ELSEVIER

Physica A 285 (2000) 1–17

PHYSICA A

www.elsevier.com/locate/physa

Exotic statistical physics: Applications to biology, medicine, and economics

H. Eugene Stanley*

Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA

Abstract

This manuscript is based on four opening lectures, which were designed to offer a brief and somewhat parochial overview of some "exotic" statistical physics puzzles of possible interest to biophysicists, medical physicists, and econophysicists. These include the statistical properties of DNA sequences, heartbeat intervals, brain plaques in Alzheimer brains, and fluctuations in economics. These problems have the common feature that the guiding principles of scale invariance and universality appear to be relevant. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

One challenge of biology, medicine, and economics is that these sciences have no metronome in time and no evident architecture – crystalline or otherwise. As if by magic, out of nothing but *randomness* we find remarkably fine-tuned processes in time and fine-tuned structures in space. To understand this "miracle", we should put aside the human tendency to see the universe as a machine. Our task is to find out how, through pure (albeit, as we shall see, strongly correlated) randomness, we can arrive at the structures in biology we all know exist. These introductory lectures are not general, but rather concern four specific examples under investigation by cross-disciplinary researchers in Boston. I apologize in advance to those whose related work I do not include, due to constraints of time and space, and thank those whose research provided the basis of this short lecture summary: L.A.N. Amaral, P. Bernaola, S.V. Buldyrev, A. Bunde, P. Cizeau, L.C. Cruz, N.V. Dokholyan, A.L. Goldberger, P. Gopikrishnan, I. Grosse, S. Havlin, B.T. Hyman, P.Ch. Ivanov, T.H. Keitt, H. Leschhorn, Y. Lee, Y. Liu, P. Maass, M. Meyer, R.N. Mantegna, M. Meyer, C.-K. Peng, V. Plerou, B. Rosenow, M.A. Salinger, M. Simons, M.H.R. Stanley, R.H.R. Stanley, B. Urbanc, and C. Wyart.

* Fax: +1-617-3533783.
E-mail address: hes@bu.edu (H.E. Stanley).

PII: S0378-4371(00)00341-1

2. Is there hidden information in DNA sequences?

2.1. Noncoding "Junk" DNA

Human DNA has become a fascinating topic for physicists to study. One reason for this fascination is the fact that when living cells divide – which they are doing all the time – the DNA is replicated exactly. Why is this so interesting? Because approximately 95% of human DNA is called "junk", even by biologists who specialize in DNA. The Human Genome Project will soon be completed, having carefully identified 3 gigabases at a cost of approximately 3 gigadollars. The raw data in the Human Genome Project covers one million pages; every page is chocker-block full with 3000 letters, and each letter is seemingly drawn at random from a set of four letters. But only 90 megabases (3%) have any known meaning, even though they are copied faithfully at each cell division.

One practical task for physicists is simply to identify which sequences within the molecule are the coding sequences. If we could efficiently separate the junk from the rest, we could save time and money. It also would be an important first step toward finding and hopefully understanding the genetic components associated with the onset of three big gene-related killers: heart disease, cancer, and Alzheimer disease.

Another scientific interest is to discover why the junk DNA is there in the first place. Almost everything in biology has a purpose that, in principle, is discoverable. If something in biology does not have a purpose, we often can discover at least why it is there, i.e., where it came from. For example, the appendix has no purpose that we can discern, but at least we know why it is there.

2.2. Distinguishing coding and noncoding DNA

Noncoding DNA has statistical properties that distinguish it from coding DNA (see, e.g. Refs. [1–6] and references therein). In noncoding DNA, there are correlations between successive base pairs, correlations exhibiting a specific power-law form. These power-law correlations are very similar to those near a critical point or the percolation threshold. These same power-law correlations are *not* present in the coding DNA.

One big question in DNA research is whether there is some meaning to the order of the base pairs in DNA. One can easily map each DNA sequence onto a landscape (a "DNA walk") [7,8]. To do the map, we distinguish between the DNA base pairs that are single-ringed pyrimidines (made up of cytosine and thymine) and those that are double-ringed purines (made up of adenine and guanine). For our map we take an up-step for every pyrimidine and a down-step for every purine, producing a landscape that resembles a cartoon mountain range. Just as with all scale-free phenomena, we can measure the width of this mountain range as a function of the length scale over which we measure it. This is the analog of measuring the characteristic size of the structure of a polymer as a function of the length scale over which that size is studied by, e.g. X-ray scattering. One finds that the width depends on this length scale with a power-law relation [9,10].

If we look at a particular set of real data that are on the muscle gene of a rat, and make a landscape corresponding to the 20,000 base pairs that constitute this gene, we can immediately see that it is not an uncorrelated random walk. Instead there are huge stretches of "up" and huge stretches of "down".

It turns out that it is only the noncoding part of the DNA that displays this long-range power law correlation. The coding part of the DNA does not. We have devised a kind of virtual "machine" that walks along a DNA sequence and measures the correlation with some index α such that $\alpha = 0.5$ corresponds to coding (noncorrelated) base pairs and $\alpha > 0.5$ corresponds to noncoding (correlated) DNA. As the machine moves down the chromosome, the correlation signal dips whenever the machine locates a region of coding DNA [11]. The accuracy of finding coding regions is comparable to that of other methods – so long as the coding region is above ≈ 1000 base pairs in length.

2.3. Linguistic features of noncoding DNA

Why is this long-range power-law correlation present in the noncoding DNA? There is another system that displays long-range power-law correlations: language. All languages have some correlation. In English, for example, every "q" is followed by a "u". When I type my e-mail in English and make many errors, nearly everyone can still understand what I am trying to communicate – reflecting the fact that language has a built-in correlation. This correlation is in fact long-range, similar to that found in junk DNA [12].

Fifty years ago – before computers were available – Zipf analyzed the frequency of word use in a large written text. To construct his histogram, he counted the number of times each word appeared, ranked them in order of their frequency of use, and graphed the ranked data on log–log paper. The data were remarkably linear with a slope of about -1. This is sometimes called Zipf's law of language texts, a law that has been confirmed by numerous people and for many different human languages (and even for the size of business firms [13]). Today, Zipf's law can be demonstrated using virtually any personal computer. This remarkable law implies the presence of a hierarchical ordering in the structure of languages; if we want to move a word up in the word-use hierarchy, we know exactly how much to increase its frequency of use in order to change its rank order.

Mantegna and collaborators [14] used Zipf's technique to analyze noncoding DNA – with one "word" of noncoding DNA defined as a subsequence with $n = 6$ base pairs. A graph of 4000 such words yields the same linearity and slope as the Zipf graph. In marked contrast, when they analyze the coding regions, the graph is not at all linear. If we change the log–log plot to a linear-log plot, the coding data become linear. Thus, the coding and noncoding regions of the DNA obey different statistical laws differently when analyzed in this linguistic fashion. That does not necessarily imply that junk DNA is a language and coding DNA is not. It does suggest that we might study DNA using the same statistical techniques that have been used to analyze language, and that if information is contained in the junk DNA, it is less likely to be in the form of a code and more likely in the form of a structured language.

3. The challenge of heartbeat interval time series

3.1. The nonstationarity problem

The time differences between successive heartbeats – the interbeat intervals – can easily be measured with millisecond accuracy. The traditional method of taking a patient's pulse measures only the *average* interbeat interval over a period of, typically, a minute or less. The *fluctuations* around that average, however – supplied by a cardiogram – constitute a body of data that is of interest to statistical physicists. But when statistical physicists begin to study these admirably accurate data, almost immediately they encounter a roadblock. Unlike similar data associated with critical phenomena, these data have the property of nonstationarity. This means that the statistical properties of those interbeat intervals are not constant in time; they are not served up neatly as independent, identically distributed (i.i.d.) random variables.

Thus analyzing the heartbeat is a challenge because the statistical properties of heartbeat intervals are not stationary functions of time. The heartbeat signal is anything but stationary. As a function of time, it changes its statistical properties constantly. At one point in time it has one average, and a little later it has another average [15]. At one point in time it has one standard deviation or fluctuation, and a little later it has another standard deviation or fluctuation [16]. At one point in time it has one fractal Hurst exponent, and a little later it has another fractal Hurst exponent [17].

3.2. Detrended fluctuation analysis

In order to get around this roadblock, there are a number of techniques we can use. One of them is detrended fluctuation analysis (DFA), which was developed by Peng while working on his Ph.D. at Boston University [16,18]. We take a typical time-dependent interbeat interval function – which is clearly nonstationary. We plot that function and then divide the entire time-interval into "window boxes" of, e.g. 100 heartbeats. Within each window box we calculate the local linear trend in that box. These local trends will differ from each other. If we subtract the actual function from each local trend we get something that fluctuates much less, because the trend is subtracted out. By forming a function that is the original heartbeat and subtracting the local trend we can calculate a fluctuating quantity that has the same statistical properties as the function we are trying to understand. This technique helps clarify the behavior of nonstationary time series for healthy and unhealthy heartbeat intervals.

3.3. Interbeat anticorrelations

When we apply DFA to interbeat intervals, we find something surprising: the interbeat intervals of a healthy heart have remarkably long-range "memory". Specifically, when we analyze a sequence of $\approx 60,000$ beats over a 15 h period, we see that these fluctuations are around a characteristic average value and that they are fairly big – some at least 20% the size of the average. Remarkably, one finds that these fluctuations are *anticorrelated* in time [16,19].

This simpler interbeat pattern seems to leave the unhealthy heart more vulnerable to trauma. Lacking interbeat correlations on a variety of scales, it becomes analogous to a suspension bridge with only a simple primary resonance period – a resonance period that, given the right wind velocity, allows the structure to sway at progressively increasing amplitudes until it pulls itself apart. This actually happened to a bridge 60 years ago in the United States, near Tacoma, Washington. Since that time we have been careful to design into our bridges a variety of resonance periods on a variety of scales.

It seems possible that, in a similar manner, a healthy heart could perhaps require its own kind of scale invariance – fluctuations on a variety of scales – in its interbeat intervals.

3.4. The wavelet method

The wavelet method, pioneered by mathematicians, has been systematically applied to heartbeat analysis by Ivanov [20] and collaborators. If we take two heartbeat interval data sets, one from a healthy heart and one from an unhealthy heart, an untrained eye might not see the difference between the healthy and the unhealthy. How do we devise a mechanism such that anyone can distinguish the difference? In wavelet analysis, we take, e.g. 30 min of data, choose a scale over which to examine the fluctuations in the signal, and select the amplitude of some statistical quantity that we calculate. In the unhealthy heartbeat, we find regularly repeating patterns of a characteristic scale. In the healthy heartbeat the patterns are nonrepeating and scale-free. Wavelet analysis sometimes requires fewer than 30 min of data in order to distinguish between healthy and unhealthy heartbeat patterns. A much smaller data set, say 3 min, is sometimes sufficient to detect the difference.

3.5. Multifractals

In a nonstationary signal, the Hurst exponent can have a variety of values. In order to see this, we color-code the values of the Hurst exponent. Small values are at the red end of the color spectrum and large values are at the blue end of the spectrum. In a healthy heart, the color spectrum of the Hurst exponents constitute a “Joseph’s-coat-of-many-colors” display – there are a very large number of different Hurst exponents; the value fluctuates wildly. Since there are so many fractal dimensions, it is convenient to record these in a quantitative fashion called multifractal analysis. We encounter multifractals in many different areas of modern statistical physics.

Take a red filter, place it on top of the color spectrum to separate out the red hues, and determine the fractal dimension of the red pattern. Repeat the procedure with a yellow filter, a blue filter, and several others for perhaps a total of 10 different color, determining the 10 different fractal dimensions (Fig. 1).

On the other hand, the color spectrum of the Hurst exponents of an unhealthy heart's interbeat intervals covers a much smaller region. It is immediately apparent to the eye that there is a real difference.

4. The tragedy of Alzheimer disease

As we age, the probability that we will ultimately die of Alzheimer disease increases sharply. It is a terrible disease about which we know very little. We are beginning to believe that it may be associated with the appearance of aggregates of amyloid β protein molecules called plaques, which appear in the brains of Alzheimer patients statistically more frequently than in the brains of healthy people, but we do not know how these plaques are formed or their morphology.

We are beginning to apply our understanding of scale-free phenomena to this problem of plaque shape and formation. Using confocal microscopy and computer imaging [21–25], we obtain a three-dimensional portrait of a plaque. It is a remarkable picture, because it appears to be full of holes; indeed, it has a sponge-like porosity. This is significant because if plaques are porous they can be understood using aggregation models that come from the theory of scale-free phenomena, and these aggregation models in turn give us some insight into how the plaque is formed. Through these holes traverse the nerve cells that are perhaps damaged in the course of Alzheimer disease.

We can measure quantitatively various correlation functions, such as the familiar $g(r)$ that we have measured in statistical physics since Van Hove's time. For example, if we look at a picture of a 3 mm $\times$ 3 mm section of the brain, we see lamina about 300 μm in diameter [26,27]. By quantifying the correlations $g(x, y)$ in both x and y directions (see Fig. 2), Buldyrev and co-workers discovered the existence of little columns (resembling little "polymers" of about 11 "monomers") positioned at right angles to the lamina [28]. At present, we are studying these columns as they occur in both the healthy brain and the Alzheimer brain [28]. The same sort of correlation analysis that we had applied to plaques can also be applied to neurons. We found that the morphology of specific dendrites is disrupted in the Alzheimer brain [23].

In summary, statistical physics approaches to Alzheimer disease have yielded three relatively firm results thus far:

(i) A log-normal distribution of senile plaque size [29].

(ii) The plaque is a remarkably porous object [21], which can be quantified by a model characterized by both aggregation and *dis*aggregation [21,24,25].

(iii) Quantifiable neuron architecture that suggests the existence of microcolumns positioned at right angles to known lamina [28].

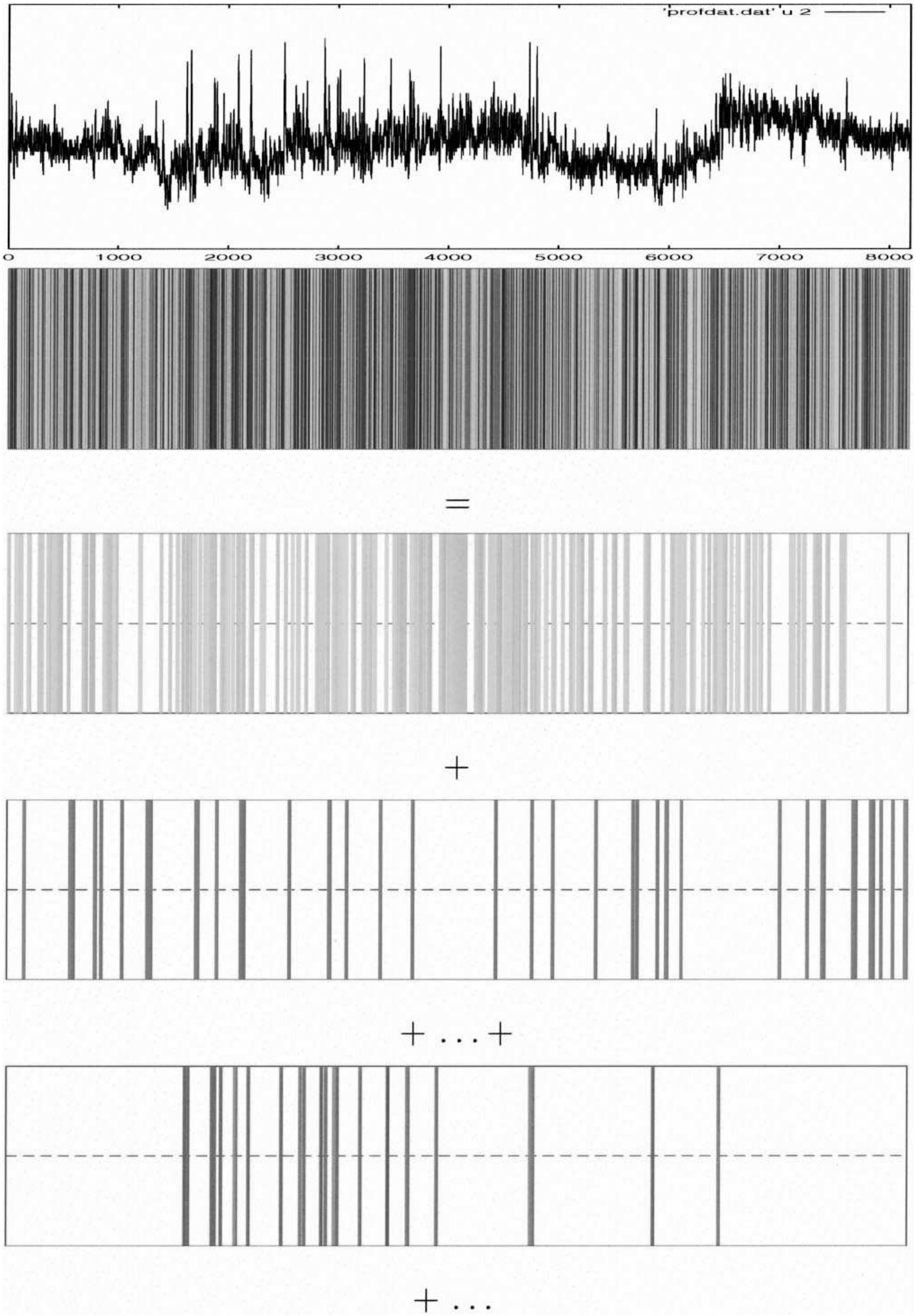

Fig. 1. Decomposition of a heartbeat interval time series into subsets, each characterized by a different Hurst exponent h. Here the Hurst exponent is color coded by the rainbow, so that each subset has a different color. The "multifractal spectrum" $D(h)$ is a function giving the fractal dimension D of each of the different subsets. This figure is kindly contributed by Zbigniew Struzik.

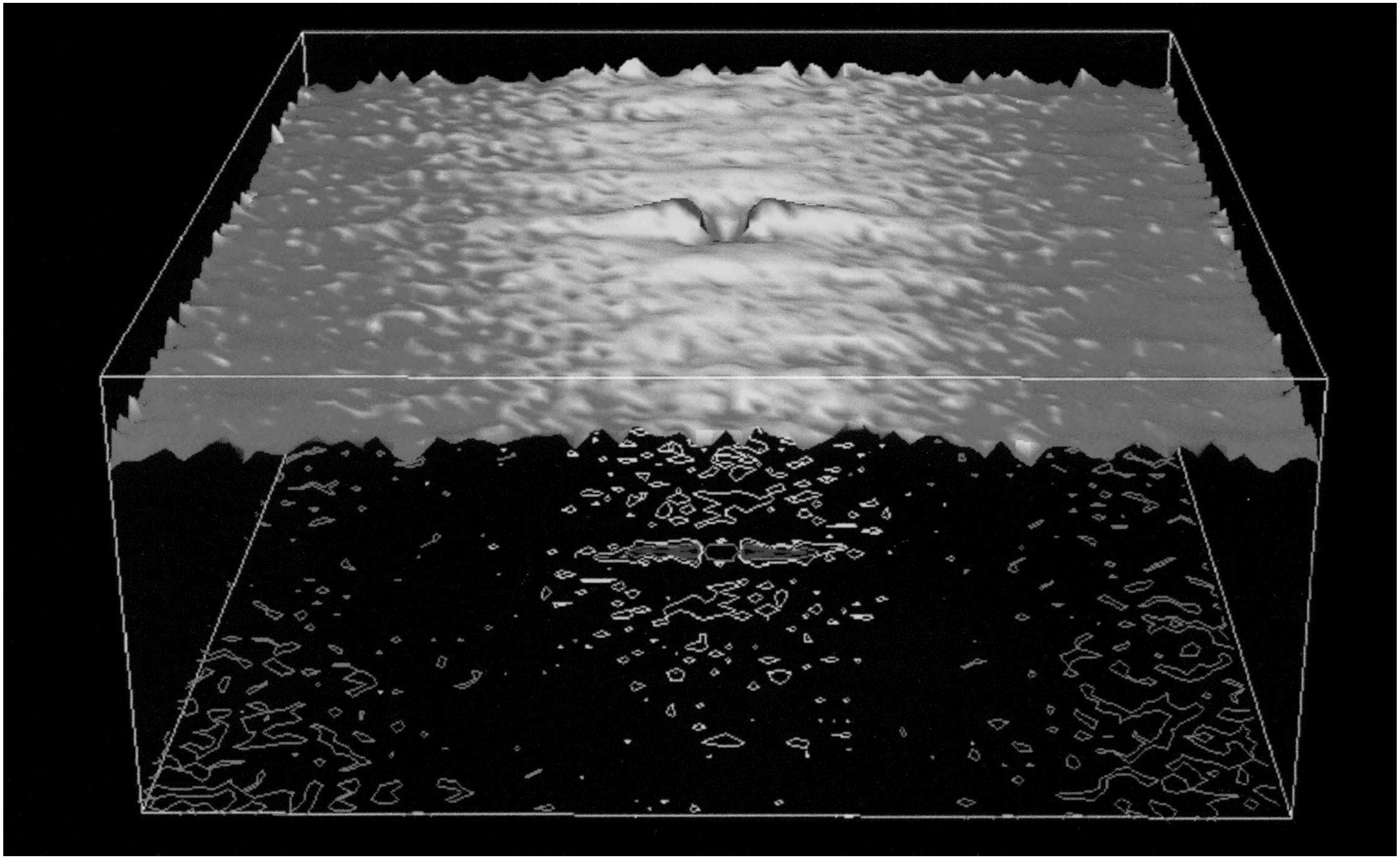

Fig. 2. The neuronal density in the human brain cortex, indicated in green (low density) and red (high density). This result was obtained by Buldyrev et al. [28] by applying quantiative methods of statistical physics to brain architecture. The extended region of high density suggests the presence of vertical chains of cells forming repeating units called neuronal microcolumns. Such microcolumns appear to contain about 11 neurons and have a periodicity of about 80 microns – a periodicity that is disrupted in Alzheimer disease and in another neurodegenerative disease termed Lewy body dementia. This figure is kindly contributed by S.V. Buldyrev.

5. The intrigue of econophysics

Very recently, scale-free phenomena have been used to describe economic systems [30]. This approach was pioneered by Mandelbrot almost 40 years ago. In 1963 he introduced the Lévy distribution as an example of a scale-invariant distribution that seemed to fit commodity (cotton) data that were available at the time [31]. Since that time, working with larger databases, we have found that the Lévy distribution does not apply to noncommodity data.

5.1. What is the question that physicists find interesting?

Statistical physicists, myself included, are extremely interested in fluctuations. In the field of economics, we find ourselves surrounded by fluctuations – were it not for economic fluctuations, economists would have no work to do. Moreover, these fluctuations are thoroughly documented in large databases. If we take a graph of the S&P 500 index comprising the 500 largest US firms according to market capitalization, and place it above a graph of an uncorrelated biased random walk with the same overall bias, at first glance they seem almost identical. When we look closer, however, we notice the graph of the S&P 500 has occasional large fluctuations (e.g. the huge drop that took place on Black Monday in October, 1987 – when most world markets lost 20–30% of their value over a period of 1–2 days). We do not see this kind of large fluctuation in the biased random walk graph because the probability of taking a very large number random steps in the same direction – which would be necessary for a large fluctuation – is exponentially small.

So the question we physicists find so interesting is "how do we quantify these economic fluctuations?"

5.2. Why should physicists want to quantify economic fluctuations?

One reason physicists might want to quantify economic fluctuations is in order to help our world financial system avoid "economic earthquakes". Not too long ago, a particular financial firm overextended itself and was unable to pay its loans – and other firms reached into their pockets and gave them the money they needed. This was not a charitable gesture, but one made out of self-interest. The other firms were aware that, in the world economic system, everything depends on everything else. If one major financial firm is unable to cover its debts, perhaps investors in other similar firms will begin to lose their confidence and withdraw their money – and then the other firms would be unable to pay their debts, and the phenomenon would propagate across the world. Indeed, bailing out the company in trouble was a prudent action.

We want to be able to quantify fluctuations in order to discover how to insure against this kind of economic earthquake. When a localized economic earthquake occurred several years ago in Indonesia, everyone in the region was affected. It caused the bankers to lose money, but it also caused a rice shortage in the country – and some people actually starved.

A second reason physicists might want to quantify economic fluctuations is that economic fluctuations and the clustering of economic fluctuations has become very much a part of our modern experience. The topic is in the newspapers, on television, and in barbershop conversations. Almost anyone can understand *what* we are doing even if they do not understand the specifics of *how* we are doing it.

A third reason is that it is a new topic, one that the American Physical Society journals initially resisted recognizing as legitimate. Only after the rapid expansion of the field (through Physica A, European Physical Journal, and other journals) and articles on page 1 of *The Wall Street Journal* did the APS allow econophysics articles to be published in Phys. Rev. E and Phys. Rev. Lett. Be forewarned that if you submit articles on econophysics to physics journals, economists will be among the referees – and the turn-around time for referee reports in that research community seems to be an order of magnitude longer than in physics circles (on the other hand, their referee reports are usually very polite, thorough, and gracefully written).

A fourth reason is because of cross-disciplinary interests. In the field of turbulence, we may find some crossover with certain aspects of financial markets. Colloquially, people say "It's been a turbulent day on Wall Street". But there could also be some serious analogies between the two fields. In 1994, it was proposed that in the same way that a glass of water, when stirred, dissipates a large influx of energy on successively smaller scales – and we attempted to quantify that dissipation – so also in economics, fluctuations driven by the influx of information (the energy source in economics) is dissipated on successively smaller scales [32,33]. This is hardly a rigorous argument, but we do notice that when highly publicized information that people consider important spills through the news channels – the resignation of Boris Yeltsin as President of Russia, for example – the financial market response is sharp and immediate (in this case, it meant a 23% drop in the Russian stock market). It is obvious that important information causes big changes, but even seemingly unremarkable bits of news can drive price fluctuations, e.g. the news that Apple Computer's stock, though moving steadily upward, might not peak *quite* as high as initially anticipated often causes the price to fluctuate down. Most people want to make a trade because they believe they have some information that gives them an advantage in the trade. The analog to price changes in the stock market is velocity changes in turbulence, but this analogy is only qualitative, not quantitative [33], as the actual form of the functions that describe fluctuations are quite different.

5.3. How can physicists quantify fluctuations in finance?

What do we do when we carry out research on economic fluctuations? Our approach has been to use our experience in critical phenomena research and assume that when we see fluctuations, correlations may be present. If it were known that similar correlations are present in economic fluctuations, each bit of information would perhaps itself cause further fluctuations. There would be a feedback event by virtue of the fact of this correlation. Using this observation, perhaps economic fluctuations are not correlated after all. But since there are lines of reasoning both for and against correlations in economic fluctuations, why not try this approach? We did it using the S&P 500 stock index over a 13-year period. The basic quantity is the price change $G(t)$ defined over a certain time horizon or "window". If we open a window of a time-width Δt, and ask how much the price changes, the answer will depend on the size of the window.

What sort of correlations can we find in $G(t)$? On a log-linear graph, the data for the autocorrelation in $G(t)$ fall approximately on a straight line, which means the decay is exponential. The slope tells us something about the characteristic decay time, the time in which the autocorrelation drops by a ratio $1/e$ to what it was about τ minutes earlier. We find $\tau \approx 4$ min, and after ≈ 30 min the autocorrelation function disappears into the noise [34,35].

Suppose we look not at the correlation of the $G(t)$ values, but the correlation of the *absolute value* of the $G(t)$ values. Here we ignore whether the price change is up or down and simply measure the volatility. If we make a log–log plot of the same autocorrelation function of the absolute values of $G(t)$, we see an approximate linearity over 1.5–2.0 decades with a slope ≈ -0.3 [35].

This behavior of the autocorrelation in both $G(t)$ values and in the absolute value of $G(t)$ has already been discussed in the economics literature. What we have added to the discussion is that we have tested these results using a variety of different methods, and uncovered an interesting crossover between two different power laws.

We have calculated the Fourier transform of the autocorrelation function, i.e., the power spectrum, and we find a straight line only up to frequencies of about 10^{-3} – corresponding to a characteristic time on an order of magnitude of approximately one day. There is another area of linear behavior corresponding to larger frequencies, i.e., shorter times. It would appear that there are two different power laws, one shorter than $\approx$ 1–2 days ("high-frequency data") and one longer ("low-frequency data").

In order to verify this behavior, we created a random walk of these fluctuations and then analyzed the result using the detrended fluctuation analysis method. Once again we find two power laws – with characteristic exponents that are related algebraically to the characteristic exponents of the power-spectrum data.

To summarize: the fluctuations themselves have no correlations of any interest, but the absolute values of the fluctuations have long-range power-law correlations. Since the fluctuations have no correlations, we dump them on the floor (so to speak) and pick them up in random order. Since we have destroyed the order of the numbers, we cannot compute correlations, but we can make a histogram of the $G(t)$ values.

Mandelbrot did this in 1963 with the price fluctuations of cotton [31]. Some of them were on a daily time scale and some were on a monthly time scale. He found that on log–log paper the cumulative distribution of these price fluctuations was approximately linear. The slope of this line (over approximately one decade) was ≈ 1.7, consistent with the possibility that those price fluctuations are described by a Lévy stable distribution. These results enjoyed widespread acceptance until recently, when better data became available.

Mandelbrot used data sets of approximately 2000 data points. Mantegna and I acquired data sets of the S&P 500 stock index comprising ≈ 1 million data points [34]. In studies of critical phenomena, an increase in data points usually means, at most, small corrections to the previous result. This appears not to be the case in price fluctuations. Mantegna found on a log-linear graph that the data fit neither a Gaussian nor a Lévy distribution beyond a few standard deviations, but are intermediate. This is good news, since it implies that "rare events" are much less likely than if the data tails followed a Lévy distribution.

These data exhibit a kind of scale invariance. This means that, if I know the fluctuations on one time horizon – e.g. 1 min – and I know the exponent α of the distribution, I already know something about the fluctuations on a much longer time scale, i.e., up to approximately 3 trading days. How robust is this scale invariance? Skeltorp did this same calculation for the Norwegian stock market, which is much smaller (5%) and less active than the US, and found virtually the same behavior [36].

Plerou, Gopikrishnan, and collaborators acquired a database that recorded every transaction in all major US stock markets for two years [37–39]. They extracted 40 million data points and analyzed them. They found that, if they constructed the same sorts of histograms that Mantegna utilized, there was no indication of a Lévy regime at all. The slope α is not ≈ 1.4, as Mantegna found, or ≈ 1.7, as Mandelbrot found, but ≈ 3.0. To check this, they did a "best-fit" straight line and found the value of α for every stock. We constructed a histogram of α-values and found that the typical data for all the stocks tended to scatter around $\alpha \approx 3.0$, with almost all 1000 values falling between $\alpha = 2$ and 5. If we average all the firms together on a final graph, we see that out to 100 standard deviations the data are linear with a slope 2.84 for the positive tail and 2.73 for the negative tail, with error bars that may bring those values closer to 3. This means that events that are rare by 8 orders of magnitude – events that occur once in every 100 million trades – fall on the same curve as everyday events.

Individual companies seem to exhibit $\alpha = 3$ behavior all the way from one standard deviation out to 100 standard deviations. The stock averages appear to be $\alpha < 2$ (Lévy regime) out to 2–3 standard deviations, and then resemble a "power-law truncated" Lévy distribution with $\alpha = 3$.

There is currently a great deal of work that is being done on cross-correlations between companies. That work would be scientifically interesting (as are cross-correlation functions in almost any topic area), as well as have an obvious practical application that would interest a great many people [40,41]. Specifically, their approach is to compare the properties of cross-correlation matrices calculated from price fluctuations to that of a matrix with random entries (a random matrix). These studies determine precisely the amount of nonrandom correlations present in the correlation matrix. In particular, one can identify modes (eigenvectors of the correlation matrix) of correlations which are stable in time, and determine their relative stability (Fig. 3).

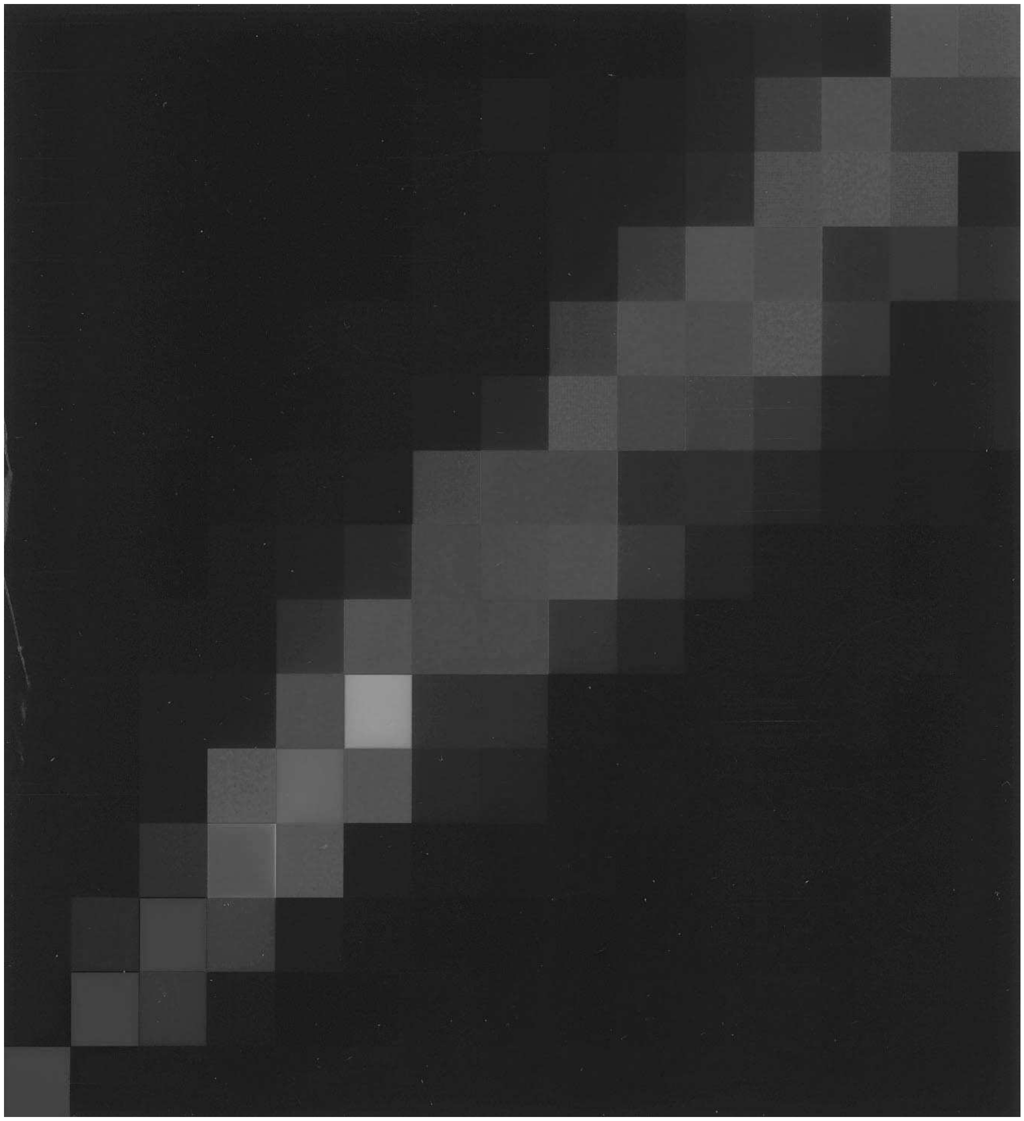

Fig. 3. Stability in time of the eigenvectors of the correlation matrix (calculated from 30 min returns), that deviate from random-matrix bounds [40,41]. Two partially overlapping time periods A, and B, of four months each were analyzed, January 1994–April 1994 and March 1994–June 1994. Each of the 225 squares has a rainbow color proportional to the scalar product ("overlap") of the largest 15 eigenvectors of the correlation matrix in period A with those of the same 15 eigenvectors from period B. Perfect stability in time would imply that this pixel-representation of the overlaps has ones (the red end of the rainbow spectrum) in the diagonal and zeros (violet) in the off-diagonal. The eigenvectors are shown in inverse rank order (from smallest to largest), and we note that the pixels near the lower right corner have colors near the red end of the spectrum corresponding to the fact that the largest 6–8 eigenvectors are relatively stable; in particular, the largest 3–4 eigenvectors are stable for very long periods of time. Further, we note that the remainder of the pixels are distributed toward the violet end of the spectrum, corresponding to the fact that the overlaps are not statistically significant, and corroborating the finding their corresponding eigenvalues are random [40,41]. This figure is kindly contributed by P. Gopikrishnan and V. Plerou.

5.4. Quantifying fluctuations of economic organizations

How do organizations grow and shrink? The traditional approach in economics is equivalent to a cluster approximation in critical phenomena [42]. The crudest approach is the mean-field approximation, in which you represent the interaction of all the systems with one spin using some effective field. This can be improved by taking not one spin as your cluster, but by taking a sequence of larger and larger spins as your cluster, treating exactly the interactions within that cluster, and then treating by mean-field approximation the interaction of all the other spins with the spins in that cluster. That is approximately what is done in the theory of the firm in which one divides the entire economy into sectors – e.g. manufacturing, food, automotive, and computers – treating the firms within each sector as strongly interacting with each other, and assuming that firms between sectors have no interaction at all.

The problem in critical phenomena is that no matter how much time we spend working with larger and larger clusters, we always end up with exponents that are wrong. Near a critical point, everything interacts with everything else, either directly or indirectly. So also in the economy.

Suppose the news media announce that Ford has a serious design defect in its vehicles. People react by buying GM cars, and GM experiences a sales increase because of Ford's downturn in sales. Suppose the sales increase in GM is so great that more workers are hired to help manufacture the cars, workers who then flood into the McDonald's across the street from the GM plant during lunchtime. So interactions take place not just between firms within one sector (automotive), but also between firms in different sectors (automotive and food).

This is a kind of indirect interaction. The first company goes down and, as a result, a McDonald's located near another company in a different city goes up, a kind of second-order interaction. Just as in the situation near a critical point, all these indirect interactions are significant [43]. Since there are $\approx 10^4$ business firms in the US, to be completely honest in our inquiry we must consider all $\approx 10^8$ interactions between these firms. Since that task is much too complex, we go back to our histograms.

We take a database of all 10^4 firms and calculate the growth rate of each firm, i.e., the size of the firm this year (using some measurement, often sales) divided by the size of the firm last year. We put these calculations into 10 smaller data "bins" as a function of the size of the firm. We know intuitively that growth rates will vary inversely with the size of the firm. Larger firms will have a narrower distribution of growth rates and smaller firms will have a wider distribution of growth rates. How does the width depend on the firm's size? If we calculate the width and plot it as a function of firm size on log–log paper, we get an approximate straight line that extends all the way from tiny kilodollar firms to gigantic teradollar firms – 9 orders of magnitude and a very robust scaling [44,45]. The slope is approximately 0.2. This result seems to be universal in that if we change the standard of measurement for firm size – using number of employees instead of annual sales, for example – the results are the same: same approximate straight line and same slope 0.2.

Takayasu did a parallel firm-size analysis on firms in other countries, and got results that agree with our findings [46]. On a suggestion from Jeffrey Sachs, we did a parallel analysis with the economies of entire countries, and got the same results, the same tent-shaped distribution for all the countries in the world over 42 years, where the size of a country was measured by its GDP [47,48].

Plerou and collaborators found similar results for the changes in size of university research budgets in the US [49]. In this situation, we researchers are the sellers and the granting institutions are the customers. Keitt did a similar analysis on changing bird populations, which fluctuate in size from year to year, with similar results [50]. Thus, it appears that the pillars of scaling and universality may be relevant to a range of "social" phenomena [51].

Acknowledgements

I thank all those under whose tutelage I learned what little I understand of statistical physics. These include the students, postdocs, and faculty visitors to our research group with whom I have enjoyed the pleasure of scientific collaboration. Also, I thank those who supported this research, especially NSF, NIH, ONR, NATO, BP Amoco, the NIH National Center for Research Resources (grant P41 13622), and the Massachusetts Alzheimer Disease Research Center Brain Bank (NIA AG05134, Dr. E.T. Hedley-Whyte, director) for tissue samples.

References

[1] N. Dokholyan, S.V. Buldyrev, S. Havlin, H.E. Stanley, Distribution of base pair repeats in coding and noncoding DNA sequences, Phys. Rev. Lett. 79 (1997) 5182–5185.
[2] R.H.R. Stanley, N.V. Dokholyan, S.V. Buldyrev, S. Havlin, H.E. Stanley, Clumping of identical oligonucleotides in coding and noncoding DNA sequences, J. Biomol. Struct. Des. 17 (1999) 79–87.
[3] N. Dokholyan, S.V. Buldyrev, S. Havlin, H.E. Stanley, Distributions of dimeric tandem repeats in noncoding and coding DNA sequences, J. Theor. Biol. 202 (2000) 273–282.
[4] I. Grosse, H. Herzel, S.V. Buldyrev, H.E. Stanley, Species independence of mutual information in coding and noncoding DNA, Phys. Rev. E 61 (2000) 5624–5629.
[5] P. Bernaola-Galvan, I. Grosse, P. Carpena, J.L. Oliver, H.E. Stanley, Identification of DNA coding regions using an entropic segmentation method, Phys. Rev. Lett. 84 (2000).
[6] D. Holste, I. Grosse, H.E. Stanley, S.V. Buldryev, H. Herzel, Optimization of protein coding measures using positional dependence of nucleotide frequencies, J. Theor. Biol. 206 (2000).
[7] M.Y. Azbel, Random two-component one-dimensional Ising model for heteropolymer melting, Phys. Rev. Lett. 31 (1973) 589–592.
[8] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Long-range correlations in nucleotide sequences, Nature 356 (1992) 168–171.
[9] S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, M.E. Matsa, C.-K. Peng, M. Simons, H.E. Stanley, Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis, Phys. Rev. E 51 (1995) 5084–5091.
[10] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, H.E. Stanley, Fractals in biology and medicine: from DNA to the heartbeat, in: A. Bunde, S. Havlin (Eds.), Fractals in Science, Springer, Berlin, 1994, pp. 48–87 (Chapter 2).
[11] S.M. Ossadnik, S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, C.-K. Peng, M. Simons, H.E. Stanley, Correlation approach to identify coding regions in DNA sequences, Biophys. J. 67 (1994) 64–70.
[12] A. Schenkel, J. Zhang, Y.C. Zhang, Long-range correlations in human writings, Fractals 1 (1993) 47.

[13] M.H.R. Stanley, S.V. Buldyrev, S. Havlin, R. Mantegna, M.A. Salinger, H.E. Stanley, Zipf plots and the size distribution of firms, Econ. Lett. 49 (1995) 453–457.
[14] R.N. Mantegna, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, H.E. Stanley, Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics, Phys. Rev. E 52 (1995) 2939–2950.
[15] P. Bernaola-Galván, P.Ch. Ivanov, L.A.N. Amaral, A.L. Goldberger, H.E. Stanley, Scale invariance in the nonstationarity of physiological signals, preprint.
[16] C.K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, in: L. Glass (Ed.), Proceedings of the NATO Dynamical Disease Conference, Chaos 5 (1995) 82–87.
[17] P.Ch. Ivanov, L.A.N. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z. Struzik, H.E. Stanley, Multifractality in human heartbeat dynamics, Nature 399 (1999) 461–465.
[18] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (1994) 1685–1689.
[19] C.K. Peng, J. Mietus, J. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, Non-equilibrium behavior of the heartbeat, Phys. Rev. Lett. 70 (1993) 1343–1346.
[20] P.Ch. Ivanov, M.G. Rosenblum, C.-K. Peng, J. Mietus, S. Havlin, H.E. Stanley, A.L. Goldberger, Scaling behavior of heartbeat intervals obtained by wavelet-based time-series analysis, Nature 383 (1996) 323–327.
[21] L. Cruz, B. Kutnjac-Urbanc, S.V. Buldyrev, R. Christie, T. Gómez-Isla, S. Havlin, M. McNamara, H.E. Stanley, B.T. Hyman, Aggregation and disaggregation of senile plaques in Alzheimer disease, Proc. Nat. Acad. Sci. 94 (1997) 7612–7616.
[22] H.E. Stanley, S.V. Buldyrev, L. Cruz, T. Gomez-Isla, S. Havlin, B.T. Hyman, R. Knowles, B. Urbanc, C. Wyart, Statistical physics and Alzheimer's disease, Physica A 249 (1998) 460–471.
[23] R.B. Knowles, C. Wyart, S.V. Buldyrev, L. Cruz, B. Urbanc, M.E. Hasselmo, S. Havlin, H.E. Stanley, B.T. Hyman, Plaque-induced neural network disruption in Alzheimer's disease, Proc. Nat. Acad. Sci. 96 (1999) 5274–5279.
[24] B. Urbanc, L. Cruz, S.V. Buldyrev, S. Havlin, B.T. Hyman, H.E. Stanley, Dynamic feedback in an aggregation-disaggregation model, Phys. Rev. E 60 (1999) 2120–2126.
[25] B. Urbanc, L. Cruz, S.V. Buldyrev, S. Havlin, H.E. Stanley, B.T. Hyman, Dynamics of plaque formation in Alzheimer disease, Biophys. J. 76 (1999) 1330–1334.
[26] V.B. Mountcastle, The columnar organization of the neocortex, Brain 120 (1997) 701–722.
[27] W.Y. Ong, L.J. Garey, Neuronal architecture of the human temporal cortex, Anat. Embryol. 181 (1990) 351–364.
[28] S.V. Buldyrev, L. Cruz, T. Gomez-Isla, S. Havlin, H.E. Stanley, B. Urbanc, B.T. Hyman, Description of microcolumnar ensembles in association cortex and their disruption in Alzheimer and Lewy body dementias, Proc. Nat. Acad. Sci. 97 (2000) 5039–5043.
[29] B.T. Hyman, H.L. West, G.W. Rebeck, S.V. Buldyrev, R.N. Mantegna, M. Ukleja, S. Havlin, H.E. Stanley, Quantitative analysis of senile plaques in Alzheimer disease: observation of log-normal size distribution and of differences associated with Apolipoprotein E genotype and Trisomy 21 (Down syndrome), Proc. Nat. Acad. Sci. 92 (1995) 3586–3590.
[30] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations & Complexity in Finance, Cambridge University Press, Cambridge, 1999.
[31] B.B. Mandelbrot, The variation of certain speculative prices, J. Business 36 (1963) 394.
[32] R.N. Mantegna, H.E. Stanley, Turbulence and financial markets, Nature 383 (1996) 587.
[33] R.N. Mantegna, H.E. Stanley, Turbulence and exchange markets, Nature (Scientific Correspondence) 383 (1996) 587–588.
[34] R.N. Mantegna, H.E. Stanley, Scaling behavior in the dynamics of an economic index, Nature 376 (1995) 46–49.
[35] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, The statistical properties of the volatility of price fluctuations, Phys. Rev. 60 (1999) 1390–1400.
[36] J. Skjeltorp, Scaling in financial markets, Physica A 283 (2000) 486.
[37] P. Gopikrishnan, V. Plerou, X. Gabaix, H.E. Stanley, Statistical properties of share volume traded in financial markets, Phys. Rev. E (Rapid Communications) 62 (2000), to appear.
[38] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, X. Gabaix, H.E. Stanley, Diffusion and economic fluctuations, Phys. Rev. E. (Rapid Communications) 62 (2000), to appear.

[39] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, Scaling of the distribution of price fluctuations of individual companies, Phys. Rev. E 60 (1999) 6519.
[40] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (1999) 1469.
[41] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Universal and non-universal properties of cross-correlations in financial time series, Phys. Rev. Lett. 83 (1999) 1471.
[42] J.S. Smart, Effective Field Theories of Magnetism, Saunders, Philadelphia, 1966.
[43] H.E. Stanley, Scaling, universality, and renormalization: three pillars of modern critical phenomena, Rev. Mod. Phys. 71 (1999) S358.
[44] M.H.R. Stanley, L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, Scaling behavior in the growth of companies, Nature 379 (1996) 804.
[45] L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, Scaling behavior in economics: I. Empirical results for company growth, J. Phys. I France 7 (1997) 621.
[46] H. Takayasu, K. Okuyama, Country dependence on company size distributions and a numerical model based on competition and cooperation, Fractals 6 (1998) 67.
[47] D. Canning, L.A.N. Amaral, Y. Lee, M. Meyer, H.E. Stanley, A power law for scaling the volatility of GDP growth rates with country size, Econ. Lett. 60 (1998) 335.
[48] Y. Lee, L.A.N. Amaral, D. Canning, M. Meyer, H.E. Stanley, Universal features in the growth dynamics of complex organizations, Phys. Rev. Lett. 81 (1998) 3275.
[49] V. Plerou, L.A.N. Amaral, P. Gopikrishnan, M. Meyer, H.E. Stanley, Similarities between the growth dynamics of university research and of competitive economic activities, Nature 400 (1999) 433.
[50] T. Keitt H.E. Stanley, Scaling in the dynamics of North American breeding-bird populations, Nature 393 (1998) 257.
[51] L.A.N. Amaral, S.V. Buldyrev, S. Havlin, M.A. Salinger, H.E. Stanley, Power law scaling for a system of interacting units with complex internal structure, Phys. Rev. Lett. 80 (1998) 1385.