Information content of turbulence

R. T. Cerbus[*] and W. I. Goldburg

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15221, USA

We treat a turbulent velocity field as a message in the same way as a book or a picture. All messages can be described by their entropy per symbol h, defined as in Shannon's theory of communication. In a turbulent flow, as the Reynolds number Re increases, more correlated degrees of freedom are excited and participate in the turbulent cascade. Experiments in a turbulent soap film suggest that the spatial entropy density h is a decreasing function of Re, namely $h \propto -\log \text{Re} + \text{const}$. In the logistic map, also analyzed here, increasing the control parameter r increases h. A modified logistic map with additional coupling to past iterations suggests the significance of correlations.

I. INTRODUCTION

Any physical system is an information channel, since it is "communicating its past to its future through its present" [1]. A series of data measured for such a system is thus a message, a sequence of symbols. Information theory is the natural framework for the quantitative study of messages [1–4]. The entropy H, also called the information, plays a central role in the theory. It is a measure of uncertainty or disorder. Shannon's theory of communication has found wide application in genetics [5], dynamical systems [6,7], and a variety of other fields in physics [8]. It has also been used extensively in statistical inference problems [9] and in this light has provided an interesting way of interpreting the maximization of entropy in statistical mechanics [10]. In this work the entropy density h, the entropy per symbol, is used as a measure of the information content in a 2D turbulent flow [11,12].

Treating a physical system as a source of information has its roots in the early development of nonlinear dynamics and chaos [6]. This connection is nicely illustrated by Pesin's theorem, which for many chaotic systems equates the sum of positive Lyapunov exponents λ with the Kolmogorov-Sinai entropy h_{KS} [7]. The Lyapunov exponent λ is the rate at which phase-space trajectories separate, while h_{KS} can be thought of a special case of h. As two initially almost indistinguishable trajectories separate, new details are uncovered. Thus if λ is large, new information is revealed faster, i.e., h is also larger.

When a physical system is probed, it reports to the experimenter an ordered sequence of signals $(s_1, s_2, s_3, \dots)$. In the present experiments, the measured signal is a sequence of velocities $(u_1, u_2, u_3, \dots)$ in a turbulent 2D flow. These fluctuate in magnitude about a mean flow speed.

Presumably, the disorder h of fluid flow is relatively small if the flow is almost laminar. In this limit of small Reynolds number Re, one expects h to increase with Re. In the opposite limit of large Re, a so-called inertial range of correlated eddies of various sizes develops. Increased correlations implies added constraints or redundancies, which always decreases the uncertainty and information content of any message [3]. One therefore expects that after passing through a maximum, h

will decrease with increasing Re. (This issue is addressed in Ref. [13] from a thermodynamic point of view.)

Although turbulence is both a temporal and spatial phenomenon, the fundamental work of Kolmogorov deals only with the spatial structure of turbulence [14,15]. The work of Kraichnan and others has shown that many of the essential features of turbulence are retained if one throws away temporal correlations but keeps spatial ones [16–18]. Thus, the expectation that h decreases ought to be true for a spatial series but may not be true for a time series. When it is necessary to make a distinction, the spatial entropy density will be denoted h_S while the temporal entropy rate will be denoted h_T.

The present experiments probe a turbulent system at high Re, where h_S is indeed seen to decrease with Re for a spatial velocity sequence. The near-laminar regime, where one expects h_S to increase with Re, is not experimentally accessible.

The temporal entropy rate h_T has often been used to probe turbulence [7,19–21]. Several novel experiments have shown that the onset of turbulence in many systems can be described by a low-dimensional strange attractor. The evidence for this is provided by probing h_T at the transition. In the case of Taylor-Couette flow [19], h_T and the largest λ increase with the Reynolds number Re, which can be thought of as a measure of the nonlinearity or strength of the flow. The usual expectation is that this trend continues as Re increases, as suggested by some models and analytic work [20,21]. This may be true for h_T, but this work provides evidence that h_S follows a different path.

The motivation for this study is twofold. First, this seems to be the first study of the spatial disorder of a turbulent velocity field as a function of Re. By characterizing the flow with the entropy density h_S, the fundamental role of the cascade in producing correlations is clearly manifested. Second, h is one of several fundamental quantities necessary to describe how a system creates and communicates information [1,22]. This approach appears to provide a new and interesting description of nature but has not yet been applied to many physical systems [1].

II. PARTITIONS

In order to treat a physical system as a message, the experimental data must be converted to symbols [23]. A partition is defined which separates the data into disjoint slices
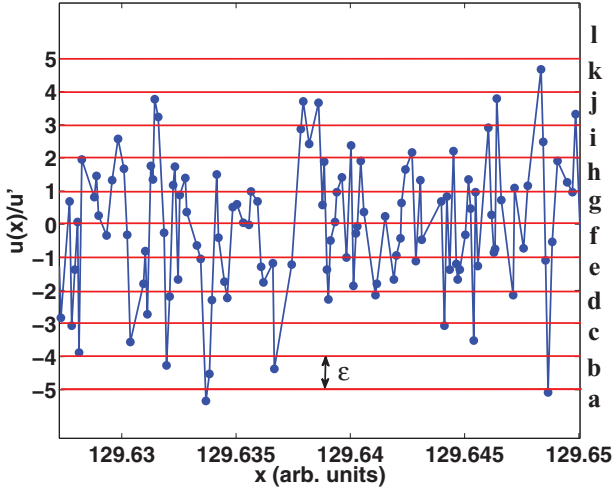
---

*rtc17@pitt.edu

FIG. 1. (Color online) An example of a partition used with turbulence data. The dividers are a distance ϵ apart. Data points that fall inside a section are assigned to that section's symbol. As ϵ decreases, the number of symbols required to describe the data increases.

of size ϵ. The data values in each specified range (slice) are then assigned to a unique symbol [23,24]. That is, if data points are ϵ apart, they correspond to different symbols. (In some sense, all experiments do this because of their limited precision.) The size and location of the divisions can be chosen to faithfully represent the original system even for seemingly coarse partitions [23].

Figure 1 shows an example of how a partition is used for some of the turbulence data where the mean speed has been subtracted out and the result divided by the rms velocity. The distance between the horizontal dividers is the partition size ϵ. Data falling between the same walls (or on the boundary as prescribed) are assigned to the same symbol. The number of symbols required to describe the data is inversely related to the magnitude of ϵ. For all the data used here, if $\epsilon \gtrsim 10$ and a divider line kept at 0, then the data are effectively binarized.

Correctly identifying those partitions which completely describe the system (called generating) can be extremely difficult [24]. However, much can and has been learned about complex systems such as the brain or turbulence even after converting a data series into a simple binary alphabet [23,25–27]. Approximate treatments are usually necessary and often useful, as long as they still represent the underlying system [23]. For a chaotic time series, the entropy rate h_T may approach a constant value (h_{KS}) as ϵ decreases [24,28]. For a spatially extended system, one may expect a similar asymptotic behavior.

III. ENTROPY AND ENTROPY ESTIMATION

The entropy of a message is usually defined as [2]

$$H[X] \equiv -\sum_x p(x) \log p(x), \tag{1}$$

where $p(x)$ is the probability of the symbol x occurring in the message. The argument $[X]$ indicates for which set of variables the entropy is calculated and is often dropped unless

necessary. The natural logarithm is used, giving the entropy in "nats." One may consider $-\log p(x)$ as a measure of the information gained from any one symbol. Thus the entropy is the average information of the message. If the message is completely random, then the surprise and the amount of new information H is maximal. H is generally large for broad distributions [3]. By contrast, a constant, unchanging stream of data will have zero entropy. The message contains no new information and no uncertainty.

However, one must take correlations into account since these always reduce the amount of information a message contains. Consider sequential blocks of symbols of length L. The probability of any unique block x^L is $p(x^L)$. The Shannon entropy of single symbols can then be generalized to define the block entropy

$$H_L \equiv -\sum_{x^L} p(x^L) \log p(x^L), \tag{2}$$

where the sum is over all blocks x^L. This block entropy will diverge as L goes to infinity. Therefore one defines a quantity h [3,29],

$$h = \lim_{L \to +\infty} h_L = \lim_{L \to +\infty} H_{L+1} - H_L. \tag{3}$$

This h is the extra information one gets from measuring one more symbol. The limit exists for stationary processes [3] and may be reached much sooner than $L = \infty$. In spatially extended systems, such as these turbulence measurements, h is called the entropy density h_S. For a time series, h is called the entropy rate or metric entropy h_T [29–32]. Although this distinction does not affect the calculation, it does influence the interpretation of the turbulence results for h_S. (For instance, there is no Pesin's theorem for the entropy density.) The h_S estimated here thus differs markedly from that considered in previous work where it has been estimated for time series [19–21,33].

The above definition already suggests problems one might have in estimating h, since the infinite limit is impossible for a finite data set. Fortunately, h_L for real data reaches an asymptote sooner than infinity since correlations are usually finite in scale. Some of the techniques designed to overcome the finite data issues can be found in Ref. [24]. Most methods involve making an assumption about the distribution of rare events. A technique proposed in Ref. [24] is used here, although the results are not changed much by its use (see Fig. 2).

In this work the block entropies are used to make an estimate of h by looking for the asymptote of h_L defined by Eq. (3) as shown in Fig. 2. This estimate will be called h^*. In Fig. 2, the naive (frequency count) estimate of h_L is plotted vs L along with the Grassberger estimate from Ref. [24]. The dotted line is the value of h determined by the inflection point of h_L. The asymptote is usually reached around $L \simeq 10$.

An alternative method for determining a message's information content is based on data compression. The lower limit for the length S to which a message can be compressed from its original length S_0, for any compression algorithm, is its entropy: $S \geqslant H$ [3,24,34,35]. Compression algorithms operate by finding redundancies and correlations in data and re-expressing the message in a shorter form. Compression
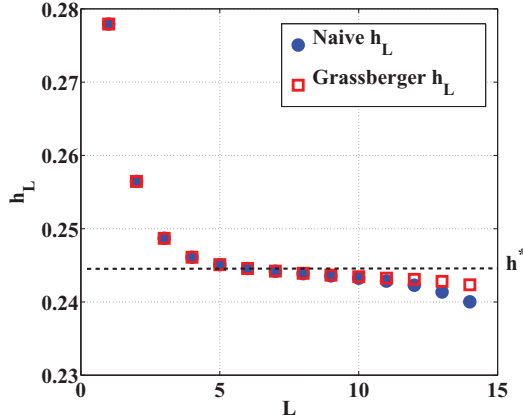
FIG. 2. (Color online) h_L as a function L for binarized turbulence data. h_L initially decreases until correlations are no longer important. The value of h_L at the point of inflection is taken to be h^*. For larger L, h_L decreases again due to undersampling. The naive estimate ($\bullet$) calculates probabilities based on frequency counts. The Grassberger estimate ($\square$) takes the undersampling bias into account [24].

provides a nice way of thinking about how much information is contained in a message, since it reduces the message to its "essentials." There is no way to shorten a completely random message since each symbol is independent of the others and there can be no compression. For a repetitive stream of symbols (like "... 111111 ... ") the message is trivially compressed to almost zero size.

The information content is then [34]

$$c = \log D\left(\frac{S}{S_0}\right), \qquad (4)$$

where D is the alphabet size (e.g., for a binary alphabet $D = 2$). The Lempel-Ziv algorithm is optimal in the sense that c converges to h in the limit of infinite S_0, so it can be used as another estimate of h and a check on h^*. The value of c is independent of file type but does require that the compression program be based on the Lempel-Ziv algorithm. In order to account for the "overhead" (file headers, etc.), a random data set is compressed and that compression ratio is used to normalize the real data [36]. Just as with h, the compression ratio will be denoted as c_S for a spatial series and c_T for a temporal one as necessary.

Traditionally, c has been given the name of algorithmic or Kolmogorov complexity [3,34,36]. This is a measure of the computational complexity of the data set in question. Even if c is not equal to h, it is still a measure of the information content of the data [36–38]. It is important to recognize the many limitations involved in calculating information content. At best h^* and c are approximations to h, but this does not make them meaningless and they can still be used for comparison [23].

IV. RESULTS

A. Logistic map

The estimates h^* and c are first applied to the logistic map as a test of the method as well as to illustrate some principles regarding h_T for chaotic systems. The logistic map is a simple one-dimensional nonlinear map which nicely illustrates
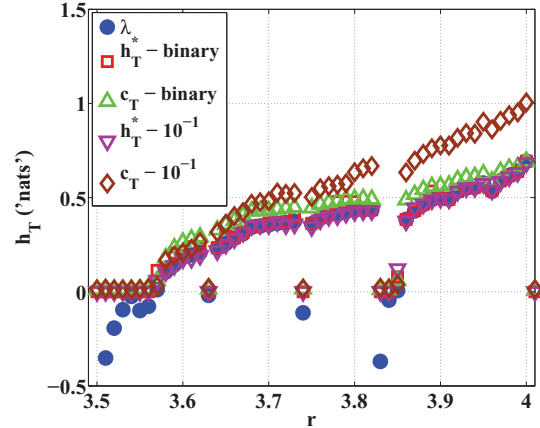

FIG. 3. (Color online) Lyapunov exponents λ ($\bullet$) for the logistic map plotted as a function of r (see text). The entropy rates were estimated using block entropies h_T^* (binary: $\square$, 10^{-1}: $\triangle$) and compression c_T (binary: $\triangledown$, 10^{-1}: $\diamond$). Although h_T^* performs better as an estimate of h_T, all estimates show the same trend.

chaotic behavior [28],

$$x_{n+1} = r x_n (1 - x_n), \qquad (5)$$

where r is a parameter that increases the strength of the nonlinearity. As r increases, the system goes through a series of period-doubling bifurcations and eventually becomes chaotic at $r \simeq 3.56995$ ($\lambda > 0$). As usual, $x \in (0,1)$ and $r \in [0,4]$. As mentioned earlier, Pesin's theorem states that the sum of the positive Lyapunov exponents λ is equal to h_{KS}, as long as the system satisfies certain conditions [7]. For the logistic map, which is one-dimensional, there is only one λ for each value of r. The value of λ has been calculated as a function of r, using the algorithm in Ref. [39], and is compared with h_T^* and c_T in Fig. 3. For each value of r, a randomly chosen initial condition is iterated 10^6 times.

Two partitions are shown in Fig. 3. A binary partition is used where $x = 1$ if $x \geqslant 0.5$ and $x = 0$ if $x < 0.5$, so $\epsilon = 0.5$. (The location of this partition divider is important [40].) The second partitioning involves simply rounding the data to the first decimal point ($\epsilon = 10^{-1}$) and assigning a symbol to each distinct data value. The estimate h_T^* performs very well, while c_T shows significant deviations for the 10^{-1} partition. Despite its shortcomings in estimating h_T, c_T is nonetheless useful as a measure of the information contained in these finite sequences, as discussed above. It follows the same trend as λ and reveals the logistic map's information dependence on r. The values of r for which λ is negative have $h_T \simeq 0$, since it is positive definite.

Partitions with as few as 2 slices or as many as 1000 slices give essentially the same h_T^* and c_T. This is because the partitions are all generating, i.e., they represent the dynamics faithfully and the entropy calculated for any of them is the Kolmogorov-Sinai entropy h_{KS} [24]. In other words, anything smaller than a binary partition is overkill. This is not always true but suggests that crude representations of data can still capture important features. This emboldens us to do the same for turbulence, to be discussed in Sec. IV C.

Once the transition to chaos occurs, λ and the estimates of h_T increase almost monotonically. There are several isolated
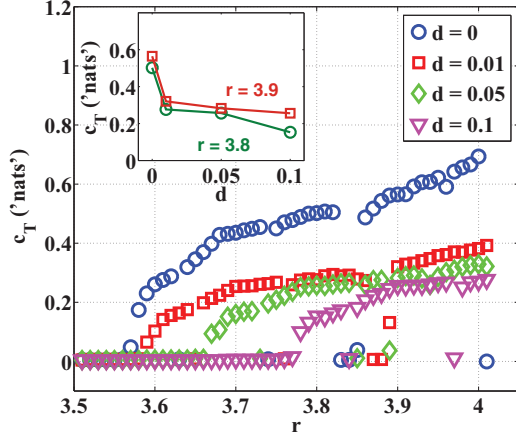
FIG. 4. (Color online) Entropy rate estimate c_T for the modified logistic map using a binary partition. (Recall that a binary partition is generating for the logistic map.) As d increases, c is lowered considerably. ($d = 0$: $\bigcirc$, $d = 0.01$: $\square$, $d = 0.05$: $\diamond$, $d = 0.1$: $\triangledown$.) Inset: c_T vs d for fixed r in the chaotic regime.
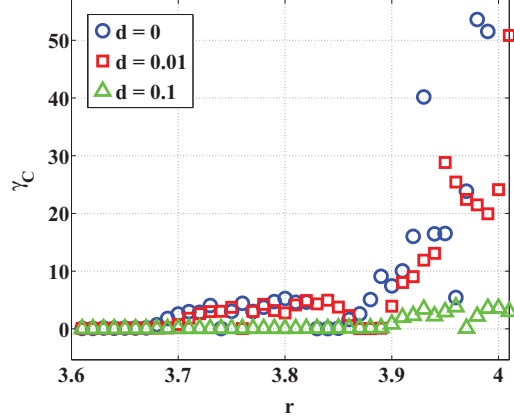

FIG. 5. (Color online) Decay rate of mutual information for logistic map and modified logistic map as a function of r using a binary partition ($d = 0$: $\bigcirc$, $d = 0.01$: $\square$, $d = 0.1$: $\triangle$). The decay rate here was calculated as the reciprocal of the area underneath the mutual information curve: $1/\sum_0^\infty I(\Delta)$. If this is large, then the correlations are weak. As the coupling d increases, the correlations get stronger.

regions where the logistic map returns to periodic behavior [28] and so $\lambda < 0$ and $h_T \simeq 0$. The general behavior appears to be that as the strength of the nonlinearity increases (see Fig. 3), so does h_T. Chaos creates information. Similar behavior was observed at the onset of turbulence in Taylor-Couette flow [19]. This increase in h_T for the logistic map is accompanied by a decrease in the strength of correlations, as will be shown shortly.

B. Modified logistic map

In order to get a better picture of the importance of correlations, a modified logistic map is introduced to explicitly increase correlations through a term that couples to previous iteration values further back than one. Denoting

$$f(x) = rx(1 - x), \tag{6}$$

the modified logistic map is defined as

$$x_{n+1} = f(x_n) + d\left[\frac{f(x_{n-2}) + f(x_{n-1})}{2} - f(x_n)\right], \tag{7}$$

where d is the coupling strength. This modification is really a kind of logistic delay map [28]. Now using three random intial conditions, this map is also iterated 10^6 times and the compression estimate is used to compare h_T for different values of d. The results are shown in Fig. 4.

Even for small d, c_T is changed drastically. As d is increased, c_T is decreased more and the transition to chaos shifts to larger values of r. This suggests that in addition to decreasing h_T, correlations can also act to suppress the chaotic transition.

In order to quantify correlations for messages, it is useful to introduce the mutual information I [3]. This is a measure of the information shared between two variables. For two variables X and Y it is

$$I(X;Y) \equiv \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = H(X) - H(X|Y), \tag{8}$$

where $p(x,y)$ is the joint probability. Here uppercase letters denote the variable and lowercase letters denote specific values of that variable. The second equality shows that the mutual information may also be thought of as the information about variable X minus the information about X given knowledge of Y. When X and Y are uncorrelated, $I(X;Y) = 0$. When the two variables X and Y are symbols separated by a certain number of symbols Δ ($X(i)$, $Y = X(i + \Delta)$), then $I(\Delta)$ becomes like an autocorrelation function for symbolic sequences [41,42]. For the logistic map and modified logistic map, Δ is a temporal interval while for the turbulence measurements Δ is a spatial interval.

The mutual information is observed to decay exponentially for more than a decade for the chaotic regime of the logistic map and the logistic delay map, with a decay rate γ_C that increases with r (see Fig. 5). Put another way $1/\gamma_C$, which can be thought of as a correlation time, decreases with r. The correlations are thus decreasing as the strength of the nonlinearity increases, which corresponds well with the understanding that h_T is reduced by correlations. Figure 5 shows γ_C as a function of r for three different values of d. The addition of coupling has increased the strength of the correlations and mirrors the drop in c_T.

C. Turbulence

Now consider the real physical system of a turbulent soap film, which is a good approximation to 2D turbulence since the film is only several μm thick [11,12]. The soap solution is a mixture of Dawn (2%) detergent soap and water with 1.5-μm particles added for laser doppler velocimetry (LDV) measurements. Figure 6 is a diagram of the experimental setup.

The soap film is suspended between two vertical blades connected to a nozzle above and a weight below by nylon fishing wire. The nozzle is connected by tubes to a valve and a reservoir which is constantly replenished by a pump that brings the soap solution back up after it has flowed through. The flow is gravity driven. Typical centerline speeds u are several
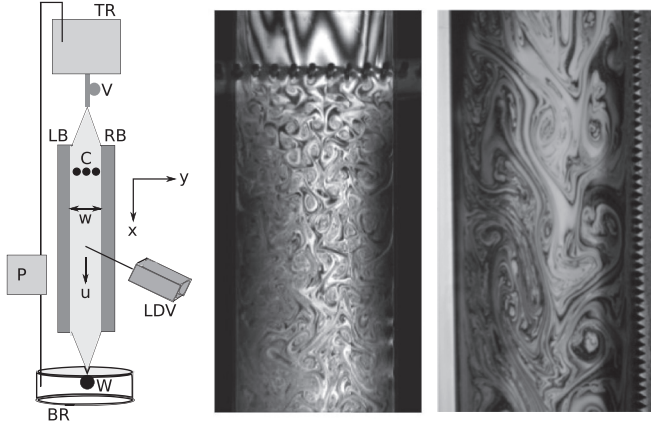
FIG. 6. Left: Experimental setup showing the reservoirs (TR and TB), pump (P), valve (V), comb (C), blades (LB and RB), weight (W), and LDV. Center: Fluctuations in film thickness from turbulent velocity fluctuations for smooth walls and a comb. (Here the thin blades have been replaced by fishing wires.) Right: Fluctuations created by a rough wall. The width w of the channels is several cm.

FIG. 7. (Color online) Representative one-dimensional energy spectra in a log-log plot of $E(k)$ vs k. The enstrophy cascade ($\triangle$) has a slope close to -3 while the energy cascade ($\square$) has a slope close to $-5/3$. The flat curve ($\bigcirc$) has no cascade.

hundred cm/s with rms fluctuations u' ranging from roughly 1 to 30 cm/s. The channel width w is usually several cm.

Turbulence in the soap film is generated by either (1) inserting a row of rods (comb) perpendicular to the film or (2) replacing one or both smooth walls with rough walls (saw blades) with the comb removed. When protocol (1) is used decaying 2D turbulence results which is almost always accompanied by the direct enstrophy cascade [11,12]. If procedure (2) is used, then forced 2D turbulence can be generated with an inverse energy cascade [11,12]. The ability to see the inverse energy cascade depends sensitively on the flux and channel width. This sensitivity is decreased if two rough blades are used. The type of cascade is determined by measuring the one-dimensional velocity energy spectrum $E(k)$, where $\frac{1}{2}u'^2 = \int_0^\infty E(k)dk$.

Although a condensate has been observed in some 2D turbulent systems [12], it is not present in this one. A condensate is revealed by a sharp spike in $E(k)$, which is never observed. In other experimental arrangements, two slopes are seen in a log-log plot of $E(k)$ vs k, indicating a dual cascade of both energy and enstrophy [11,12,43]. For these experiments only one slope is observed.

Measurements of the velocity are usually taken near the vertical middle of the channel. In all cases, the data are obtained for the longitudinal velocity component at the horizontal center of the channel. The data rate is $\simeq 5000$ Hz and the time series typically had more than 10^6 data points. For this system, the time series should be thought of as a spatial series by virtue of Taylor's frozen turbulence hypothesis [11,12,15]. Its validity has been thoroughly tested for this system [44]. The fact that these measurements involve a spatial series rather than a time series is a crucial point.

With this high data rate, the smallest turbulent scales are easily resolved. A number of measurements were taken near the top of the channel where the flow is still quite slow. In this case there is no power-law scaling in $E(k)$ and so apparently no cascade, although the flow is not laminar ($u' \neq 0$). Some representative spectra are shown in Fig. 7.
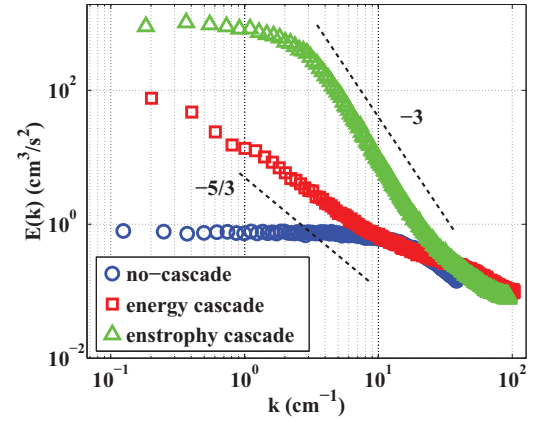
Before converting the velocity data into symbols, the mean velocity is subtracted out and the result divided by u'. This was done to have a similar alphabet size for different Re and seems a natural way to treat the data. The velocity data were then partitioned in a similar way to the logistic map. That is, the data were separated into slices of various sizes and then converted into symbols. In the turbulence case a binary partition means that a 1 is assigned if the velocity is above the mean value and 0 if below.

The main results of this paper appear in Fig. 8, which is a plot of h_S^* and c_S vs Re, where $\mathrm{Re} \equiv u'w/v$ and v is the kinematic viscosity. The Reynolds number Re is a measure of the nonlinearity of the system, much like r for the logistic map. Four different estimates of h_S are shown in Fig. 8. The open circles ($\bigcirc$) and squares ($\square$) show h_S^* and c_S, respectively, for the binary partition. The two upper data sets ($\triangle, \triangleright$) are h_S^* and c_S for a finer partition where the velocity data are distinguished
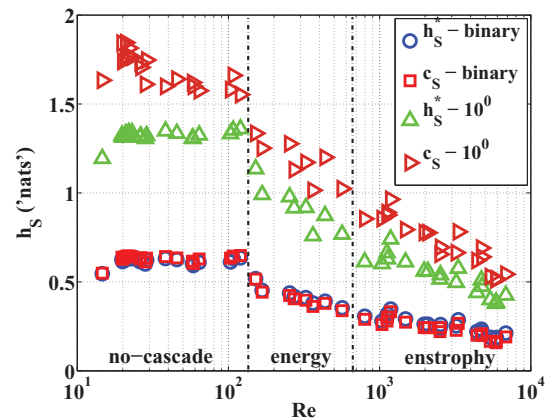

FIG. 8. (Color online) Entropy density estimates of h_S vs Re for the 2D turbulent data. h_S is a decreasing function of Re. The flat, higher h_S region corresponds to the no-cascade data. The decay begins with the emergence of the cascade. A binary partition (h_S^*: $\bigcirc$, c_S: $\square$) and a second partition where the data is saved to the nearest integer (h_S^*: $\triangle$, c_S: $\triangleright$) are shown here. Dividing lines are shown that separate the data into no-cascade, energy, and enstrophy regions, respectively.

by their first significant figure ($\epsilon = 1$). The same trend is shown by both partitions and for all partitions studied, namely that h_S^* and c_S are decreasing functions of Re.

Note that h_S^* and c_S are very weakly dependent on Re: $h_S^* \propto -\log \text{Re} + \text{const}$ after an initial plateau. This very slow decay invites an explanation. The decrease begins as soon as a cascade appears, as seen in Fig. 8. The decrease is independent of the type of cascade, as both the energy and enstrophy cascade data are present in the figure. The flat region at low Re corresponds to the data without a cascade.

At first glance this result seems surprising, but the decrease is in accord with the common picture of the turbulent cascade [11,12,15]. The energy (enstrophy) flows from one scale r to nearby spatial scales. The eddies participating in the cascade are necessarily correlated and the extent of the inertial range (cascade region) increases with Re. Since laminar flow is presumably not disordered at all, this implies that h_S passes through a local maximum at an intermediate value of Re. It is regrettable that the soap film is not stable at low Re, thus hindering the observation of this local maximum.

Although the system under study here is two dimensional, the same decrease in h_S should also hold for three-dimensional turbulence in the fully developed regime. It should be noted that the results in Fig. 8 appear to be somewhat similar to that of Wijesekera *et al.* in their study of spatial density fluctuations in the ocean [45]. However, the behavior of the spatial entropy density h_S vs Re observed here differs substantially from that of the temporal entropy rate h_T studied elsewhere [19–21,33].

The spatial correlations in the flow are becoming increasingly important as Re increases. This is evidenced by the decrease in the (spatial) decay rate γ_C for the mutual information $I(\Delta)$ as shown in Fig. 9. Note the similarities between Figs. 9 and 8. The increased strength of the correlations is responsible for the decrease in h_S, which follows from its definition in Eq. (3).

Unlike the logistic map, the turbulence data are more sensitive to the size of the partition ϵ when converting to

FIG. 10. (Color online) Entropy density h_S^* as a function of the partition size ϵ for three different Re. (The largest partition size corresponds to binarized data.) The three curves correspond to the no-cascade ($\bigcirc$), energy cascade ($\square$), and enstrophy cascade ($\triangle$) data. Despite the significant ϵ dependence, none of the curves intersect. This means that the Re dependence is ϵ independent.

symbols as shown in Fig. 10. Here h_S^* is plotted as a function of ϵ for three different values of Re. Although h_S^* increases as ϵ decreases, the curves never cross for the different Re. The reason for the general inverse relationship between ϵ and h_S^* is that as ϵ decreases, more detailed information is described by the symbols. The location of the dividers is important for the coarser partitions [40], but as the partition size decreases the results are not sensitive to this placement.

The compression ratio c_S is not a reliable estimate of h_S at the finer partitions, but it is still an indicator of the information content of the data streams and also shows the same decrease with Re [36]. The important point is that the general behavior of c_S and h_S^* is the same for partitions of all sizes.

Although the selection of a correct partition is trickier for the turbulent data since h_S^* and c_S are more sensitive to ϵ for the turbulence data, they show that the spatial disorder decreases with Re at each level of descriptive precision. An estimate can be made for the smallest size of the partition needed to capture the entire inertial range, based on the smallest eddy's characteristic velocity u_η [15,46,47]. Simple estimates show that the smaller partitions are fine enough to resolve a fluctuation of this magnitude for all Re. It is surprising that even a binary partition captures the main features: $h_S \propto -\log \text{Re} + \text{const}$. This suggests that one may fruitfully study turbulence just by looking at these 1s and 0s. Similar studies of complex systems such as the brain and heart and even turbulence [23,25–27] have also used very coarse partitions.

As an additional test of the validity of this coarse-graining approach, the decay rates calculated from the raw data using the autocorrelation method and with the mutual information were compared for various partitions (data not shown). The Re dependence was almost exactly the same, although there is a shift by a factor of $1/e$ for the mutual information method. Since the entropy is fundamentally connected to correlations, this is strong evidence for the validity of this coarse-graining approach.
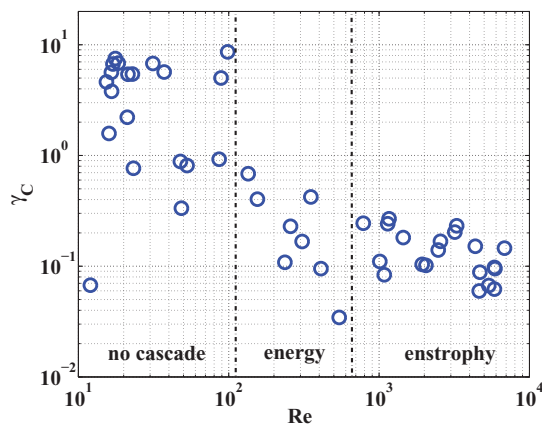

FIG. 9. (Color online) Spatial decay rate of mutual information for turbulence data using a partition where data are saved to the nearest integer. The decay rate here was calculated as the reciprocal of the area underneath the mutual information curve: $1/\sum_0^\infty I(\Delta)$. If this is large, then the correlations are weak. Dividing lines are shown that roughly separate the data into no-cascade, energy, and enstrophy regions, respectively. As Re increases, the correlations get stronger.
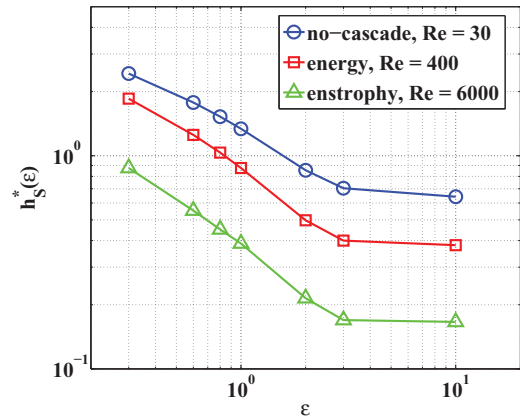
Entropy maximization is a familiar principle for solving a variety of problems and is a fundamental principle in equilibrium statistical mechanics [9,10]. In these problems, understanding the constraints is of paramount importance. In a turbulent system the constraints are correlations that span a wider range of scales as Re increases. These constraints, combined with some variational principle, may explain the decrease in h_S with Re observed here. Perhaps the organization of the cascade is the response to the system's effort to more efficiently transfer energy (enstrophy) between scales.

V. CONCLUSION

Treating turbulence as a message enables one to quantify the information content in the system through the entropy density h_S. Estimates show that h_S, a measure of disorder, is a decreasing function of Re at large Re. The cascade reduces spatial randomness by introducing correlations.

Cascades in turbulence are often thought to arise naturally because of the wide separation between the forcing scale and the dissipative scale, as well as because of some essential features of the Navier-Stoke's equation [15,48]. However, this may not be the only way of looking at the issue, just as in mechanics one can use either Newton's laws or a variational principle and reach the same conclusion. Perhaps the underlying reason for the development of a cascade can be connected to the decrease of h_S as Re increases.

ACKNOWLEDGMENTS

[1] J. P. Crutchfield, Nat. Phys. **8**, 17 (2012).

[2] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1964).

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[4] L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1962).

[5] H. P. Yockey, *Information Theory, Evolution, and the Origin of Life* (Cambridge University Press, New York, 2010).

[6] R. Shaw, Z. Naturforsch. A **36**, 80 (1981).

[7] J. P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[8] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, New York, 2009).

[9] E. T. Jaynes, Proc. IEEE **70**, 939 (1982).

[10] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[11] H. Kellay and W. I. Goldburg, Rep. Prog. Phys. **65**, 845 (2002).

[12] G. Boffetta and R. E. Ecke, Ann. Rev. Fl. Mech. **44**, 427 (2012).

[13] Yu. L. Klimontovich, *Turbulent Motion and the Structure of Chaos* (Springer, Dordrecht, 1991).

[14] A. N. Kolmogorov, Dokl. Akad. Nauk. SSSR **30**, 299 (1941); [Proc. R. Soc. Lond., Ser. A **434**, 9 (1991) (reprinted)].

[15] P. A. Davidson, *Turbulence: An Introduction for Scientists and Engineers* (Oxford University Press, Oxford, 2004).

[16] R. H. Kraichnan, Phys. Rev. Lett. **72**, 1016 (1994).

[17] B. I. Shraiman and E. D. Siggia, Nature **405**, 639 (2000).

[18] G. Falkovich, K. Gawedzki, and M. Vergassola, Rev. Mod. Phys. **73**, 913 (2001).

[19] A. Brandstater, J. Swift, H. L. Swinney, A. Wolf, J. D. Farmer, E. Jen, and J. P. Crutchfield, Phys. Rev. Lett. **51**, 1442 (1983).

[20] M. Yamada and K. Ohkitani, Phys. Rev. Lett. **60**, 983 (1988).

[21] D. Ruelle, Comm. Math. Phys. **87**, 287 (1982).

[22] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney, Phys. Rev. Lett. **103**, 094101 (2009).

[23] C. S. Daw, C. E. A. Finney, and E. R. Tracy, Rev. Sci. Instrum. **74**, 915 (2003).

[24] T. Schürmann and P. Grassberger, Chaos **6**, 414 (1996).

[25] A. J. Palmer, C. W. Fairall, and W. A. Brewer, IEEE Trans. Geo. Remote Sensing **38**, 2056 (2000).

[26] M. Lehrman, A. B. Rechester, and R. B. White, Phys. Rev. Lett. **78**, 54 (1997).

[27] M. Lehrman and A. B. Rechester, Phys. Rev. Lett. **87**, 164501 (2001).

[28] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, New York, 2002).

[29] J. P. Crutchfield and D. P. Feldman, Chaos **13**, 25 (2003).

[30] D. P. Feldman, Information Theory Lecture Notes [http://hornacek.coa.edu/dave/Tutorial/notes.pdf]

[31] D. P. Feldman and J. P. Crutchfield, Phys. Rev. E **67**, 051104 (2003).

[32] J. P. Crutchfield and D. P. Feldman, Phys. Rev. E **55**, R1239 (1997).

[33] H. L. Swinney and J. P. Gollub, Physica D **18**, 448 (1986).

[34] P. Grassberger, arXiv:physics/0207023v1.

[35] D. Salomon, *Data Compression: The Complete Guide* (Springer-Verlag, London, 2007).

[36] A. Baronchelli, E. Caglioti, and V. Loreto, Eur. J. Phys. **26**, S69 (2005).

[37] F. Kaspar and H. G. Schuster, Phys. Rev. A **36**, 842 (1987).

[38] W. Ebeling, T. Pöschel, and A. Neiman, *PhysComp96*, edited by T. Toffoli, M. Biafore, and J. Leao (New England Complex Systems Institute, Cambridge, MA, 1996).

[39] E. I. Olivares, R. Vazquez-Medina, M. Cruz-Irisson, and J. L. Del-Rio-Correa, *12th International Conference on Mathematical Methods in Electromagnetic Theory, Odesa, Ukraine* (IEEE, Odesa, 2008), p. 409.

[40] R. Steuer, L. Molgedey, W. Ebeling, and M. A. Jiménez-Montaño, Euro. Phys. J. B **19**, 265 (2001).

[41] W. Ebeling, T. Poschel, and K. F. Albrecht, Int. J. Bifurcat. Chaos **5**, 51 (1995).

[42] W. Li, J. Stat. Phys. **60**, 823 (1990).

[43] M. A. Rutgers, Phys. Rev. Lett. **81**, 2244 (1998).

[44] A. Belmonte, B. Martin, and W. I. Goldburg, Phys. Fluids **12**, 835 (2000).

[45] H. W. Wijesekera and T. M. Dillon, J. Geophys. Res. **102**, 3279 (1997).

[46] P. Gaspard and X. J. Wang, Phys. Rep. **235**, 291 (1993).

[47] X. J. Wang and P. Gaspard, Phys. Rev. A **46**, R3000 (1992).

[48] K. R. Sreenivasan and G. Stolovitzky, J. Stat. Phys. **78**, 311 (1995).