General formula for stationary or statistically homogeneous probability density functions

Emily S. C. Ching
Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

(Received 23 January 1996)

A general formula is derived for the probability density function (PDF) of fluctuating physical quantities measured in any stationary or statistically homogeneous process. For stationary processes, the formula relates the PDF to two conditional means: two averages involving a general function of the quantity and its time derivatives, the time derivative of this function and the time derivative of the quantity, taken when the fluctuating quantity is at a certain value. A previous result by Pope and Ching [Phys. Fluids A **5**, 1529 (1993)] is a special case of this general formula when the function is chosen to be the time derivative of the fluctuating quantity. An analogous formula is obtained for the PDF of fluctuating physical quantities measured in statistically homogeneous processes with spatial derivatives in place of time derivatives. [S1063-651X(96)01206-8]

PACS number(s): 47.27.Gs, 05.40.+j

In turbulent fluid flows, the physical quantities of interest, such as the velocity and temperature fields, display highly irregular fluctuations both in space and time. To study such processes, using a statistical approach is most natural. The statistics of any fluctuating quantity are described by its probability density function (PDF). One would, therefore, hope to be able to calculate the PDF's of the turbulent quantities directly from the equations of motion, i.e., for example, to calculate the PDF of velocity fluctuations from the Navier-Stokes equation. However, this task is highly nontrivial and, as of today, has not yet been accomplished, not even for the relatively simpler problem of temperature being a passive scalar advected by a random velocity field. Because of the difficulty in obtaining explicit results for the PDF's, it is useful to have exact implicit formulas relating the PDF's to other physical quantities.

One such formula was derived by Pope and the present author [1]. In this work, it is found that the PDF of any physical quantity measured in a stationary process can be expressed in terms of two conditional means of the time derivatives of the quantity. Denoting the physical variable by $X(t)$ and its PDF by $P(x)$, the formula found is

$$P(x)=\frac{C_N}{\langle \dot{X}^2|x\rangle}\exp\left(\int_0^x \frac{\langle \ddot{X}|x'\rangle}{\langle \dot{X}^2|x'\rangle}dx'\right), \tag{1}$$

where $\langle\cdots\rangle$ denotes ensemble average and an overdot indicates time derivative. The conditional means $\langle \ddot{X}|x\rangle$ and $\langle \dot{X}^2|x\rangle$ (short hand notations for $\langle \ddot{X}|X=x\rangle$ and $\langle \dot{X}^2|X=x\rangle$) represent, respectively, the averages of $\ddot{X}$ and $\dot{X}^2$, subject to the condition that $X(t)$ is at a certain value x. These conditional means are, therefore, functions of x. The constant C_N is not arbitrary but is determined by the normalization condition: $\int_{-\infty}^{\infty}P(x)dx=1$. For simplicity, we take $\langle X\rangle=0$ and $\langle X^2\rangle=1$ in the remaining of this paper.

When studying the PDF's of fluctuating physical quantities in turbulent flows, an interesting question is what their shapes are. The statistics of velocity and temperature derivatives, which are believed to be small-scale characteristics, have been known to deviate significantly from Gaussian and this is related to the problem of intermittency [2], which is a fundamental problem in turbulence. More recent interest stems from the discovery that the PDF of temperature fluctuations in turbulent Rayleigh-Bénard convection changes from Gaussian in the lower Rayleigh-number regime (known as soft turbulence) to exponential-like in the higher Rayleigh-number regime (known as hard turbulence) [3,4]. This discovery has prompted various studies which try to understand the non-Gaussianity of scalar fluctuations in turbulent flows [5–15].

From (1), we see that the shape of the PDF, especially its tails, is governed by the functional dependence of the two conditional means on x. For turbulent temperature fluctuations, both in thermal convection and in the wake of a slightly heated cylinder, the conditional mean $\langle \ddot{X}|x\rangle$ is found [1,11] to be reasonably well approximated by a linear function of x, which implies

$$r(x)\equiv\frac{\langle \ddot{X}|x\rangle}{\langle \dot{X}^2\rangle}\approx -x. \tag{2}$$

This linearity of $r(x)$ has also been found to hold approximately for spanwise vorticity data taken in several different turbulent shear flows [16]. The existence of such simple and general statistical feature in turbulence, a complicated phenomenon, is quite surprising. In Ref. [16], it is noted that linearity is one possible form of $r(x)$ such that the statistical constraints $\langle r(x)\rangle=0$ and $\langle xr(x)\rangle=-1$ are satisfied, regardless of what the PDF of X is. Nevertheless, the physics behind this simple statistical feature remains to be understood.

As a result of (2), the tails of the PDF of turbulent temperature fluctuations is mainly determined by the functional form of the other conditional mean $\langle \dot{X}^2|x\rangle$. Indeed, the change of statistics from Gaussian in soft turbulence to exponential-like in hard turbulence can be directly attributed to the change in behavior of $\langle \dot{X}^2|x\rangle$ in the two turbulent regimes [11]. In Fig. 1, we plot $q(x)\equiv\langle \dot{X}^2|x\rangle/\langle \dot{X}^2\rangle$ with $X(t)\equiv[T(t)-\langle T(t)\rangle]/\sqrt{\langle[T(t)-\langle T(t)\rangle]^2}$ for temperature measurements $T(t)$ taken, as described in Ref. [4], in the two turbulent regimes. It can be seen clearly that $q(x)$ changes from being approximately independent of x (and is, there-

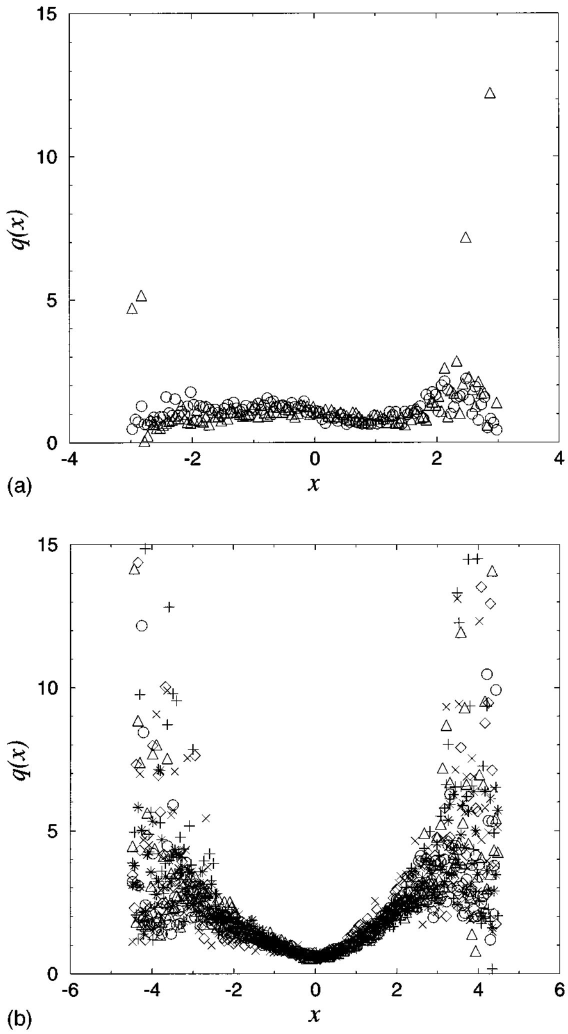


FIG. 1. The normalized conditional mean $q(x)=\langle \dot{X}^2|x\rangle/\langle \dot{X}^2\rangle$ as a function of x for turbulent temperature data in (a) soft turbulence [$\mathrm{Ra}=6.9\times10^6$ (circles) and $\mathrm{Ra}=2.1\times10^7$ (triangles)] and (b) hard turbulence [$\mathrm{Ra}=6.0\times10^8$ (pluses), $\mathrm{Ra}=4.0\times10^9$ (crosses), $\mathrm{Ra}=7.3\times10^{10}$ (triangles), $\mathrm{Ra}=6.0\times10^{11}$ (asterisks), $\mathrm{Ra}=6.7\times10^{12}$ (squares), $\mathrm{Ra}=4.1\times10^{13}$ (diamonds), and $\mathrm{Ra}=5.8\times10^{14}$ (circles)]. The large scatter in $q(x)$ when $|x|$ is large is due to the less frequent occurence of large $|x|$ fluctuations.

fore, approximately equal to 1) in soft turbulence to being approximately linear in $|x|$ in hard turbulence.

In this paper, I generalize the work of Pope and Ching [1] and derive a formula for the PDF. This formula contains (1) as a special case and involves two conditional means $\langle f\dot{X}|x\rangle$ and $\langle \dot{f}|x\rangle$, with f being a general function of X and its time derivatives. Equation (1) is recovered when f is taken to be $\dot{X}$. An analogous derivation of the PDF of fluctuating quantities measured in statistically homogeneous processes is also carried out. The result obtained is similar to that in the stationary case with spatial derivatives in place of time derivatives. With these exact formulas for the PDF's of fluctuating quantities, insights of the underlying physics determining the PDF's can be gained via the study of the conditional means. In fact, some of the conditional means can be calculated directly from the equations of motion in certain special cases of statistically homogeneous flows [17]. Moreover, these general formulas imply that conditional means involving different choices of the general function are related. Studying these relations may improve our understanding of the statistics of fluctuations in turbulence.

In the following, the general formulas are derived. First consider a stationary process. Let $X(t)$ be a physical variable measured as a function of time t at a certain fixed spatial location. For example, $X(t)$ can be the temperature or a component of the velocity measured in a stationary turbulent flow. Then the PDF of X, $P(x)$, is given formally by [18]

$$P(x)=\langle p(x,t)\rangle, \quad p(x,t)\equiv\delta(X(t)-x), \tag{3}$$

where $\langle\cdots\rangle$ is the ensemble average and is equivalent to the time average in stationary processes.

Differentiate $p(x,t)$ with respect to time gives

$$\frac{\partial p}{\partial t}=-\frac{\partial p}{\partial x}\dot{X}. \tag{4}$$

Multiplying a general function $f(X,\dot{X},\ddot{X},\ldots)$ of X and its time derivatives to both sides and taking ensemble average, we get

$$\langle \dot{f}|x\rangle P(x)=\frac{d}{dx}[\langle f\dot{X}|x\rangle P(x)]. \tag{5}$$

To obtain (5), we have used the fact that $\langle\partial(fp)/\partial t\rangle$ vanishes for stationary processes and that $\langle p(x,t)F(X,\dot{X},\ddot{X},\ldots)\rangle=\langle F(X,\dot{X},\ddot{X},\ldots)|x\rangle P(x)$ for any function $F(X,\dot{X},\ddot{X},\ldots)$.

We can also follow the method used in Ref. [5] to derive (5). Because $X(t)$ is stationary, we have

$$\langle \dot{f}h(X)\rangle=-\langle h'(X)f\dot{X}\rangle \tag{6}$$

for any differentiable function f of X and its time derivatives and for any well-behaved function h of X. In (6), $h'(X)$ denotes the derivative of h with respect to X. Writing the ensemble averages in terms of the PDF, we get

$$\int h(x)\langle \dot{f}|x\rangle P(x)dx=-\int h'(x)\langle f\dot{X}|x\rangle P(x)dx. \tag{7}$$

Integrating by parts the expression on the right-hand side gives

$$\int h(x)\langle \dot{f}|x\rangle P(x)dx=\int h(x)\frac{d}{dx}[\langle f\dot{X}|x\rangle P(x)]dx. \tag{8}$$

Since (8) is valid for any arbitrary well-behaved $h(x)$, this implies the integrands [apart from $h(x)$] have to be equal to each other, which is exactly (5).

Equation (5) is valid for any differentiable function f. To obtain a formula for $P(x)$, we divide both sides by $\langle f\dot{X}|x\rangle P(x)$ and integrate with respect to x:

$$P(x)=\frac{C_N}{\langle|f\dot{X}||x\rangle}\exp\left(\int_0^x\frac{\langle \dot{f}|x'\rangle}{\langle f\dot{X}|x'\rangle}dx'\right), \tag{9}$$

where C_N is again the normalization constant. Equation (9) is a general formula expressing the PDF of X in terms of two conditional means $\langle f\dot{X}|x\rangle$ and $\langle\dot{f}|x\rangle$, where $f(X,\dot{X},\ddot{X},\ldots)$ is any general differentiable function of X and its time derivatives such that

$$\langle f\dot{X}|x\rangle\neq 0 \quad \text{for all values of } x. \tag{10}$$

Since the same $P(x)$ can be expressed in terms of conditional means involving different f's, this implies that for any function f satisfying (10), we have

$$\frac{\langle\dot{f}|x\rangle}{\langle f\dot{X}|x\rangle}-\frac{d}{dx}\left[\ln\left(\frac{\langle f\dot{X}|x\rangle}{\langle\dot{X}^2|x\rangle}\right)\right]=\frac{\langle\ddot{X}|x\rangle}{\langle\dot{X}^2|x\rangle}. \tag{11}$$

In Ref. [1], (4) is differentiated once more and the ensemble average of the resulting equation leads to an ordinary differential equation for $P(x)$ which is then solved to give (1). A natural question one may ask is: what result one would get if (4) is differentiated n times? The answer is, we get

$$\sum_{j=1}^{n}(-1)^j\frac{d^j}{dx^j}[\langle\dot{f}_{n,j}|x\rangle P(x)]+(-1)^{j+1}\frac{d^{j+1}}{dx^{j+1}}[\langle f_{n,j}\dot{X}|x\rangle P(x)]=0 \tag{12}$$

with

$$f_{n+1,1}=\dot{f}_{n,1},$$
$$f_{n+1,j}=f_{n,j-1}\dot{X}+\dot{f}_{n,j},\quad \forall j=2,\ldots,n,$$
$$f_{n+1,n+1}=f_{n,n}\dot{X}, \tag{13}$$

and $f_{11}=\dot{X}$. We see that (12) is already contained in (5). Moreover, (1) is a special case of the general formula (9) when we take $f=\dot{X}$. Interestingly, such a choice of f is the simplest nontrivial choice that one can make since $\langle f\dot{X}|x\rangle$ vanishes for f being any polynomial in X in stationary processes.

For the next simplest choice, we take $f=\dot{X}^3$ and the formula for the PDF becomes

$$P(x)=\frac{C_N}{\langle\dot{X}^4|x\rangle}\exp\left(3\int_0^x\frac{\langle\dot{X}^2\ddot{X}|x'\rangle}{\langle\dot{X}^4|x'\rangle}dx'\right). \tag{14}$$

Comparing with (1), (14) contains conditional means of higher powers of $\dot{X}$. As the effect of noise in the data is more pronounced when one computes such conditional means, one expects the PDF obtained from (14) is noiser than that from (1). This is indeed the case, as shown in Fig. 2. The data used are the temperature measurements taken at Ra$=7.3\times10^{10}$ [4] in turbulent convection. The solid line is the PDF computed directly from the data, which is almost indistinguishable from the solid triangles which are obtained from (1). The squares are obtained from (14) and scatter much more, as expected. Nonetheless, we see that (14) is verified.

Since the tail of the PDF is mainly determined by the exponential factor in (9), we expect the ratio of the two conditional means, $\langle\dot{f}|x\rangle/\langle f\dot{X}|x\rangle$, to be approximately the same for all allowed f's and, in particular, approximately equal to

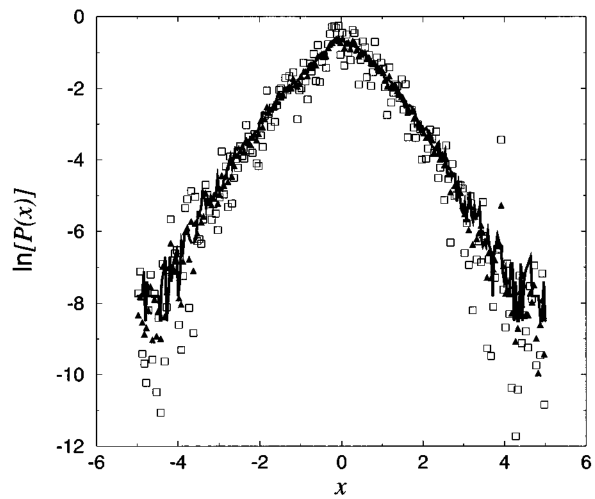


FIG. 2. Comparison of the PDF directly computed from the temperature data of turbulent convection taken at Ra$=7.3\times10^{10}$ (solid line) with the PDF's constructed from (1) (solid triangles) and (14) (squares) using the conditional means computed from the data.

$\langle\ddot{X}|x\rangle/\langle\dot{X}^2|x\rangle$. [This means that in the left-hand side of (11), the x dependence of the second term is much weaker than that of the first term.] We find that not only this is true, but that the two conditional means, when normalized,

$$r_f(x)\equiv\frac{\langle\dot{f}|x\rangle}{\langle f\dot{X}\rangle},\quad q_f(x)\equiv\frac{\langle f\dot{X}|x\rangle}{\langle f\dot{X}\rangle}, \tag{15}$$

are also approximately the same for $f=\dot{X}$ and $f=\dot{X}^3$. [Note that $r_f(x)$ and $q_f(x)$ for $f=\dot{X}$ are just $r(x)$ and $q(x)$.] The comparison is shown in Fig. 3. As expected, $r_f(x)$ and $q_f(x)$ for $f=\dot{X}^3$ scatter more.

We next consider statistically homogeneous processes. The derivation of a general formula for the PDF in this case exactly parallels the derivation in the case of stationary processes described above. Suppose now $Y(\mathbf{r})$ is a physical variable measured as a function of position $\mathbf{r}$ at a certain time in a statistically homogeneous fluid flow. The PDF of Y, denoted as $Q(y)$, is given by

$$Q(y)=\langle q(y,\mathbf{r})\rangle,\quad q(y,\mathbf{r})=\delta(Y(\mathbf{r})-y), \tag{16}$$

where the ensemble average $\langle\cdots\rangle$ is equivalent to spatial average in this case.

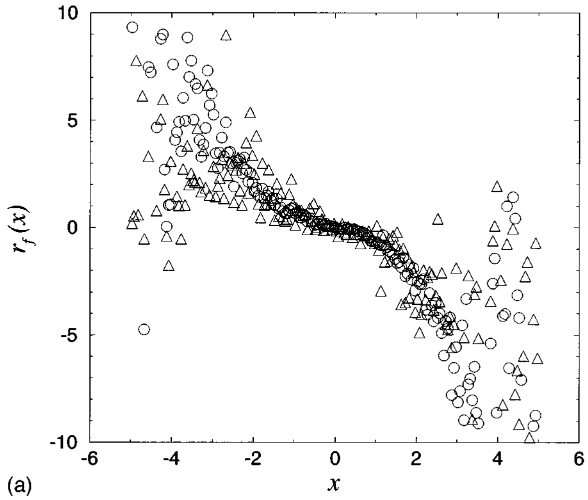
Differentiate $q(y,\mathbf{r})$ with respect to $\mathbf{r}$ gives

$$\nabla q(y,\mathbf{r})=-\frac{\partial q}{\partial y}\nabla Y. \tag{17}$$

Taking the scalar product of both sides with a general differentiable vector function $\mathbf{g}(Y,\nabla Y,\nabla^2 Y,\ldots)$ to both sides and then ensemble (or spatial) averaging, we have

$$\langle\nabla\cdot\mathbf{g}|y\rangle Q(y)=\frac{d}{dy}[\langle\mathbf{g}\cdot\nabla Y|y\rangle Q(y)]. \tag{18}$$

For homogeneous systems, there cannot be any forcing from the boundaries on Y and its spatial gradient so the term



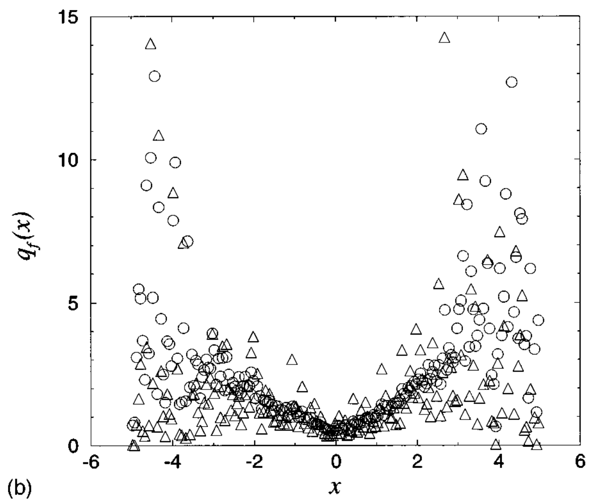


FIG. 3. Comparison of (a) $r_f(x)$ and (b) $q_f(x)$ [see (15) for definitions] for $f=\dot{X}$ (circles) and $f=\dot{X}^3$ (triangles) using the same data as in Fig. 2.

$\langle \nabla \cdot (q\mathbf{g}) \rangle$ vanishes. As in the case of stationary processes, (18) can be derived by another method. Solving (18) for those functions $\mathbf{g}$'s such that

$$\langle \mathbf{g} \cdot \nabla Y | y \rangle \neq 0 \quad \text{for all values of } y, \tag{19}$$

we find

$$Q(y) = \frac{C_N}{\langle |\mathbf{g} \cdot \nabla Y| | y \rangle} \exp\left(\int_0^y \frac{\langle \nabla \cdot \mathbf{g} | y' \rangle}{\langle \mathbf{g} \cdot \nabla Y | y' \rangle} dy' \right). \tag{20}$$

Again, since (20) is valid for any general $\mathbf{g}$ satisfying (19), this implies that the conditional means for different choices of $\mathbf{g}$ are related. The relation is similar to (11):

$$\frac{\langle \nabla \cdot \mathbf{g} | y \rangle}{\langle \mathbf{g} \cdot \nabla Y | y \rangle} - \frac{d}{dy}\left[\ln\left(\frac{\langle |\mathbf{g} \cdot \nabla Y| | y \rangle}{\langle |\nabla Y|^2 | y \rangle} \right) \right] = \frac{\langle \nabla^2 Y | y \rangle}{\langle |\nabla Y|^2 | y \rangle}. \tag{21}$$

A special result of the general formula (20) is obtained by choosing $\mathbf{g} = \nabla Y$:

$$Q(y) = \frac{C_N}{\langle |\nabla Y|^2 | y \rangle} \exp\left(\int_0^y \frac{\langle \nabla^2 Y | y' \rangle}{\langle |\nabla Y|^2 | y' \rangle} dy' \right), \tag{22}$$

which is the exact analog to (1) for stationary processes [15]. This result is partcularly interesting because we find that, in certain circumstances, the conditional means $\langle |\nabla Y|^2 | y \rangle$ and $\langle \nabla^2 Y | y \rangle$ can be evaluated directly from the equations of motion [17].

In summary, general formulas in terms of conditional means are obtained for the PDF of fluctuating physical variables in both stationary and statistically homogeneous processes, including turbulent fluid flows. We do not expect these formulas themselves will lead to deeper understanding of turbulence as their validity does not rely on the actual physics of the processes. However, physical insights can be gained through the study of the conditional means, especially when these conditional means have simple features that hold generally in various different turbulent flows. Moreover, these general formulas may also provide a means for evaluation of conditional means involving high-order time derivatives of the variable, which cannot be accurately computed directly from the data when the latter is not sampled fast enough.

ACKNOWLEDGMENTS

The author is grateful to A. Libchaber and X.Z. Wu for providing her with the turbulent convection data and acknowledges discussion with Y.K. Tsang. This work is supported in part by the Hong Kong Research Grant Council, Grant No. 458/95P.

[1] S.B. Pope and E.S.C. Ching, Phys. Fluids A **5**, 1529 (1993).
[2] See, for example, M. Nelkin, Adv. Phys. **43**, 143 (1994), and references therein.
[3] B. Castaing, G. Gunaratne, F. Heslot, L. Kadanoff, A. Libchaber, S. Thomas, X.Z. Wu, S. Zaleski, and G. Zanetti, J. Fluid Mech. **204**, 1 (1989).
[4] X.Z. Wu, Ph.D. thesis, University of Chicago, 1991 (unpublished).
[5] Y.G. Sinai and V. Yakhot, Phys. Rev. Lett. **63**, 1962 (1989).
[6] H. Chen, S. Chen, and R.H. Kraichnan, Phys. Rev. Lett. **63**, 2657 (1989).
[7] Z.-S. She and S.A. Orszag, Phys. Rev. Lett. **66**, 1701 (1991).
[8] A. Pumir, B.I. Shraiman, and E.D. Siggia, Phys. Rev. Lett. **66**, 2984 (1991).
[9] Jayesh and Z. Warhaft, Phys. Rev. Lett. **67**, 3503 (1991).
[10] J.P. Gollub, J. Clarke, M. Gharib, B. Lane, and O.N. Mesquita,

Phys. Rev. Lett. **67**, 3507 (1991).
[11] E.S.C. Ching, Phys. Rev. Lett. **70**, 283 (1993).
[12] H. Takayazsu and Y.-H. Taguchi, Phys. Rev. Lett. **70**, 782 (1993).
[13] R. Bhagavatula and C. Jayaprakash, Phys. Rev. Lett. **71**, 3657 (1993).
[14] E.S.C. Ching and Y. Tu, Phys. Rev. E **49**, 1278 (1994).
[15] L. Valino, C. Dopazo, and J. Ros, Phys. Rev. Lett. **72**, 3518 (1994).
[16] J. Mi and R.A. Antonia, Phys. Rev. E **51**, 4466 (1995).
[17] E.S.C. Ching and R. Kraichnan (unpublished).
[18] See, for example, S.B. Pope, Prog. Energy Combust. Sci. **11**, 1991 (1985).