

Reliability of the two-point measurement of the spatial correlation length from Gaussian-shaped fluctuating signals in fusion-grade plasmas

Jaewook Kim,^{1,*} Y.U. Nam,² M. Lampert,³ and Y.-c. Ghim^{1,†}

¹*Department of Nuclear and Quantum Engineering, KAIST, Daejeon 34141, Korea*

²*National Fusion Research Institute, Daejeon, Republic of Korea*

³*Wigner RCP, Euratom Association-HAS, Budapest, Hungary*

(Dated: June 20, 2021)

A statistical method for the estimation of spatial correlation lengths of Gaussian-shaped fluctuating signals with two measurement points is examined to quantitatively evaluate its reliability (variance) and accuracy (bias error). The standard deviation of the correlation value is analytically derived for randomly distributed Gaussian shaped fluctuations satisfying stationarity and homogeneity, allowing us to evaluate, as a function of fluctuation-to-noise ratios, sizes of averaging time windows and ratios of the distance between the two measurement points to the true correlation length, the goodness of the two-point measurement for estimating the spatial correlation length. Analytic results are confirmed with numerically generated synthetic data and real experimental data obtained with the KSTAR beam emission spectroscopy diagnostic. Our results can be applied to Gaussian-shaped fluctuating signals where a correlation length must be measured with only two measurement points.

I. INTRODUCTION

Plasma turbulence is an intellectually interesting phenomenon as we do not have full capability of predicting a cause, an evolution and a consequence of it; yet it heavily influences the world around us. For instance, in both laboratory and astrophysical plasmas, a possible explanation for the amplification of magnetic fields with *turbulence* is provided [1] (and the references therein); while plasma *turbulence* is believed to be one of the major obstacles for harnessing nuclear fusion power economically [2].

It is, therefore, important to acquire and provide statistically accurate (precision to the true value or bias error) and reliable (variance of the measured values) experimental measurements of turbulence. Probably, the best and easiest way to do so is by using a large number of temporally fast detectors covering the space of interest. Temporally fast detectors are becoming widely available; on the other hand, using a large number of such detectors may not be a feasible solution due to many reasons such as installation difficulties, bandwidth processing and limited resources on budgets depending on applications. Also, we do not want to cover the whole space just with the detectors if they are in-situ types. For this reason, there has been previous studies on obtaining the spatial structure of fluctuating signals with two measurement points in laboratory plasmas [3, 4] and with four measurement points in solar winds [5]. Note that this four-point measurement in solar winds is really a two-point measurement as the four points are not aligned in a straight line (or to a background magnetic field line).

As we do not find any systematic studies on the reliability of the obtained two-point correlation function,

i.e., variance of the correlation function, which is equally as important as the correlation function itself, we investigate how accurately and reliably one can measure the correlation function with only two measurement points. This is then used to calculate the accuracy and reliability of the estimated spatial correlation length.

We focus on the *two-point* measurement because it is the smallest necessary number to get the spatial structure unless Taylor's hypothesis [6] can be validly applied where temporal information at one spatial position contains upstream spatial information. We focus on the *spatial correlation length* because it is one of the basic properties of turbulence. Decorrelation rate in time and fluctuation levels are also important characteristics, however these suffer less from system hardware due to availability of many time points. Furthermore, turbulence is believed to be critically balanced [7–10] as reported in observations of solar winds [11–13] and a gyro-kinetic simulation of ion-temperature-gradient driven turbulence [14] in a fusion relevant geometry. However, in fusion experiments the parallel correlation length has never been measured in the core of the plasmas, consequently only an experimental ‘signature’ of critically balanced turbulence in MAST (Mega Amp Spherical Tokamak) is reported [15]. The parallel correlation length of fluctuations are typically measured using the probes only at the edge of the fusion-grade plasmas [16–18]. As the diagnostic systems being developed, there is a possibility of measuring parallel correlation lengths in the core of KSTAR (Korea Superconducting Tokamak Advanced Research) plasmas with the beam emission spectroscopy (BES) [19] and microwave imaging reflectometry (MIR) [20]. Both systems are 2D (poloidal and radial) but installed at different toroidal locations measuring the same physical quantity - density fluctuations. Again, this motivates us to study the reliability and accuracy of the correlation lengths obtained with the two measurement points.

This paper is organized as follows: in Sec. II, the cor-

* ijwkim@kaist.ac.kr

† ycghim@kaist.ac.kr

relation function and its variance of randomly distributed moving Gaussian-shaped (both in time and space) fluctuations are analytically derived with a finite averaging time window and noise. This expression is, then, compared with the numerically generated synthetic data where we find good agreement between the analytic and numerical results. This correlation function is used to obtain the accuracy (bias error) and reliability (variance) of the estimated correlation length in Sec. III both analytically and numerically whose results are also confirmed quantitatively with the experimental data obtained from the KSTAR BES system. Our conclusions are provided in Sec. IV.

II. CORRELATION FUNCTION OF FLUCTUATING SIGNALS AND ITS VARIANCE

We model a time (t) dependent 1D fluctuating signal at the spatial location $x = x_a$ as the sum of ‘eddies’ :

$$S_a(t) = \sum_{i=1}^N S_{a_i}(t), \quad (1)$$

where N is the total number of eddies and $S_{a_i}(t)$ is the signal due to i^{th} ‘eddy’ at the location $x = x_a$. As has been done previously [21–23] motivated by the experimental observations [24, 25] of ion-scale turbulence in fusion-grade plasmas, we use a Gaussian-shaped structure in both time and space for each eddy:

$$S_{a_i}(t) = A_i \exp \left[-\frac{(t - t_i)^2}{2\tau_{\text{life}}^2} - \frac{(x_a - v(t - t_i) - x_i)^2}{2\lambda_x^2} \right], \quad (2)$$

where the coherent properties of each eddy in time and space are parameterized by the characteristic temporal scale (τ_{life}) and the spatial scale (λ_x). Furthermore, we allow the eddy to move with the speed of v to mimic observed eddy motions due to background plasma flows [22] or wave-like propagations. We let the i^{th} eddy have the maximum amplitude A_i at $x = x_i$ and $t = t_i$ where random number A_i is selected from a normal distribution with zero mean and variance of A^2 . x_i and t_i are selected from uniformly distributed random numbers within the finite domain of $[-\frac{\Delta L}{2}, \frac{\Delta L}{2}]$ and $[-\frac{\Delta T}{2}, \frac{\Delta T}{2}]$, respectively. This means that our result is strictly valid within a flux-surface where the radial gradients of various equilibrium quantities correlated with the turbulence [26] are constant.

A spatio-temporal filling factor F is defined as [22]

$$F \equiv N \left(\frac{\lambda_x}{\Delta L} \right) \left(\frac{\tau_{\text{life}}}{\Delta T} \right), \quad (3)$$

and we control the total number of eddies N such that $F \sim \mathcal{O}(1)$ not to have too frequent or too rare eddies.

Later on, we show that $F \sim \mathcal{O}(1)$ ensures the generated signal to have the square of the fluctuation level indeed of the order of A^2 as specified. More detailed descriptions on the model of the fluctuating signal can be found elsewhere [21, 23].

For the readers who question the validity of the Gaussian-shaped eddies as the Lorentzian eddies are also observed in the scrape-off-layer region of the magnetically confined plasmas [27–29], we provide the discussions with the Lorentzian eddies in Appendix D. This section is recommended to be read after reading Sec. III.

A. Correlation function and its variance

In this section, we analytically derive the correlation value between the two spatial positions at $x = x_a$ and $x = x_b$ at the correlation time delay $\tau = 0$ (as we are interested in obtaining the correlation length of the fluctuating signals) and its associated variance.

To analytically derive the correlation function following Kim et al. [23], ΔL and ΔT are assumed to be infinitely large compared to λ_x and τ_{life} . A correlation value $C_{a,b}$ between $x = x_a$ and x_b is calculated as $E[S_a(t) S_b(t)]$, where $E[\]$ is the time averaging operator over the ‘sub-time window’ whose size is set by ΔT_{sub} . Then, the ensemble averaged correlation value $\langle C_{a,b} \rangle$ is [23]

$$\begin{aligned} \langle C_{a,b} \rangle &\approx \pi A^2 F \left(\exp \left[-\frac{(x_a - x_b)^2}{2(\sqrt{2}\lambda_x)^2} \right] \right. \\ &\quad \left. - 2\sqrt{\pi} \frac{\tau_{\text{ac}}}{\Delta T_{\text{sub}}} \exp \left[-\frac{(x_a - x_b)^2}{4(\lambda_x^2 + \tau_{\text{life}}^2 v^2)} \right] \right), \\ &\approx \pi A^2 F \left(\exp \left[-\frac{(x_a - x_b)^2}{2(\sqrt{2}\lambda_x)^2} \right] \right), \end{aligned} \quad (4)$$

where the approximation in the last line is valid for $\tau_{\text{ac}}/\Delta T_{\text{sub}} \ll 1$. ΔT_{sub} is much smaller than ΔT allowing many of ΔT_{sub} to exist within ΔT (for an ensemble average [23]) but still much larger than τ_{ac} , the usual auto-correlation time of the fluctuating signal in the lab frame defined as [30]

$$\tau_{\text{ac}} \equiv \frac{\lambda_x \tau_{\text{life}}}{\sqrt{\lambda_x^2 + \tau_{\text{life}}^2 v^2}}. \quad (5)$$

Note that the square of the fluctuation level $\langle C_{a,a} \rangle$ is $\sim \mathcal{O}(A^2)$ for the spatio-temporal filling factor $F \sim \mathcal{O}(1)$.

The correlation values, i.e., $C_{a,b}$, have a certain distribution resulting from the fact that the amplitude (A_i) and the spatio-temporal location (x_i and t_i) of the eddies are randomly distributed. This distribution gives us randomness in $C_{a,b}$, i.e., the variance $\sigma_{C_{a,b}}^2$ (see Appendix

A), which is

$$\sigma_{C_{a,b}}^2 \approx \sqrt{2}\pi^{5/2}A^4F^2\frac{\tau_{ac}}{\Delta T_{\text{sub}}}\left(1 + \exp\left[-\frac{(x_a - x_b)^2}{2\lambda_x^2}\right]\right) + \frac{3}{\sqrt{2}}\pi^{3/2}A^4F\frac{\tau_{ac}}{\Delta T_{\text{sub}}}\exp\left[-\frac{(x_a - x_b)^2}{2\lambda_x^2}\right], \quad (6)$$

where the approximation is valid for $\Delta L \gg \lambda_x$, $\Delta T \gg \Delta T_{\text{sub}} \gg \tau_{\text{life}}$ (or τ_{ac}), and $N \gg 1$ with the condition of spatio-temporal filling factor $F \sim \mathcal{O}(1)$ consistent with the assumptions we made to obtain $\langle C_{a,b} \rangle$ in Eq. (4). Note that large N is, in fact, a consequence of having $F \sim \mathcal{O}(1)$ with large ΔL and ΔT .

B. Normalized correlation function and its variance with and without noise

We now calculate the *normalized* correlation functions with and without noise and their associated variances. Again, we are interested in the correlation value between the two spatial points x_a and x_b at the correlation time delay $\tau = 0$. Although the correlation length can be estimated with the ensemble averaged correlation value derived above, we derive the normalized correlation value because the normalized one estimates the correlation length more accurately (smaller bias error) and reliably (smaller variance) compared to the unnormalized one, which we show later in Sec. IIIB.

The ensemble averaged normalized correlation value without noise, $\langle \bar{C}_{a,b} \rangle$, is

$$\begin{aligned} \langle \bar{C}_{a,b} \rangle &\equiv \left\langle \frac{C_{a,b}}{\delta_a \delta_b} \right\rangle \approx \left\langle \frac{C_{a,b}}{\delta_f^2} \right\rangle \\ &\approx \frac{\langle C_{a,b} \rangle}{\langle \delta_f^2 \rangle} = \exp\left[-\frac{(x_a - x_b)^2}{2(\sqrt{2}\lambda_x)^2}\right], \end{aligned} \quad (7)$$

where $\delta_a = \sqrt{C_{a,a}}$ and $\delta_b = \sqrt{C_{b,b}}$ are the fluctuation levels at the locations $x = x_a$ and x_b , respectively. We have denoted this fluctuation level as δ_f and assumed $\delta_a \approx \delta_f$ and $\delta_b \approx \delta_f$ based on the homogeneity of the data. Note that $\langle \delta_f^2 \rangle = \pi A^2 F$ from Eq. (4). The validity of the approximation in the second line of Eq. (7) is provided in Appendix B.

As for the case of $C_{a,b}$, $\bar{C}_{a,b}$ also has a distribution resulting in a finite variance solely due to the randomness of the eddies. This variance denoted as $\sigma_{\bar{C}_{a,b}}^2$ is

$$\sigma_{\bar{C}_{a,b}}^2 \approx \frac{\sigma_{C_{a,b}}^2}{\pi^2 A^4 F^2} \left(1 - \exp\left[-\frac{(x_a - x_b)^2}{2\lambda_x^2}\right]\right)^2. \quad (8)$$

The detailed derivation of $\sigma_{\bar{C}_{a,b}}^2$ and assumptions we made are provided in Appendix C 1.

Let us now consider the signal with uncorrelated noise. Our model signal with noise $S_a^{\text{noise}}(t)$ is, then,

$$S_a^{\text{noise}}(t) = S_a(t) + n_a(t), \quad (9)$$

where $n_a(t)$ is the noise as a function of time at $x = x_a$. A normalized correlation value with noise $\bar{C}_{a,b}^{\text{noise}^*}$ between the signals from two spatial positions x_a and x_b is

$$\bar{C}_{a,b}^{\text{noise}^*} = \frac{\text{E}[S_a(t)S_b(t)] + \text{E}[n_a(t)n_b(t)]}{\sqrt{\text{E}[S_a^2(t)] + \text{E}[n_a^2(t)]}\sqrt{\text{E}[S_b^2(t)] + \text{E}[n_b^2(t)]}}. \quad (10)$$

Note that $\text{E}[n_a(t)n_b(t)] = 0$ unless $x_a = x_b$ due to the uncorrelated noise assumption. The normalized correlation value at $x_a = x_b$ is by definition one whether or not there exists noise, i.e., $\bar{C}_{a,a} = 1$ and $\bar{C}_{a,a}^{\text{noise}^*} = 1$. It is obvious that the effect of $\text{E}[n_a(t)n_b(t)]$ on $\bar{C}_{a,b}^{\text{noise}^*}$ for $x_a = x_b$ can be removed for our model signal, but it is also possible to do so with experimental data by considering the auto-correlation function as a function of the correlation time delay τ [31]. The idea is, basically, interpolating the auto-correlation signal around $\tau = 0$ to eliminate the effect of $\text{E}[n_a(t)n_b(t)]$ present only around very short interval of $\tau = 0$.

Once we remove the noise effect at $x_a = x_b$ both for our model data and experimental data, we can define $\bar{C}_{a,b}^{\text{noise}}$ as

$$\begin{aligned} \bar{C}_{a,b}^{\text{noise}} &= \frac{\text{E}[S_a(t)S_b(t)]}{\sqrt{\text{E}[S_a^2(t)] + \text{E}[n_a^2(t)]}\sqrt{\text{E}[S_b^2(t)] + \text{E}[n_b^2(t)]}} \\ &= \frac{\delta_f^2}{\delta_f^2 + \delta_n^2} \frac{C_{a,b}}{\delta_f^2}, \end{aligned} \quad (11)$$

where we assumed the the noise level δ_n is homogeneous as is the case for δ_f . Then, the ensemble averaged normalized correlation value with noise is

$$\begin{aligned} \langle \bar{C}_{a,b}^{\text{noise}} \rangle &= \left\langle \frac{\delta_f^2}{\delta_f^2 + \delta_n^2} \right\rangle \left\langle \frac{C_{a,b}}{\delta_f^2} \right\rangle \\ &\approx \frac{\langle \delta_f^2 \rangle}{\langle \delta_f^2 + \delta_n^2 \rangle} \langle \bar{C}_{a,b} \rangle \\ &= \frac{\langle \delta_f^2 \rangle}{\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle} \langle \bar{C}_{a,b} \rangle \\ &= \frac{\langle \delta_f^2 \rangle}{\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle} \exp\left[-\frac{(x_a - x_b)^2}{2(\sqrt{2}\lambda_x)^2}\right]. \end{aligned} \quad (12)$$

The first line equality is owing to the fact the normalized correlation value $C_{a,b}/\delta_f^2$ is not correlated with a function of fluctuation levels, i.e., $\delta_f^2/(\delta_f^2 + \delta_n^2)$. The second line approximation is based on the same rationale for $\langle \frac{C_{a,b}}{\delta_f^2} \rangle \approx \frac{\langle C_{a,b} \rangle}{\langle \delta_f^2 \rangle}$ in Eq. (7) (see Appendix B). Notice that the effective role of the noise is to decrease the normalized correlation value, and it is not even unity at $x_a = x_b$, rather it is a function of fluctuation(δ_f)-to-noise(δ_n) ratio (FNR).

The variance $\sigma_{\overline{C}_{a,b}^{\text{noise}}}^2$ of the ensemble averaged normalized correlation value with noise calculated in Appendix C 2 is

$$\sigma_{\overline{C}_{a,b}^{\text{noise}}}^2 \approx \left\langle \overline{C}_{a,b}^{\text{noise}} \right\rangle^2 \left[\left(\frac{\sigma_{\delta_f^2}}{\langle \delta_f^2 \rangle} - \frac{\sigma_{\delta_n^2}}{\langle \delta_n^2 \rangle + \langle \delta_f^2 \rangle} \right)^2 + \frac{\sigma_{\overline{C}_{a,b}}^2}{\langle \overline{C}_{a,b} \rangle^2} \right], \quad (13)$$

where $\sigma_{\delta_f^2}^2$ and $\sigma_{\delta_n^2}^2$ are the variances of δ_f^2 and δ_n^2 , respectively. Since $\delta_f^2 = C_{a,a}$, we have $\sigma_{\delta_f^2}^2 = \sigma_{C_{a,a}}^2$ which can be calculated using Eq. (6).

Note that Eq. (12) and Eq. (13) which are the ensemble averaged normalized correlation value and its variance with noise, respectively, become Eq. (7) and Eq. (8) for $\delta_n^2 = 0$, and none of these equations contain unknowns for our model fluctuation data.

C. Comparisons between analytic expressions and numerical results

To confirm that the analytically derived unnormalized and normalized correlation values and their associated variances are indeed correct, we generate 50 sets of 1D synthetic data based on Eq. (1) and Eq. (2). These sets of synthetic fluctuation data are generated such that they are guaranteed to be homogeneous and stationary [23].

The basic properties of each set of the synthetic data are: the characteristic spatial scale $\lambda_x = 0.3$ m, the characteristic temporal scale $\tau_{\text{life}} = 15$ μs , the velocity of the eddies $v = 5,000$ m/s, the total time window for the data $\Delta T = 15,000$ μs with the sub-time window $\Delta T_{\text{sub}} = 585$ μs (giving 25 sub-time windows) and the spatial domain size $\Delta L = 2.7$ m. To make sure that the synthetic data are homogeneous, δ_f^2 is kept to be constant in space as well as satisfying the criteria on the domain size according to reference [23].

From the synthetic data we estimate the correlation value $\langle C_{a,b} \rangle^*$ (we use a superscript asterisk to denote that it is estimated with the synthetic data) and its standard deviation, i.e., square root of the variance, $\sigma_{C_{a,b}}^*$ as

$$\langle C_{a,b} \rangle^* = \frac{1}{50} \sum_{i=1}^{50} \underbrace{\left[\frac{1}{25} \sum_{j=1}^{25} C_{a,b}^{ij*} \right]}_{\equiv \langle C_{a,b} \rangle^{i*}}, \quad (14)$$

$$\sigma_{C_{a,b}}^* = \frac{1}{50} \sum_{i=1}^{50} \underbrace{\left[\sqrt{\frac{1}{25} \sum_{j=1}^{25} \left(C_{a,b}^{ij*} - \langle C_{a,b} \rangle^{i*} \right)^2} \right]}_{\equiv \sigma_{C_{a,b}}^{i*}},$$

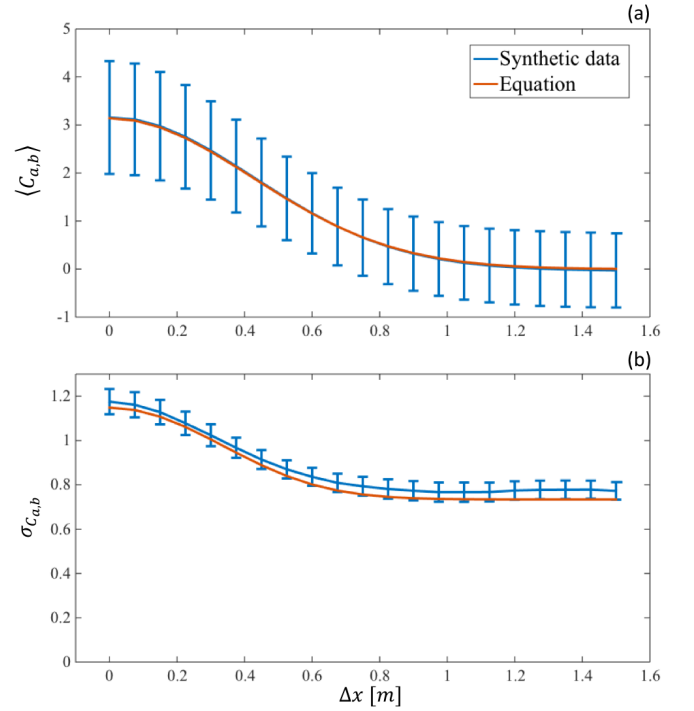


Figure 1. (a) Correlation values and (b) their standard deviations as a function of $\Delta x = |x_a - x_b|$ estimated from the synthetic data, $\langle C_{a,b} \rangle^*$ and $\sigma_{C_{a,b}}^*$ (blue lines), and calculated analytically, $\langle C_{a,b} \rangle$ and $\sigma_{C_{a,b}}$ (red lines). The error bars in (a) represent $\sigma_{C_{a,b}}^*$, and those in (b) show 95% confidence interval.

where the superscripts i and j denote the i^{th} set and the j^{th} sub-time window within a set of the synthetic data, respectively. Here, $C_{a,b}^{ij*}$ is estimated as

$$C_{a,b}^{ij*} = \frac{1}{\Delta T_{\text{sub}}} \int_{\Delta T_{\text{sub}}} S_a^{ij}(t) S_b^{ij}(t) dt. \quad (15)$$

Figure 1(a) shows analytically calculated $\langle C_{a,b} \rangle$ (red line) using Eq. (4) and $\langle C_{a,b} \rangle^*$ (blue line) using Eq. (14) on the synthetic data as a function of $\Delta x = |x_a - x_b|$. They show a good agreement. The error bars on $\langle C_{a,b} \rangle^*$ are $\sigma_{C_{a,b}}^*$. Furthermore, Figure 1(b) illustrates a good agreement within the 95% confidence interval between the analytically calculated $\sigma_{C_{a,b}}$ (red line) using Eq. (6) and estimated $\sigma_{C_{a,b}}^*$ from the synthetic data using Eq. (14) as a function of Δx .

Similar comparisons are shown in Figure 2 for the normalized correlation values with various fluctuation-to-noise ratios (FNR= δ_f/δ_n): FNR = 1.0 (blue), 2.0 (red) and 4.0 (yellow). Normalized correlation values and their standard deviations with noise, $\langle \overline{C}_{a,b}^{\text{noise}} \rangle$ and $\sigma_{\overline{C}_{a,b}^{\text{noise}}}$, are calculated using Eq. (12) and Eq. (13); while they are also obtained from the synthetic data, $\langle \overline{C}_{a,b}^{\text{noise}} \rangle^*$ and $\sigma_{\overline{C}_{a,b}^{\text{noise}}}^*$, using Eq. (14) with normalized $C_{a,b}^{ij*}$. Again,

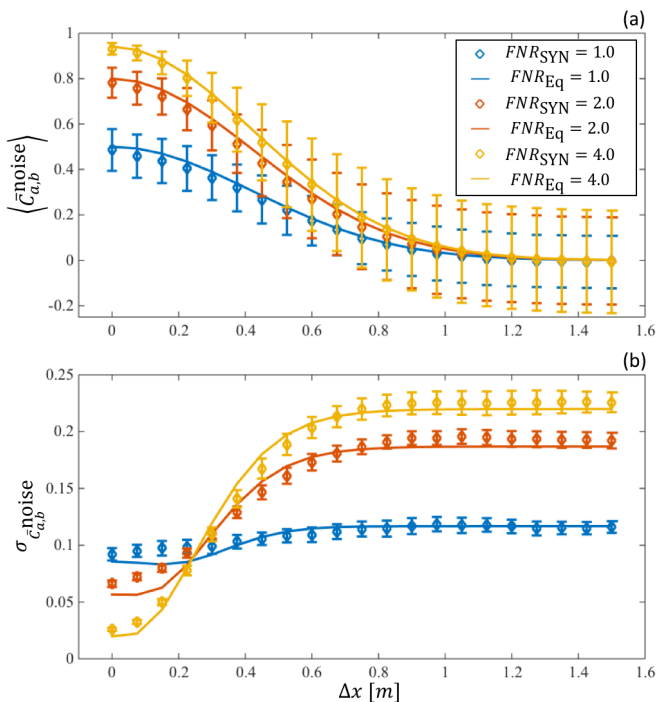


Figure 2. Same as Figure 1 for (a) normalized correlation values and (b) their standard deviations estimated from the synthetic data (diamonds) and calculated analytically (lines) with the fluctuation-to-noise ratio (FNR) of 1.0 (blue), 2.0 (red) and 4.0 (yellow).

we find good agreements between the analytical and numerical results.

Note that the value of $\langle \bar{C}_{a,b}^{\text{noise}} \rangle$ with $x_a = x_b$, i.e., $\Delta x = 0$ is $\langle \delta_f^2 \rangle / (\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle)$ according to Eq. (12). Thus, we expect this value to be 0.50, 0.80 and 0.94 for FNR = 1.0, 2.0 and 4.0, respectively. These values coincide with the results from synthetic data which are obtained by removing the noise in the signal by using the interpolation technique around the correlation time delay $\tau = 0$ of the auto-correlation function [31] as mentioned earlier.

As vindicated by the numerical results, we can use the analytically obtained normalized correlation values with and without noise and their associated variances to examine the accuracy (bias error) and reliability (variance) of the two-point measurement of the spatial correlation length.

III. ACCURACY AND RELIABILITY OF CORRELATION LENGTH MEASUREMENT WITH TWO SPATIAL POINTS

With experimentally obtained fluctuation data one may estimate the normalized auto- and cross-correlation values at the correlation time delay $\tau = 0$ denoted as C_1 and C_2 , respectively; while the true normalized correla-

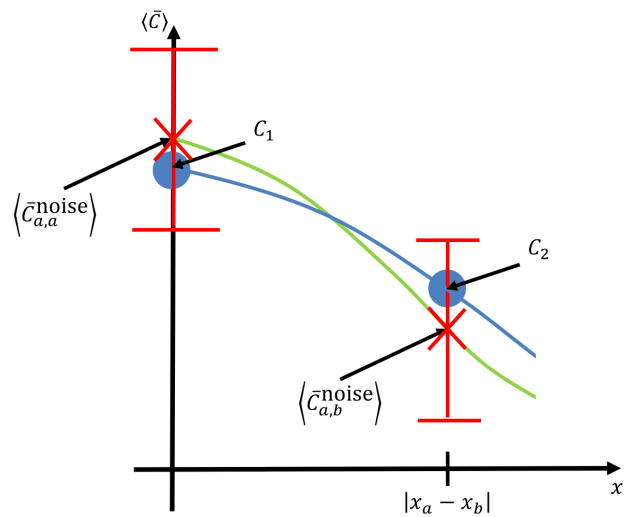


Figure 3. A cartoon picture of the two points fitting. The normalized auto- and cross-correlation values C_1 and C_2 marked as blue filled circles, respectively, may be obtained experimentally; while the true values are $\langle \bar{C}_{a,a}^{\text{noise}} \rangle$ and $\langle \bar{C}_{a,b}^{\text{noise}} \rangle$ marked as red X. Fitting C_1 and C_2 (blue curve) results in overestimation of the true correlation length (green curve) in this case.

tion values are $\langle \bar{C}_{a,a}^{\text{noise}} \rangle$ and $\langle \bar{C}_{a,b}^{\text{noise}} \rangle$ as shown in Figure 3. Here, C_1 is the normalized auto-correlation value after removing the noise by using the technique explained earlier [31]. As we are restricted to use only two spatial points, one may fit any arbitrary functions to the data. We have chosen to fit a Gaussian function to the experimentally obtained values C_1 and C_2 motivated by the experimentally observed ion-scale turbulence in tokamaks [24, 25]:

$$C_2 = C_1 \exp\left(-\frac{(x_a - x_b)^2}{2\ell_x^2}\right), \quad (16)$$

where we have only one unknown ℓ_x which is the *correlation length* that we wish to estimate. The correlation length ℓ_x is $\sqrt{2}$ times larger than the characteristic spatial scale λ_x (cf. Eq. (7) and Eq. (12)). The measured correlation length ℓ_x is, then, estimated to be

$$\ell_x = \begin{cases} \infty & C_1 \leq C_2 \\ \frac{|x_a - x_b|}{\sqrt{2}} \sqrt{\frac{1}{\ln C_1 - \ln C_2}} & C_1 > C_2 > 0 \\ 0 & C_2 \leq 0. \end{cases} \quad (17)$$

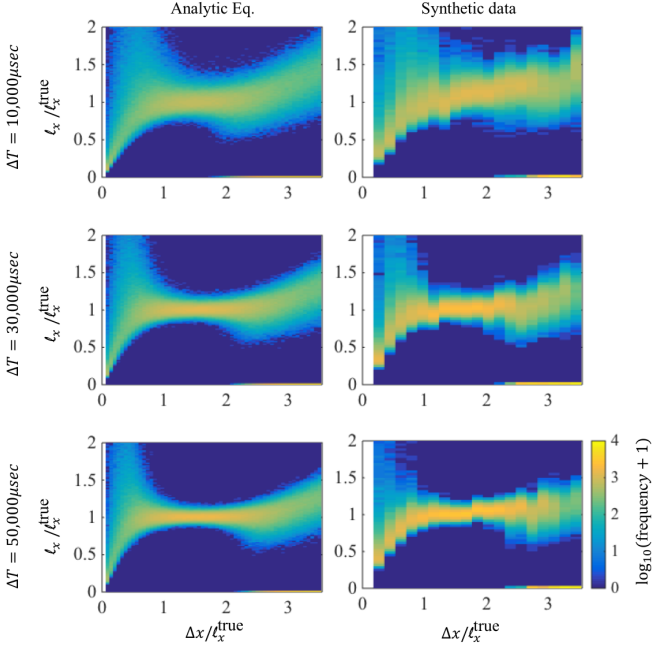


Figure 4. Histograms of the normalized correlation length $\ell_x/\ell_x^{\text{true}}$ as a function of the separation distance $\Delta x/\ell_x^{\text{true}}$ for different sizes of the total time window ΔT . Left panels are obtained from the analytic results and right panels from the synthetic data.

The expected value μ_{ℓ_x} and the variance $\sigma_{\ell_x}^2$ of the measured correlation length ℓ_x are

$$\begin{aligned}\mu_{\ell_x} &= \int \int \ell_x p(C_1) p(C_2) dC_1 dC_2 \\ \sigma_{\ell_x}^2 &= \int \int (\ell_x)^2 p(C_1) p(C_2) dC_1 dC_2 - \mu_{\ell_x}^2,\end{aligned}\quad (18)$$

where $p(C_1)$ and $p(C_2)$ are the probability density functions of obtaining C_1 and C_2 , respectively.

A. Correlation length measurement with two spatial points from the analytic expression and synthetic data

To be able to calculate μ_{ℓ_x} and $\sigma_{\ell_x}^2$, we need to know the probability density function of $\langle \overline{C}_{a,b}^{\text{noise}} \rangle$, i.e., $p(C_1)$ and $p(C_2)$ in Eq. (18). The variance of the ensemble averaged correlation value is $1/M$ times the variance of the correlation value where M is the sample number, i.e., the number of sub-time windows used to measure the ensemble average of the correlation value. Owing to the central limit theorem, this probability density function, $p(C_1)$ or $p(C_2)$, is a normal distribution function with the mean of $\langle \overline{C}_{a,b}^{\text{noise}} \rangle$ and the standard deviation of $\sigma_{\overline{C}_{a,b}^{\text{noise}}}/\sqrt{M}$ which can be estimated either analytically

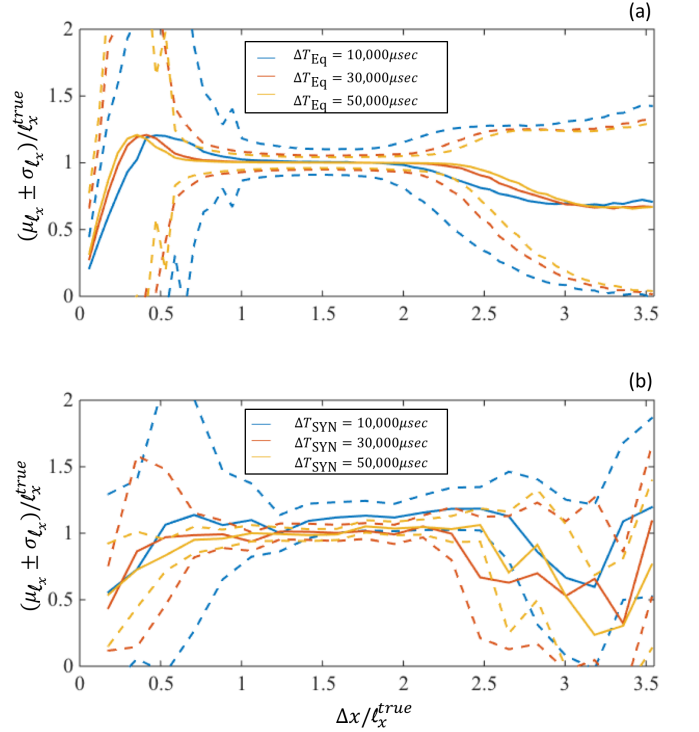


Figure 5. Averages (solid lines) and standard deviations from the averages (dashed lines) of the correlation length normalized to ℓ_x^{true} estimated (a) analytically and (b) numerically as a function of the normalized separation distance $\Delta x/\ell_x^{\text{true}}$ for different sizes of the total time window of $\Delta T = 10,000 \mu\text{s}$ (blue), $\Delta T = 30,000 \mu\text{s}$ (red) and $\Delta T = 50,000 \mu\text{s}$ (yellow).

or numerically with the synthetic data.

Once we have the probability density functions $p(C_1) = P(\langle \overline{C}_{a,a}^{\text{noise}} \rangle)$ and $p(C_2) = P(\langle \overline{C}_{a,b}^{\text{noise}} \rangle)$ either from the analytic results or numerical synthetic data, a Monte-Carlo method by generating 10,000 samples from these probability density functions is used to generate a histogram of the correlation length normalized to the true correlation length $\ell_x^{\text{true}} = \sqrt{2}\lambda_x$ as a function of the separation distance between the two measured positions Δx normalized to the true correlation length ℓ_x^{true} as well.

This histogram is, then, used to estimate the mean μ_{ℓ_x} and the variance $\sigma_{\ell_x}^2$ of the two-point correlation length measurements. Note that we ignore the case of $C_1 \leq C_2$ which corresponds to $\ell_x = \infty$. Nevertheless, if $C_1 \leq C_2$ occurs frequently, indicating that the correlation length ℓ_x is much larger than the separation distance of the two measurement positions, such a signature will be shown as a large variance according to Eq. (17) and Eq. (18). Here, we examine two major factors that affect the quality of the correlation length measurements: 1) size of the total time window and 2) noise level.

To quantify how the size of the total time window affects the measured correlation length, we generate three sets of synthetic data with $\lambda_x = 0.3 \text{ m}$, $\tau_{\text{life}} = 15 \mu\text{s}$, $v = 5,000 \text{ m/s}$ and $\Delta T_{\text{sub}} = 585 \mu\text{s}$ at the fixed FNR

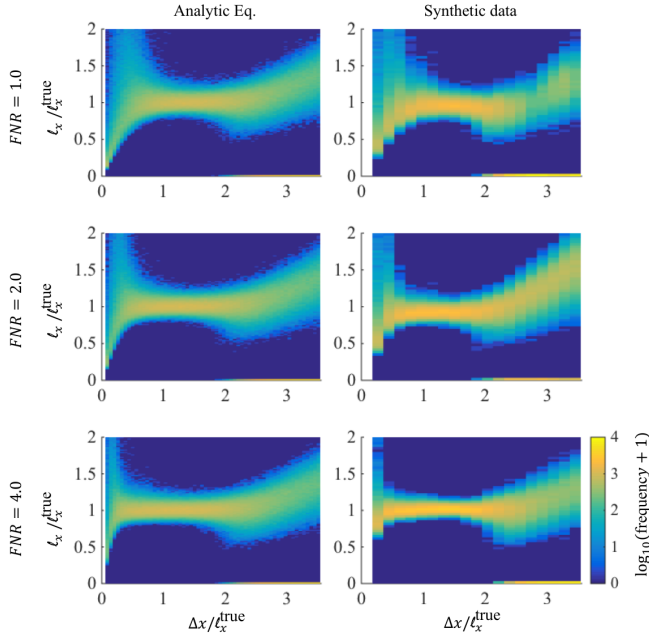


Figure 6. Same as Figure 4 for a fixed $\Delta T = 15,000 \mu\text{s}$ with the different values of FNR: 1.0 (top), 2.0 (middle) and 4.0 (bottom).

of 0.5, whereas ΔT are set to be $10,000 \mu\text{s}$, $30,000 \mu\text{s}$ and $50,000 \mu\text{s}$. Figure 4 shows the histograms of the obtained correlation length with the $p(C_1)$ and $p(C_2)$ from the analytic results (left panels) and the synthetic data (right panels). Based on these histograms, we finally obtain the average and standard deviation of the correlation length (normalized to ℓ_x^{true}) as a function of $\Delta x/\ell_x^{\text{true}}$ as shown in Figure 5. Although Figure 5 is more succinct than Figure 4, it is better to consider Figure 4 because the first (μ_{ℓ_x}) and second ($\sigma_{\ell_x}^2$) moments become less representative as the histogram deviates from a normal distribution, especially for small and large $\Delta x/\ell_x^{\text{true}}$.

The similar trend between the analytic and numerical results shows that increasing the size of the total window decreases the standard deviation of the measurements, and the two-point correlation length measurement becomes less reliable if the separation distance between the two measured points are either smaller than or more than two times larger than the actual correlation length of the fluctuation data.

Next, we examine the effect of noise level, i.e., the fluctuation-to-noise ratio (FNR), on the two-point correlation length measurements. Three sets of synthetic data with $\lambda_x = 0.3 \text{ m}$, $\tau_{\text{ife}} = 15 \mu\text{s}$, $v = 5,000 \text{ m/s}$, $\Delta T = 15,000 \mu\text{s}$ and $\Delta T_{\text{sub}} = 585 \mu\text{s}$ are generated with different values of FNR: 1.0, 2.0 and 4.0. Figure 6 shows the histograms of the obtained correlation length using the analytic results (left panel) and synthetic data (right panel), whereas the average and standard deviation of the correlation length is shown in Figure 7.

We find that the lower limit on the $\Delta x/\ell_x^{\text{true}}$ is dependent on the FNR; while the upper limit is not. This

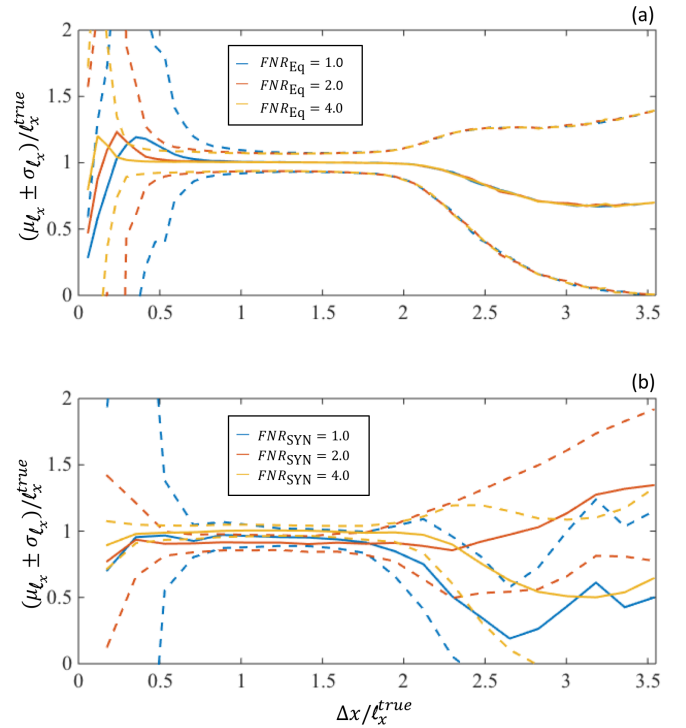


Figure 7. Same as Figure 5 for a fixed $\Delta T = 15,000 \mu\text{s}$ with the different values of FNR: 1.0 (blue), 2.0 (red) and 4.0 (yellow).

observation can be explained based on Figure 2. We see that the standard deviation levels are larger at small Δx for the smaller FNR. This results in a larger value on the lower limit of $\Delta x/\ell_x^{\text{true}}$ with the smaller FNR. In addition, larger standard deviation at the small $\Delta x/\ell_x^{\text{true}}$ is more probable to have $C_1 \leq C_2$ in Eq. (17). Figure 2 also shows that although the standard deviation levels are larger with the larger FNR at large Δx , they are smaller at $\Delta x = 0$. Therefore, we observe similar upper limit on $\Delta x/\ell_x^{\text{true}}$ for the investigated FNR cases.

B. Unnormalized vs. normalized correlation function for the correlation length measurement

The variance of the correlation length $\sigma_{\ell_x}^2$ is a function of $p(C_1)$ and $p(C_2)$ as shown in Eq. (18), i.e., the narrower the $p(C_1)$ and $p(C_2)$, the smaller the variance. Considering an extreme case of absolutely no noise, we see that $p(C_1)$ is a delta function, i.e., no variance, for the normalized correlation function (see Eq. (8)) while it always has a non-zero width, i.e., non-zero variance, for the unnormalized correlation function (see Eq. (6)). Figure 8 shows the unnormalized (blue) and normalized correlation functions with no noise (red) and FNR=4.0 (yellow) as a function of $\Delta x/\ell_x^{\text{true}}$ calculated with the analytic results. Here, the error bars represent the standard deviation of the mean, i.e., $1/\sqrt{M}$ times the standard

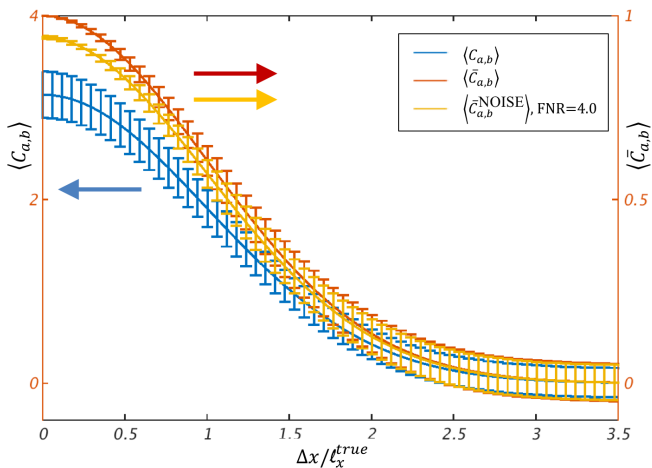


Figure 8. Unnormalized (blue) and normalized correlation functions with no noise (red) and noise level of FNR=4.0 (yellow) as a function of $\Delta x/\ell_x^{\text{true}}$ calculated with the analytic results. As the $\Delta x/\ell_x^{\text{true}}$ approaches to zero, error bars become smaller for the normalized correlation functions than the unnormalized ones. Here, the error bars represent the standard deviation of the mean, i.e., $1/\sqrt{M}$ times the standard deviation of the correlation value.

deviation of the correlation value; thus, they provide approximate widths of $p(C_1)$ and $p(C_2)$. Direct comparison of the widths of $p(C_2)$ between unnormalized and normalized correlation functions can be tricky as they are functions of Δx , but the very small width of the $p(C_1)$ at $\Delta x = 0$ for the normalized correlation function, in general, provides better correlation length measurements compared to the unnormalized correlation function.

Figure 9 shows the average and standard deviation of the correlation length estimated with the unnormalized (blue) and normalized correlation functions without noise (red) and with the noise level of FNR=4.0 (yellow) with the analytic results as in Figure 8. We see that the normalized correlation function provides smaller standard deviation and bias error of the measured correlation length especially for $\Delta x \leq \ell_x^{\text{true}}$.

C. Experimental data: two point correlation length measurement for real density fluctuations

Reliability and accuracy of the two-point correlation length measurement have been examined analytically and numerically. We now apply the two-point correlation length measurement on the density fluctuation data experimentally obtained via the BES system [19] installed in KSTAR.

Since the ‘true’ correlation length of the fluctuation data are not available, we estimate the correlation length using the four points (poloidally aligned four detectors) denoted as $\ell_{\theta(4 \text{ points})}$ and treat this value as the ‘true’ correlation length. Figure 10 shows an example of four-point measurement of the correlation values (blue) at the

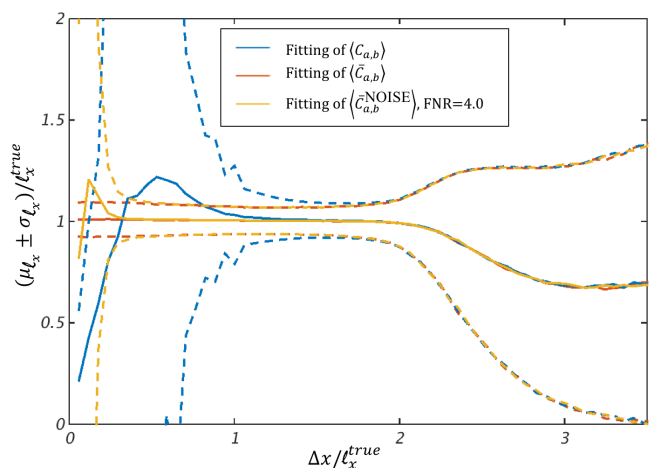


Figure 9. The two point correlation length measurement using normalized correlation function (blue line) and unnormalized correlation function (red line) when the noise level is zero. Solid lines show the mean and dashed lines show the standard deviation.

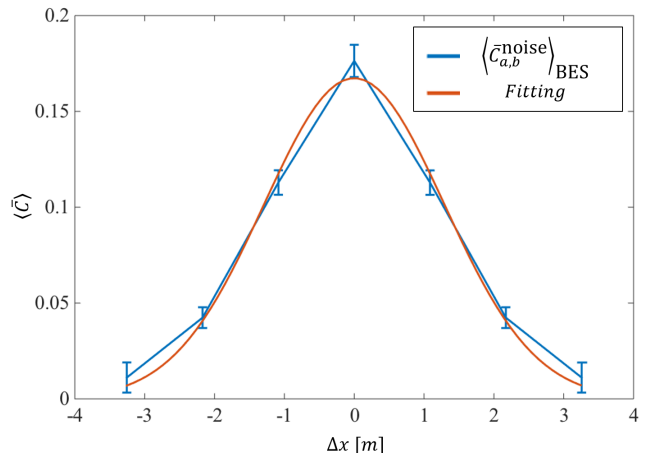


Figure 10. An example of a Gaussian fitting (red) to the correlation values (blue) from poloidally aligned 4 detectors for KSTAR shot #9133. Correlation values for $\Delta x < 0$ are copies of the values from $\Delta x > 0$ to aid visualization. This Gaussian fitting provides the ‘true’ poloidal correlation length $\ell_{\theta(4 \text{ points})}$.

time delay $\tau = 0$. Note that correlation values for $\Delta x < 0$ are just copies of the values from $\Delta x > 0$ to aid visualization. Gaussian fitting (red) to the correlation values provides the correlation length $\ell_{\theta(4 \text{ points})}$.

Then, we estimate the correlation length using only two measurement points denoted as $\ell_{\theta(2 \text{ points})}$. Figure 11 shows $\ell_{\theta(2 \text{ points})}/\ell_{\theta(4 \text{ points})}$ as a function of $\Delta x/\ell_{\theta(4 \text{ points})}$ for different sizes of the total time window $\Delta T = 10$ ms (blue), 30 ms (red) and 50 ms (yellow). The result is quantitatively similar to Figure 4 (or Figure 5), i.e., results obtained analytically and numerically: two-point correlation length measurements are not biased for

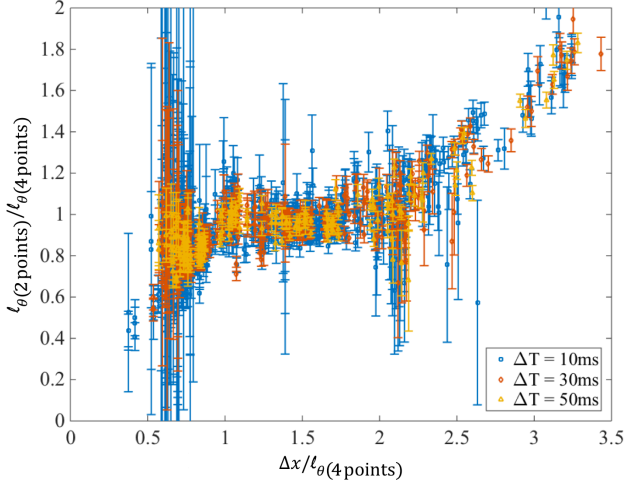


Figure 11. Two-point correlation length measurement $\ell_{\theta(2 \text{ points})}$ normalized to four-point correlation length measurement $\ell_{\theta(4 \text{ points})}$ as a function of the separation distance $\Delta x/\ell_{\theta(4 \text{ points})}$ from the KSTAR BES data for shot #9133. The result is quantitatively similar to Figure 4 (or Figure 5).

$1 \leq \Delta x/\ell_{\theta(4 \text{ points})} \leq 2$, and a larger size of total time window provides smaller variance.

IV. CONCLUSION

We have analytically derived the normalized and unnormalized correlation functions of Gaussian shaped moving fluctuation data as well as their associated variances. The analytic results are found to have good agree-

ment with the numerical results. Based on the correlation functions, the reliability and accuracy of the two-point correlation length measurement have been examined, where we have found that the separation distance between the two measurements points needs to be within one and two times of the correlation length to obtain reliable and accurate true results. The lower limit is found to be dependent on the ratio of fluctuation to noise. This indicates that if an obtained correlation length from two-point measurements are much larger or smaller than the separation distance, validity of the obtained correlation length must be questioned. Our results can also be used as design criteria when one builds a diagnostic system measuring fluctuations of any physical quantities with an aim of obtaining correlation lengths.

ACKNOWLEDGMENTS

This work is supported by National R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (grant number 2014M1A7A1A01029835) and the KUSTAR-KAIST Institute, KAIST, Korea.

Appendix A: Variance of the correlation value

The variance of the estimated correlation value $\sigma_{C_{a,b}}^2$ by definition is

$$\sigma_{C_{a,b}}^2 = \langle C_{a,b}^2 \rangle - \langle C_{a,b} \rangle^2. \quad (\text{A1})$$

By invoking a similar approach done by Kim et al. [23], we find

$$\begin{aligned} \sigma_{C_{a,b}}^2 \approx & -NA^4\pi^2 \frac{\lambda_x^2}{\Delta L^2} \frac{\tau_{\text{life}}^2}{\Delta T^2} \exp\left[-\frac{(x_a - x_b)^2}{2\lambda_x^2}\right] \\ & -4N(N-1)A^4\pi^3 \frac{\lambda_x^2}{\Delta L^2} \frac{\tau_{\text{life}}^2}{\Delta T^2} \frac{\tau_{\text{ac}}^2}{\Delta T_{\text{sub}}^2} \\ & -4N^2A^4\pi^3 \frac{\lambda_x^2}{\Delta L^2} \frac{\tau_{\text{life}}^2}{\Delta T^2} \frac{\tau_{\text{ac}}^2}{\Delta T_{\text{sub}}^2} \exp\left[-\frac{(x_a - x_b)^2}{2(\lambda_x^2 + \tau_{\text{life}}^2 v^2)}\right] \\ & +\sqrt{2}N(N-1)A^4\pi^{5/2} \frac{\lambda_x^2}{\Delta L^2} \frac{\tau_{\text{life}}^2}{\Delta T^2} \frac{\tau_{\text{ac}}}{\Delta T_{\text{sub}}} \left(1 + \exp\left[-\frac{(x_a - x_b)^2}{2\lambda_x^2}\right]\right) \\ & +4NA^4\pi^{5/2} \frac{\lambda_x^2}{\Delta L^2} \frac{\tau_{\text{life}}^2}{\Delta T^2} \frac{\tau_{\text{ac}}}{\Delta T_{\text{sub}}} \exp\left[-\frac{(2\lambda_x^2 + \tau_{\text{life}}^2 v^2)(x_a - x_b)^2}{4\lambda_x^2(\lambda_x^2 + \tau_{\text{life}}^2 v^2)}\right] \\ & +6\sqrt{2}NA^4\pi^{5/2} \frac{\lambda_x}{\Delta L} \frac{\tau_{\text{life}}}{\Delta T} \frac{\tau_{\text{ac}}^3}{\Delta T_{\text{sub}}^3} \exp\left[-\frac{(x_a - x_b)^2}{4(\lambda_x^2 + \tau_{\text{life}}^2 v^2)}\right] \\ & -6\sqrt{2}NA^4\pi^2 \frac{\lambda_x}{\Delta L} \frac{\tau_{\text{life}}}{\Delta T} \frac{\tau_{\text{ac}}^2}{\Delta T_{\text{sub}}^2} \exp\left[-\frac{(2\lambda_x^2 + \tau_{\text{life}}^2 v^2)(x_a - x_b)^2}{4\lambda_x^2(\lambda_x^2 + \tau_{\text{life}}^2 v^2)}\right] \\ & +\frac{3}{\sqrt{2}}NA^4\pi^{3/2} \frac{\lambda_x}{\Delta L} \frac{\tau_{\text{life}}}{\Delta T} \frac{\tau_{\text{ac}}}{\Delta T_{\text{sub}}} \exp\left[-\frac{(x_a - x_b)^2}{2\lambda_x^2}\right], \end{aligned} \quad (\text{A2})$$

where the approximation is for $\Delta L \gg \lambda_x$, $\Delta T \gg \Delta T_{\text{sub}} \gg \tau_{\text{life}}$ (or τ_{ac}) consistent with the conditions on

obtaining $\langle C_{a,b} \rangle$ in Eq. (4). Then, having the spatio-

temporal filling factor $F \sim \mathcal{O}(1)$ and assuming $N \gg 1$ which is well satisfied with large ΔL and ΔT , all the terms are negligible compared to the fourth and the eighth terms on the right-hand-side of Eq. (A2). Note that $N(N-1)$ is approximated to N^2 in the fourth term to get Eq. (6).

Appendix B: Ensemble average approximation of the normalized correlation function

Here, we explain the rationale and the assumptions we have made in $\left\langle \frac{C_{a,b}}{\delta_f^2} \right\rangle \approx \frac{\langle C_{a,b} \rangle}{\langle \delta_f^2 \rangle}$ which is the approximation in the second line of the right-hand-side of Eq. (7).

Suppose there are two random variables x and y as the realizations of X and Y , respectively. For any $f(x, y)$, the bivariate first order Taylor expansion about (x_1, y_1) is

$$\begin{aligned} f(x, y) &= f(x_1, y_1) \\ &+ \left. \frac{\partial f(x, y)}{\partial x} \right|_{(x_1, y_1)} (x - x_1) \\ &+ \left. \frac{\partial f(x, y)}{\partial y} \right|_{(x_1, y_1)} (y - y_1) \\ &+ \text{remainder}. \end{aligned} \quad (\text{B1})$$

Applying Eq. (B1) with $f(x, y) = x/y$ for $x = C_{a,b}$ and $y = \delta_f^2$, the ensemble average of $f(x, y)$ about $x_1 = \langle X \rangle$ and $y_1 = \langle Y \rangle$ is

$$\begin{aligned} \left\langle \frac{C_{a,b}}{\delta_f^2} \right\rangle &= \langle f(x, y) \rangle \\ &= \langle f(\langle X \rangle, \langle Y \rangle) \rangle \\ &+ \left. \frac{\partial f(x, y)}{\partial x} \right|_{(\langle X \rangle, \langle Y \rangle)} \underbrace{\langle x - \langle X \rangle \rangle}_{=0} \\ &+ \left. \frac{\partial f(x, y)}{\partial y} \right|_{(\langle X \rangle, \langle Y \rangle)} \underbrace{\langle y - \langle Y \rangle \rangle}_{=0} \\ &+ \text{remainder} \\ &\approx \left\langle \frac{\langle C_{a,b} \rangle}{\langle \delta_f^2 \rangle} \right\rangle = \frac{\langle C_{a,b} \rangle}{\langle \delta_f^2 \rangle}. \end{aligned} \quad (\text{B2})$$

Ignoring the remainder, i.e., the higher order terms, in the last line is valid when the standard deviation of a random variable (σ_y) is small compared to its mean (or expectation) value ($\langle Y \rangle$) [32]. Note that the higher order terms in x , i.e., $\partial^n / \partial x^n$ for $n \geq 2$, are zeros for $f(x, y) = x/y$. Thus, the ensemble average approximation $\left\langle \frac{C_{a,b}}{\delta_f^2} \right\rangle \approx \frac{\langle C_{a,b} \rangle}{\langle \delta_f^2 \rangle}$ is valid for $\sigma_{C_{a,b}} = \sigma_{\delta_f^2} \ll \langle \delta_f^2 \rangle = \langle C_{a,b} \rangle$ which is satisfied for $\tau_{\text{life}} \ll \Delta T_{\text{sub}}$ according to Eq. (4) and Eq. (6).

Appendix C: Variance of the normalized correlation value with and without noise

1. Without noise

To derive the variance of the normalized correlation value $\sigma_{\frac{C_{a,b}}{\delta_f^2}}^2$, we invoke an approximated form of the variance of products [33] for $g = A/B$, namely

$$\sigma_g^2 \approx g^2 \left[\left(\frac{\sigma_A}{A} \right)^2 + \left(\frac{\sigma_B}{B} \right)^2 - 2\rho_{AB} \frac{\sigma_A \sigma_B}{AB} \right], \quad (\text{C1})$$

where σ_g , σ_A and σ_B are the standard deviations of g , A and B , respectively. ρ_{AB} is the normalized correlation between A and B . For our case, $g = \langle C_{a,b} \rangle$, $A = \langle C_{a,b} \rangle$ and $B = \langle \delta_a \delta_b \rangle = \langle \delta_f^2 \rangle$ with $\sigma_A^2 = \sigma_{C_{a,b}}^2$ and $\sigma_B^2 = \sigma_{\delta_f^2}^2$. A difficulty to derive $\sigma_{\frac{C_{a,b}}{\delta_f^2}}^2$ arises due to the correlation between $C_{a,b}$ and δ_f^2 , i.e., ρ_{AB} . For this reason, we take two extreme cases where $\rho_{AB} \approx 1$ and $\rho_{AB} \approx 0$, and connect them smoothly based on an educated guess. Note that we confirm numerically in Sec. II C that our result is indeed valid.

For $x_a = x_b$, we know that $C_{a,b} = C_{a,a} = \delta_f^2$, i.e., $A = B$, resulting in $\rho_{AB} = 1$. This simply gives $\sigma_g^2 = 0$, i.e., $\sigma_{\frac{C_{a,b}}{\delta_f^2}}^2 = 0$. This result is what we expect since we know that the normalized auto-correlation at the time delay $\tau = 0$ is unity by definition, and there is no variance associated with it if the signal contains no noise.

The opposite extreme case is obtained by considering two spatial positions where $|x_a - x_b| \gg \lambda_x$ such that there is absolutely no correlation between the signals at x_a and x_b , i.e., $C_{a,b} \approx 0$. Thus, there is no correlation between $C_{a,b}$ and δ_f^2 , resulting in $\rho_{AB} \approx 0$. For this case, $\sigma_{\frac{C_{a,b}}{\delta_f^2}}^2$ becomes

$$\begin{aligned} \sigma_{\frac{C_{a,b}}{\delta_f^2}}^2 &\approx \left(\frac{\langle C_{a,b} \rangle}{\langle \delta_f^2 \rangle} \right)^2 \left[\left(\frac{\sigma_{C_{a,b}}}{\langle C_{a,b} \rangle} \right)^2 + \left(\frac{\sigma_{(\delta_f^2)}}{\langle \delta_f^2 \rangle} \right)^2 \right] \\ &\approx \left(\frac{\sigma_{C_{a,b}}}{\langle \delta_f^2 \rangle} \right)^2 = \frac{\sigma_{C_{a,b}}^2}{\pi^2 A^4 F^2}, \end{aligned} \quad (\text{C2})$$

where we have ignored the second term inside the square bracket for very small $\langle C_{a,b} \rangle$ to get the last line and used the fact that $\langle \delta_f^2 \rangle = \pi A^2 F$.

With these two results on $\sigma_{\frac{C_{a,b}}{\delta_f^2}}^2$ for $|x_a - x_b| = 0$ and $|x_a - x_b| \gg \lambda_x$, we connect them with a Gaussian shape motivated by the form of variance for the unnormalized correlation value $\sigma_{C_{a,b}}^2$ as in Eq. (6). Thus, we finally obtain the variance of the normalized correlation value as:

$$\sigma_{\frac{C_{a,b}}{\delta_f^2}}^2 \approx \frac{\sigma_{C_{a,b}}^2}{\pi^2 A^4 F^2} \left(1 - \exp \left[-\frac{(x_a - x_b)^2}{2\lambda_x^2} \right] \right)^2, \quad (\text{C3})$$

which is what we have in Eq. (8).

2. With noise

To obtain the variance of the normalized correlation value with noise $\sigma_{\overline{C}_{a,b}}^{\text{noise}}$, we again use an approximated form of the variance of products [33] for $g = AB$ which is

$$\sigma_g^2 \approx g^2 \left[\left(\frac{\sigma_A}{A} \right)^2 + \left(\frac{\sigma_B}{B} \right)^2 + 2\rho_{AB} \frac{\sigma_A \sigma_B}{AB} \right], \quad (\text{C4})$$

where $g = \langle \overline{C}_{a,b}^{\text{noise}} \rangle$, $A = \langle \delta_f^2 \rangle / \langle \delta_f^2 + \delta_n^2 \rangle$ and $B = \langle \overline{C}_{a,b} \rangle$ with $\sigma_g^2 = \sigma_{\overline{C}_{a,b}}^{\text{noise}}$, $\sigma_A^2 = \sigma_{\delta_f^2 / (\delta_f^2 + \delta_n^2)}^2$ and $\sigma_B^2 = \sigma_{\overline{C}_{a,b}}^2$. ρ_{AB} is the normalized correlation between $\delta_f^2 / (\delta_f^2 + \delta_n^2)$ and $\overline{C}_{a,b}$ which is zero because the normalized correlation value $\overline{C}_{a,b}$ is independent of the fluctuation level.

Since $\sigma_{\delta_f^2 / (\delta_f^2 + \delta_n^2)}^2$ is unknown, we calculate it using Eq. (C1) where the variances of δ_f^2 and δ_n^2 are $\sigma_{\delta_f^2}^2$ and $\sigma_{\delta_n^2}^2$, respectively. Then, the variance of $\delta_f^2 + \delta_n^2$ denoted as $\sigma_{\delta_f^2 + \delta_n^2}^2$ is $\sigma_{\delta_f^2}^2 + \sigma_{\delta_n^2}^2$ because δ_f^2 and δ_n^2 are uncorrelated. Furthermore, the normalized correlation between $\langle \delta_f^2 \rangle$ and $\langle \delta_f^2 + \delta_n^2 \rangle$ are approximately unity since δ_n^2 is a constant as long as the sub-time window size ΔT_{sub} is much larger than the auto-correlation time of the noise which is a sampling time. Thus, $\sigma_{\delta_f^2 / (\delta_f^2 + \delta_n^2)}^2$ is

$$\begin{aligned} \sigma_{\delta_f^2 / (\delta_f^2 + \delta_n^2)}^2 &\approx \left(\frac{\langle \delta_f^2 \rangle}{\langle \delta_f^2 + \delta_n^2 \rangle} \right)^2 \\ &\left[\frac{\sigma_{\delta_f^2}^2}{\langle \delta_f^2 \rangle^2} + \frac{\sigma_{\delta_f^2 + \delta_n^2}^2}{\langle \delta_f^2 + \delta_n^2 \rangle^2} - 2 \frac{\sigma_{\delta_f^2} \sigma_{\delta_f^2 + \delta_n^2}}{\langle \delta_f^2 \rangle \langle \delta_f^2 + \delta_n^2 \rangle} \right] \\ &= \left(\frac{\langle \delta_f^2 \rangle}{\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle} \right)^2 \left[\frac{\sigma_{\delta_f^2}}{\langle \delta_f^2 \rangle} - \frac{\sqrt{\sigma_{\delta_f^2}^2 + \sigma_{\delta_n^2}^2}}{\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle} \right]^2. \end{aligned} \quad (\text{C5})$$

We, then, obtain $\sigma_{\overline{C}_{a,b}}^{\text{noise}}$ as

$$\begin{aligned} \sigma_{\overline{C}_{a,b}}^{\text{noise}} &\approx \left(\frac{\langle \delta_f^2 \rangle}{\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle} \langle \overline{C}_{a,b} \rangle \right)^2 \times \\ &\left[\left(\frac{\sigma_{\delta_f^2}}{\langle \delta_f^2 \rangle} - \frac{\sqrt{\sigma_{\delta_f^2}^2 + \sigma_{\delta_n^2}^2}}{\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle} \right)^2 + \frac{\sigma_{\overline{C}_{a,b}}^2}{\langle \overline{C}_{a,b} \rangle^2} \right] \\ &\approx \langle \overline{C}_{a,b}^{\text{noise}} \rangle^2 \\ &\left[\left(\frac{\sigma_{\delta_f^2}}{\langle \delta_f^2 \rangle} - \frac{\sigma_{\delta_f^2}}{\langle \delta_f^2 \rangle + \langle \delta_n^2 \rangle} \right)^2 + \frac{\sigma_{\overline{C}_{a,b}}^2}{\langle \overline{C}_{a,b} \rangle^2} \right], \end{aligned} \quad (\text{C6})$$

where we assume that $\sigma_{\delta_f^2}^2 \gg \sigma_{\delta_n^2}^2$ in the last line approximation since δ_n^2 is almost a constant. This is the Eq. (13). Figure 2 shows how our analytical result agrees with the numerically estimated variances using the synthetic data including the noise effects.

Appendix D: The two-point measurement of the correlation length for the Lorentzian-shaped eddies

As the Lorentzian-shaped eddies at the edge (SOL) of the magnetically confined plasmas are observed [27–29], we discuss accuracy and reliability of the two-point measurement of the correlation length for the Lorentzian-shaped eddies in this section. Fluctuating signals are the sum of eddies as in Eq. (1) with moving Lorentzian eddies:

$$\begin{aligned} S_{a_i}(t) &= A_i \frac{1}{\left(\frac{t - t_i}{\tau_{\text{life}}} \right)^2 + 1} \\ &\times \frac{1}{\left(\frac{x_a - v(t - t_i)^2 - x_i}{\lambda_x} \right)^2 + 1}, \end{aligned} \quad (\text{D1})$$

where A_i is selected from a normal distribution; while x_i and t_i are generated from uniform distributions.

Once we have the synthetic data with Lorentzian eddies, we follow similar steps as in Sec. III with the Lorentzian fitting function:

$$C_2 = C_1 \frac{1}{\left(\frac{x_a - x_b}{\ell_x} \right)^2 + 1}, \quad (\text{D2})$$

and the measured correlation length (ℓ_x) is estimated to

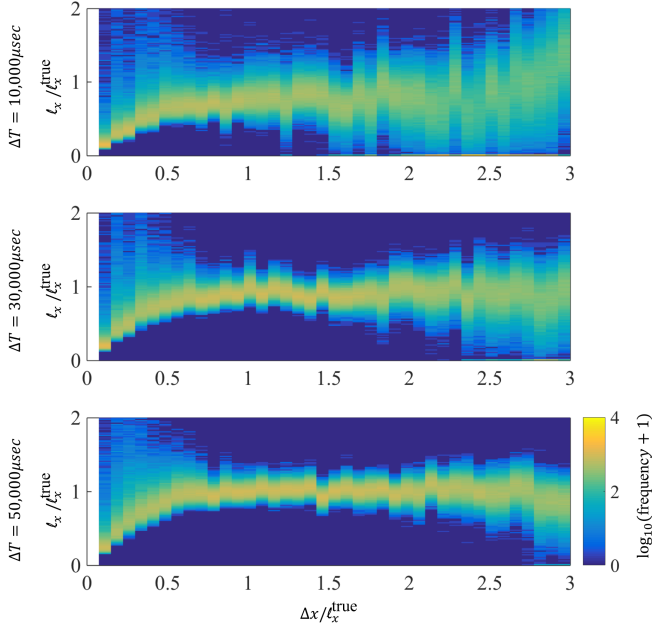


Figure 12. Histograms of the normalized correlation length $\ell_x / \ell_x^{\text{true}}$ as a function of the separation distance $\Delta x / \ell_x^{\text{true}}$ for different sizes of the total time window ΔT . They are obtained from the synthetic data.

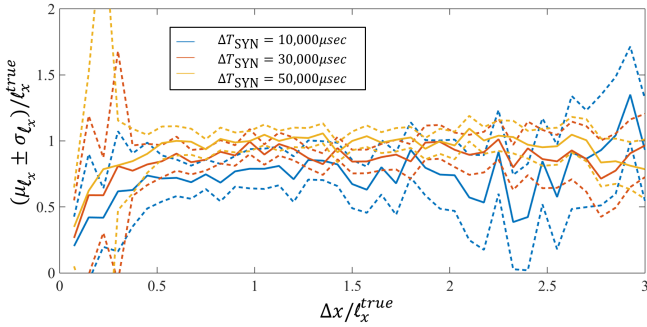


Figure 13. Averages (solid lines) and standard deviations (dashed lines) of the correlation length normalized to ℓ_x^{true} estimated numerically as a function of the normalized separation distance $\Delta x / \ell_x^{\text{true}}$ for different sizes of the total time window of $\Delta T = 10,000 \mu\text{s}$ (blue), $\Delta T = 30,000 \mu\text{s}$ (red) and $\Delta T = 50,000 \mu\text{s}$ (yellow).

be

$$\ell_x = \begin{cases} \infty & C_1 \leq C_2 \\ |x_a - x_b| \sqrt{\frac{C_2}{C_1 - C_2}} & C_1 > C_2 > 0 \\ 0 & C_2 \leq 0. \end{cases} \quad (\text{D3})$$

Note that the covariance function of the Lorentzian function shows that the correlation length (ℓ_x) is twice as

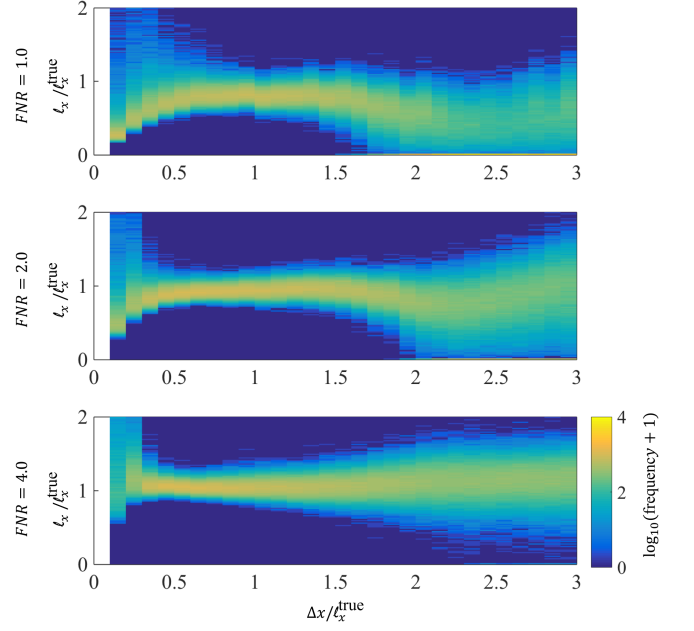


Figure 14. Same as Figure 12 for a fixed $\Delta T = 15,000 \mu\text{s}$ with the different values of FNR: 1.0 (top), 2.0 (middle) and 4.0 (bottom).

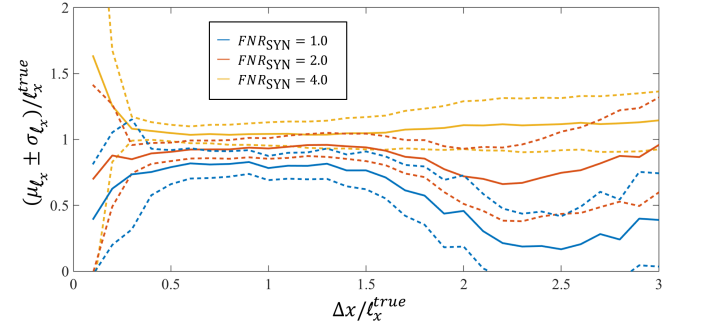


Figure 15. Same as Figure 13 for a fixed $\Delta T = 15,000 \mu\text{s}$ with the different values of FNR: 1.0 (blue), 2.0 (red) and 4.0 (yellow).

large as the characteristic length (λ_x), i.e., $\ell_x^{\text{true}} = 2\lambda_x$:

$$\int_{-\infty}^{\infty} \frac{1}{\left(\frac{x'}{\lambda_x}\right)^2 + 1} \frac{1}{\left(\frac{x-x'}{\lambda_x}\right)^2 + 1} dx' = \frac{\pi\lambda_x}{2} \frac{1}{\left(\frac{x}{2\lambda_x}\right)^2 + 1}. \quad (\text{D4})$$

We obtain the expected value μ_{ℓ_x} and the variance $\sigma_{\ell_x}^2$ using Eq. (18). The probability density functions $p(C_1) = P(\langle \overline{C}_{a,a}^{\text{noise}} \rangle)$ and $p(C_2) = P(\langle \overline{C}_{a,b}^{\text{noise}} \rangle)$ are obtained from the synthetic data, and a Monte-Carlo method is also used to generate histograms similar to Figure 4 and Figure 6. Again, we investigate the accuracy and reliability of the two-point measurement for different sizes of the total time window and different values of FNR.

To investigate how the size of the total time window

affects the measured correlation length, we generate three sets of synthetic data with $\lambda_x = 0.3$ m, $\tau_{\text{life}} = 15$ μs , $v = 5,000$ m/s and $\Delta T_{\text{sub}} = 585$ μs at the fixed FNR of 0.5, whereas ΔT are set to be 10,000 μs , 30,000 μs and 50,000 μs . Note that these values are the same as used to generate Figure 4. Figure 12 shows the histogram of the obtained correlation length with the $p(C_1)$ and $p(C_2)$ from the synthetic data, and Figure 13 shows the mean

and the standard deviation estimated using Eq. (18).

For the cases of different values of FNR, we generate the synthetic data with FNR=1.0, 2.0 and 4.0 with $\Delta T = 15,000$ μs ; whereas all other values are kept to be same as before. The results are shown in Figure 14 and Figure 15. We find that the Lorentzian-shaped eddies are not very much different from the Gaussian-shaped eddies, cf. Figure 4-Figure 7.

-
- [1] J. Meinecke, P. Tzeferacos, A. Bell, R. Bingham, R. Clarke, E. Churazov, R. Crowston, H. Doyle, R. P. Drake, R. Heathcote, M. Koenig, Y. Kuramitsu, C. Kuranz, D. Lee, M. MacDonald, C. Murphy, M. Notley, H.-S. Park, A. Pelka, A. Ravasio, B. Reville, Y. Sakawa, W. Wan, N. Woolsey, R. Yurchak, F. Miniati, A. Schekochihin, D. Lamb, and G. Gregori, *Proc. Nat. Acad. Sci. USA* **112**, 8211 (2015).
- [2] B. Carreras, *IEEE Trans. Plasma Sci.* **25**, 1281 (1997).
- [3] N. Iwama, Y. Ohba, and T. Tsukishima, *J. Appl. Phys.* **50**, 3197 (1979).
- [4] J. M. Beall, Y. C. Kim, and E. J. Powers, *J. Appl. Phys.* **53**, 3933 (1982).
- [5] T. S. Horbury, in *Cluster II Workshop: Multi-scale/Multipoint Plasma Measurements*, edited by R. A. Harris (2000).
- [6] G. Taylor, *Proc. R. Soc. London* **A164**, 476 (1938).
- [7] P. Goldreich and S. Sridhar, *Astrophys. J.* **438**, 763 (1995).
- [8] J. Cho and A. Lazarian, *Astrophys. J.* **615**, L41 (2004).
- [9] A. A. Schekochihin, S. C. Cowley, W. Dorland, G. W. Hammett, G. G. Howes, E. Quataert, and T. Tatsuno, *Astrophys. J. Suppl.* **182**, 310 (2009).
- [10] S. V. Nazarenko and A. A. Schekochihin, *J. Fluid Mech.* **677**, 134 (2011).
- [11] T. S. Horbury, M. Forman, and S. Oughton, *Phys. Rev. Lett.* **101**, 175005 (2008).
- [12] J. J. Podesta, *Astrophys. J.* **698**, 986 (2009).
- [13] R. T. Wicks, T. S. Horbury, C. H. K. Chen, and A. A. Schekochihin, *Mon. Not. R. Astron. Soc.* **407**, L31 (2010).
- [14] M. Barnes, F. I. Parra, E. G. Highcock, A. A. Schekochihin, S. C. Cowley, and C. M. Roach, *Phys. Rev. Lett.* **106**, 175004 (2011).
- [15] Y.-c. Ghim, A. A. Schekochihin, A. R. Field, I. G. Abel, M. Barnes, G. Colyer, S. C. Cowley, F. I. Parra, D. Dunai, S. Zoletnik, and the MAST Team, *Phys. Rev. Lett.* **110**, 145002 (2012).
- [16] C. P. Ritz, E. J. Powers, T. L. Rhodes, R. D. Bengtson, K. W. Gentle, H. Lin, P. E. Phillips, A. J. Wootton, D. L. Brower, N. C. Luhmann, W. A. Peebles, P. M. Schoch, and R. L. Hickok, *Rev. Sci. Instrum.* **59**, 1739 (1988).
- [17] D. L. Winslow, R. D. Bengtson, B. Richards, and W. A. Craven, *Rev. Sci. Instrum.* **68**, 396 (1997).
- [18] H. Thomsen, M. Endler, J. Bleuel, A. V. Chankin, S. K. Erents, G. F. Matthews, and C. to the EFDA-JET workprogramme, *Phys. Plasmas* **9**, 1233 (2002).
- [19] M. Lampert, G. Anda, A. Czopf, G. Erdei, D. Guszejnov, Á. Kovácsik, G. I. Pokol, D. Réfy, Y. U. Nam, and S. Zoletnik, *Rev. Sci. Instrum.* **86**, 073501 (2015).
- [20] W. Lee, I. Hong, J. Leem, M. Kim, Y. Nam, G. S. Yun, H. K. Park, Y. G. Kim, K. W. Kim, C. W. Domier, and N. C. J. Luhman, *J. Instrum.* **7**, C01070 (2012).
- [21] B. Tal, A. Bencze, S. Zoletnik, G. Veres, and G. Por, *Phys. Plasmas* **18**, 122304 (2011).
- [22] Y.-c. Ghim, A. R. Field, D. Dunai, S. Zoletnik, L. Bardóczy, A. A. Schekochihin, and the MAST Team, *Plasma Phys. Controlled Fusion* **54**, 095012 (2012).
- [23] J. Kim, M. Fox, A. Field, Y. Nam, and Y.-c. Ghim, *Comput. Phys. Commun.* **204**, 152 (2016).
- [24] R. J. Fonck, G. Cosby, R. D. Durst, S. F. Paul, N. Bretz, S. Scott, E. Synakowski, and G. Taylor, *Phys. Rev. Lett.* **70**, 3736 (1993).
- [25] G. R. McKee, C. Fenzi, R. J. Fonck, and M. Jakubowski, *Rev. Sci. Instrum.* **74**, 2014 (2003).
- [26] Y.-c. Ghim, A. R. Field, A. A. Schekochihin, E. G. Highcock, C. Michael, and the MAST Team, *Nucl. Fusion* **54**, 042003 (2014).
- [27] J. E. Maggs and G. J. Morales, *Phys. Rev. Lett.* **107**, 185003 (2011).
- [28] G. Hornung, B. Nold, J. E. Maggs, G. J. Morales, M. Ramisch, and U. Stroth, *Phys. Plasmas* **18**, 082303 (2011).
- [29] D. A. D'Ippolito, J. R. Myra, and S. J. Zweben, *Phys. Plasmas* **18**, 060501 (2011).
- [30] A. Bencze and S. Zoletnik, *Phys. Plasmas* **12**, 052323 (2005).
- [31] S. Zoletnik, S. Fiedler, G. Kocsis, G. K. McCormick, J. Schweinzer, and H. P. Winter, *Plasma Phys. Controlled Fusion* **40**, 1399 (1998).
- [32] N. L. J. Regina C. Elandt-Johnson, *Survival Models and Data Analysis* (Wiley, 1999).
- [33] L. A. Goodman, *J. Am. Stat. Assoc.* **55**, 708 (1960).