
- [28] J. Petržilka, J. Stöckel. *Contributions to Plasma Physics*, **38**, Special Issue, 74-79 (1998).
- [29] M.B. Priestley. *Spectral analysis and time series*. Academic Press (1981).
- [30] R. von Sachs, K. Schneider. *Appl. Comput. Harm. Anal.*, **3**, 268–282 (1996).
- [31] K. Schneider, M. Farge, N. Kevlahan. *Woods Hole Mathematics, Perspectives in Mathematics and Physics, Vol. 34* (Eds. N. Tongring and R.C. Penner), World Scientific, 302-328 (2004).
- [32] J. L. Terry, S. J. Zweben, K. Hallatschek, B. LaBombard, R. J. Maqueda, B. Bai, C. J. Boswell, M. Greenwald, D. Kopon, W. M. Nevins, C. S. Pitcher, B. N. Rogers, D. P. Stotler, and X. Q. Xu. *Phys. Plasmas*, **10**, 1739-1747 (2003).
- [33] A. Tsinober. *An informal introduction to turbulence*. Kluwer (2001).
- [34] A. A. Townsend. *Austr. J. Sci. Res.*, **1**, 161 (1948).
- [35] S.J. Zweben, D. P. Stotler, J. L. Terry, B. LaBombard, M. Greenwald, M. Muterspaugh, C. S. Pitcher, K. Hallatschek, R. J. Maqueda, B. Rogers, J. L. Lowrance, V. J. Mastrocola, and G. F. Renda Alcatraz C-Mod Group. *Phys. Plasmas*, **9**, 1981-1989 (2002).