Simple Models of Complex Systems

Deepak Dhar
Department of Theoretical Physics,
Tata Institute of Fundamental Research,
Homi Bhabha Road, Mumbai 400005, India
(Dated: June 12, 2010)

PACS numbers: 05.65.+b, 05.45.Df

PHYSICS ON HUMAN SCALES

The range of phenomena addressed by physicists in trying to understand nature is very wide. In length scales, it varies from the Planck scale (about 10^{-35} meters), to the size of the universe ($\sim 10^{27}$ meters). In time scales from 10^{-43} seconds to billions of years. The enormity of this range is somewhat mind-boggling. While the problems of understanding ultra-short and very-large length scales are certainly interesting and intriguing problems, it is good to remember that there are many interesting and fascinating phenomena at human length scales also, where our understanding today is still rather rudimentary. Here I will like to discuss some examples of such problems, taken the general area of statistical physics, as that is the subject I am most familiar with.

One of the main thrusts of physics research in the last century has been to understand what are the fundamental constituents of matter, and interactions between them. However, knowing these does not directly help explain the world around us. Because, most of the time, one deals with phenomena involving a large number of particles.

The mathematical problem of determining the dynamics of n interacting bodies is usually quite intractable, for $n > 2$. For example, in Newton's classical theory of gravitation, one can analytically calculate the motion of a planet around the sun. However, the three-body problem, of three massive bodies moving under each other's gravitational field (say the sun, the earth, and Jupiter) cannot be solved exactly. There are well-known examples of systems with few degrees of freedom that undergo deterministic evolution, e.g. the logistic map, or the Lorentz model, that chaos, i.e. a sensitive dependence on initial conditions, making long time predition of evolution impossible in practice. Thus, for such systems, knowledge of basic laws of interaction does not lead to effective predictability of future.

System with many more than 2 or 3 degrees of freedom are usually chaotic, but in a way , *a more chaotic evolution actually makes the system more predictable, if we agree to give only a probabilistic description of the system.* Then, by what is called the law of large numbers, the average properties of the system can be determined quite accurately. This is the domain of statistical physics, where we try to determine the properties of a system consisting of a large number of interacting parts, with the interaction between parts given.

The most interesting feature of systems with many degrees of freedom is captured in the aphorism: *the whole is bigger than the sum of its parts.* For example, from atoms and molecules, at larger scales of aggregation we go to chemistry and chemical reactions, to structure and function of large molecules like proteins, and then to cells, animals, and society. At each new level of organization, we find it useful to introduce new concepts, like life, love and culture. While the whole is more than its parts, it has to be understood in terms of the parts. The first problem with many degrees of freedom that was understood was the equilibrium properties of a box of a gas, made up of many molecules. The lessons learned there gave rise to the subject called statistical physics. It is reasonable to expect that the techniques developed there would also be useful in the more complicated cases mentioned above.

THE IMPORTANCE OF SIMPLE MODELS

The systems of many interacting subunits that I want to discuss in this article are considered 'complex'. Let us not worry about a technical definition of what is complex here. Suffice it to say that a system is complex if it can not be described adequately in terms of a few variables. Thus a box of gas in thermal equilibrium is simple, as its state is well specified by giving the values of a few thermodyanical variables like the pressure, temperature, and volume. But the same box, if stirred, can show complex turbulent motions. However, the *models* of complex systems need not be as complex. In fact, as noted by the famous mathematician V. I. Arnold :"Complex models are rarely useful (except for those writing their dissertations)."

I will try to illustrate this idea of 'simple models' with a few examples. Each example captures some important feature of the system considered. It deliberately ignores other features. The advantage of simplification is that the

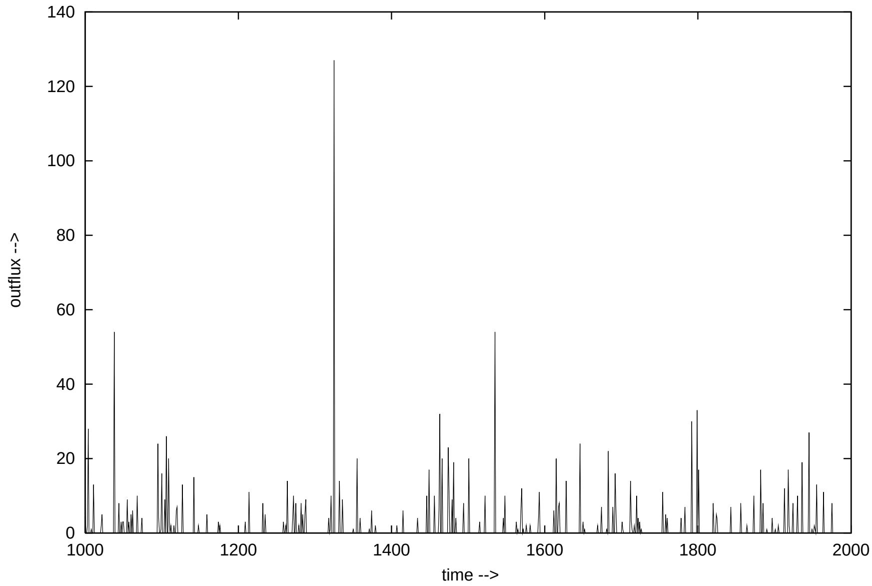

FIG. 1: A schematic representation of the outflux of sand as a function of time. The data shown was actually generated on computer by simulating the sandpile model on a square lattice of size 100×100. From [3].

problem becomes tractable. In fact, a necessary part of the process of understanding is the ability to recognize and ignore irrelevant details.

SELF-ORGANIZED CRITICALITY AND SANDPILES

Many of you may have read about 'fractals' in popular science books. Many natural systems like mountain ranges, coastlines, are said to show an effectively non-integer dimension: they are said to have a fractal spatial structure. These are produced by the natural dyanamics of the system, and are robust against changes of environmental conditions like the temperature. The occurence of fractal structure implies existence of long ranged correlations in the system. This is a rather surprizing property of non-equilibrium systems, as in equilibrium systems, one typically has to fine-tune temperature and pressure etc. to produce long-ranged correlations. Bak et al, argued that in such a system, its natural dynamics drives it towards, and then maintains it, at the critical point [1, 2]. Such systems were named self-organized critical.

Bak et. al. proposed a simple example of such a system : a sandpile. It is observed that for dry sand, one can define an angle θ_c, called the angle of repose. If we make a sandpile with local slope is smaller than θ_c everywhere, it is stable, and addition of a small amount of sand will cause only a weak response. A pile with slope larger than θ_c can be made, by careful tapping, but is unstable to perturbation: adding a small amount of sand will cause a huge avalanche. In a pile where the average slope is θ_c, the response to addition of sand is less predictable. It can cause a small or big avalanche, with a wide distribution of possible sizes. Such a state is critical. Bak and his coworkers observed that if one builds a sandpile on a finite table by pouring it very slowly, the system is invariably driven towards its critical state, and thus organizes itself into a critical state: it shows self-organized criticality.

A similar behavior is seen in earthquakes, where the build-up of stress due to tectonic motion of the continental plates is a slow steady process, but the release of stress occurs sporadically in bursts of various sizes.

Bak and coworkers also proposed a simple cellular automaton model of sandpile dynamics. The model is defined on a lattice, which we take for simplicity to be the two dimensional square lattice. There is a positive integer variable at each site of the lattice, called the height of the sandpile at that site. The system evolves in discrete time.

The rules of evolution are quite simple: At each time step a site is picked at random, and its height is increased by unity. If its height is then larger than a critical height 3, the site is said to be unstable. It relaxes by toppling whereby four sand grains leave the site, and each of the four neighboring sites gets one grain. If there is any unstable site remaining, it too is toppled. In case of toppling at a site at the boundary of the lattice, grains falling 'outside' the lattice are removed from the system. This process continues until all sites are stable. Then another site is picked

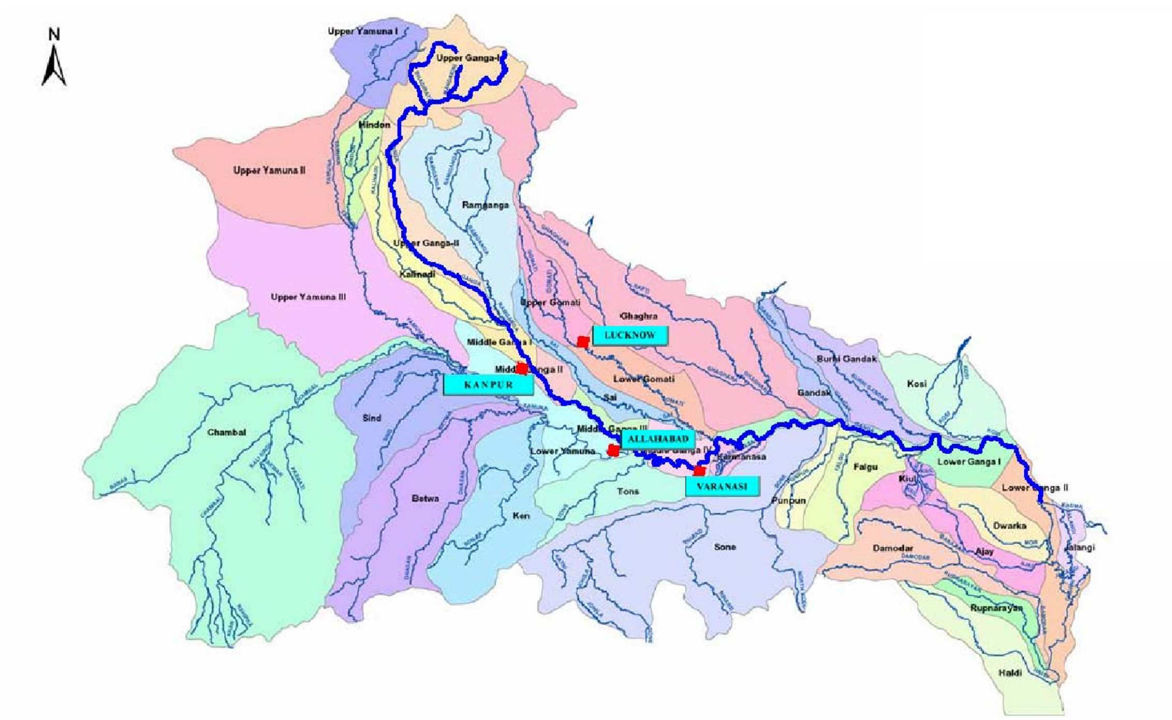

FIG. 2: A map of catchment area of the Ganga, showing some of the major tributaries. The basin has been divided into sub-basins, shown in different colors.

randomly, its height increased, and so on.

The steady state of this model is characterized by the following property: Sand is added to the system at a constant small rate, but leaves the system in a very irregular manner, with long periods of apparent inactivity interspersed by events which may vary in size and which occur at unpredictable intervals (see fig. 1).

The sandpile model is not a very good model of real sand. For realistic modelling of sand, one has to study other toppling rules, grain dynamics etc. However, it turns out that this model has a very interesting abelian group structure, which makes it analytically tractable [3]. This makes it a very good pedagogical model to illustrate the basic ideas of Sself-organized criticality. There has been a lot of work on this and related models, and many variations of the basic sandpile model have been proposed to model earthquakes, solar flares, firing of neurons in the brain, biological extinctions etc.. [2].

RIVER NETWORKS

In Fig. 2, I have shown a map of a part of the catchment area of the Ganga, and its tributaries. There is a rather complex network of smaller and bigger rivers that make up this network. If we take some other river, say the Amazon, we have a different structure. Can we *predict* the structure of the network from our knowledge of the physics and geology of the region involved?

Firstly, we realize that the question is not really well-posed. The river network is not a time-independent structure, and evolves in time. Rivers change course, or even disappear underground, as is believed to happened with the river Saraswati, more than 30 - 50 thousand years ago. Also, the course the river depends on the slope of the landscape. In turn, the river-flow modifies the landscape by eroding or depositing silt along the way. So, there is a complicated evolution in time: the water-flow depends on landscape, which then changes the landscape, which then changes the flow, etc..

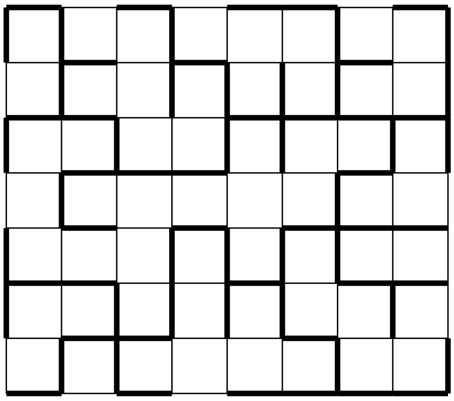

FIG. 3: A spanning tree on a 9×8 square lattice. The occupied edges are shown as thick lines.

Let us simplify the question a bit. Instead of asking for predicting the structure of a particular river at a particular time, we ask for some general properties of the network that are independent of details of geography, and are the same for different river basins.

What are these properties that can be used to describe a river network? We start by noting that rivers are formed by water, falling on land as rain or snow, which then drains out to the sea. We may imagine a land area that receives on the average a uniform rainfall per year. We also make the simplification that water-loss by evaporation and seepage into ground may be ignored. Then each drop of water that falls on land, finds its way moving downward along the land, forming a stream. If two streams meet, they join, and then move as a single stream. One can also make the simplification that a stream moving down does not break into two streams. This is usually a good approximation, away from the delta regions.

Map-makers and hydrologists have made some conventions about drawing such maps. One starts by choosing a lower cuttoff: only rivers with annual flow greater than this value are shown in the map. Then the river network becomes a tree-like structure. The nodes are the points of confluence. One assigns a rank number, called the Strahler number to each segment of the river. A river that starts from nothing is given Strahler rank 1. If two streams of rank 1 merge, we get a sream of rank 2. A sream of rank 2 joining with a stream of rank 1 remains a stream of rank 2. More generally, if a stream of rank r_1 joins a stream of rank r_2, with $r_1 > r_2$, the resulting stream has rank r_1, and retains the same name. However, if $r_1 = r_2$, the resulting stream has rank $r_1 + 1$, and in general can be given a different name. So, if a low-rank stream joins Ganga, the river remains Ganga. But if the streams Alakhnanda and Bhagirathi join, it is not clear whether the resultant stream should be called A or B. It seems fair to give the resultant a new name, in this case, Ganga.

A catchment area of a stream at any point along its length is the entire area whose rainfall reaches that point. One can take a largish catchment area, and draw a map of all the rivers in that region. Let N_r be the number of distinct streams of rank r. The numbers N_r will depend on the point selected. There are many more streams of smaller rank than larger. There is an interesting observation made by hydrologists: the ratios N_r/N_{r+1} are roughly independent of the point selected, independent of r, and roughly constant for very different river basins. We will denote this ratio by R. Its value from the field data of different river basins is found to be ≈ 3.5. There are several other such systematics seen. Let me just mention one more: The length of the longest stream in a catchment basin varies as A^h, with $h \approx 0.6$. This is known as Hack's law. See [4] for a more detailed discussion of such laws.

We would like to know what is the origin of these phenomenologically observed scaling laws? Can one make a simple model to "explain" this? Let us try to construct such a model. In our model, we discretize space into a set of grid points. At each grid point, one unit of rain falls per unit time. From each such point, water flows out to a nearest neighbor grid point. A possible river network forms a tree on this graph, which contains all the grid-points. Such a tree is called a *spanning tree* on this lattice. In Fig. 3, we have shown an example of a spanning tree on a 9×8 rectangular portion of a square lattice.

Now, it turns out that one can count how many distinct spanning trees one can construct on an $L \times L$ lattice, and that for overwhelmingly large majority of cases, R has the same value. Furthermore, this value is independent of the discretization of the grid. If one chose a triangular grid, one would get the same Ratio R for ranks $r >> 1$. Unfortunately, an analytical calulation of R has not been possible so far. In regards to other laws, one can prove that for random spanning trees in two dimensions, the exponent $h = 5/8$. The proof involves results from two-dimensional conformal field theory [3].

Thus, we see that a simple model of a river network is just a random spanning tree. The details of geology and history can be ignored, to the first approximation. The model can be improved, by taking some overall slope of the basin into account, or the fact that rivers have a finite width (quite significant for larger ranks), and cannot be represented by a line between two nodes. These improvements are certainly needed before one can fully match calculated numbers to real data. But the simple model does work quite well. There are several other models that have been proposed, for river networks [5].

STOCK MARKETS

A much more complex system than a sandpile is the cooperative behavior shown by groups of humans, say in the stock market. The stock market is more complex than the sandpile, because the behavior of a human is much more complicated than of a sand-grain. Can we make a statistical mechanical model of the stock market? There has been a lot of work recently, in applying the methods and techniques of statistical physics to financial market analysis. The subject has been called Econophysics [6, 7].

The first thing one should clarify here is the aim of modelling of the market, or the theory of the stock-markets. Many people think, wrongly, that the theory of the stock-market should help you make money at the stock-market. This is just not possible, if the other players are also equally smart, and also know the theory. Stock market is a bit like the game of poker, where agents bet, based on incomplete information, and win or lose. One can have a well-defined probabilistic description of the poker game, where one can calculate the probabilities of diffferent hands. However, if you can calculate probabilities well, it does not mean you will win in the game, even on the average. However, if you play poker with a child, who has less feel for the probabilities, you may expect to win. Knowledge of probabilities does not ensure winning, but ignorance of probabilities can be a definite disadvantage.

So, if the theory of stock-market can not help you make money against equally smart people, what is it good for? Well, it would be able make some general quantitative statements about the nature of fluctuations in prices, probabilities of large deviations etc., which are important to estimate in finance. For example, in Fig. 4, we have shown a typical variation of the price of a stock with time on a single trading day. Can one understand the general behavior of such price fluctuations?

The earliest theory of such fluctuations was proposed by Bachelier in 1900. He argued that change in price in a given time should be proportional to the price itself, hence the logarithm of the price would execute a brownian motion. This model would predict that if we consider a largish number of similar stocks, and calculate the fractional change in price, called the returns, after a fixed interval of time, the probability distributions of returns would be a gaussian distribution. While this seems to be roughly correct, in the actual distribution, the probability of deviations larger than typical root mean square deviation is much larger than that predicted by the Gaussian distribution. The theory of market would be expected to identify the origin of, and understand the nature of the correlations responsible for this behavior.

As another example of the kind of questions that are addressed in econophysics modelling, let us consider the distribution of wealth in populations. Suppose one has a group of N agents, and they all start with equal amount of money. In time, each of the agents chooses another agent to trade with, and they have a business transaction. As a result of the transaction, some money changes hands. Let money with the agents A and B be M_A and M_B before the transaction. Consider a simple nodel where both A and B pool together the total money available to both. In the transaction, a random number x uniformly distributed between 0 and 1 is drawn, and after the transaction, A gets to keep fraction x of the total, i.e. $x(M_A + M_B)$, and B keeps the rest. This model is an example of a wealth-exchange model. In general, one can study different rules for how transfer of money occurs in transactions.

For the simple model defined above, it can be shown that even if all players start with equal assets in the beginning, with time the assets of different agents will fluctuate, but at large times a steady state is reached, in which while the fortunes of individual agents will continue to fluctuate, the system reaches a statistically stationary state. The distribution of money with agents lends to a limiting distribution, in which the fractional number of agents with money $\geq M$ decreases as $\exp(-kM)$. This is like the way energy is divided amongst molecules of a gas by binary collisions.

On the other hand, one can argue that the model is very unrealistic: a rich agent has no reason to make a transaction with a poorer one. We can consider a different model where both players in a transaction bet equal amount of money, say equal to the money with the poorer agent, and then the betted amount is divided randomly between the two. The analysis of this betting game shows that at long times, almost all the money ends up with one of the players, and all others have very little money. Similar analysis can be done for more complicated rules: say there is a government

FIG. 4: A typical plot of price variation of a stock in a single trading day. The x-axis marks the clock-time.

that promotes equality, and imposes a transaction tax, and richer people have to pay more tax than poorer ones. The total tax collected each year is equally distributed amongst the agents. One can study the effect of such a government policy on the steady state wealth distribution. The long-time steady state can be determined analytically for some rules [8]. For the rest, one can study the system by numerical simulations.

Clearly, the models of financial markets described above are rather crude, and not really satisfactory. However, they do emphasize the importance of treating markets as a system of interacting agents, and the importance of properly taking account of the fluctuations in the system. These have not been adequately treated in the conventional economics theory. For example, the classical theory of markets has the Law of Supply and Demand, which says that once these match, a steady state is reached, and no price fluctuations would occur till the changes occur in demand etc.. At the very least, the physicists bring in a new approach, and new techniques of analysis, which complement the analysis of traditional economists.

To conclude, I think that there is a very rich variety of phenomena out there. Our understanding of these is still in infancy. I hope that some of the readers will be encouraged to seek.

This work is supported by the Department of Science and Technology, Government of India, by a J. B. Bose fellowship.

[1] Bak P., Tang C. and Wiesenfeld K., Phys. Rev. Lett., **59**(1987) 381.
[2] Bak P., *How Nature Works*, (Oxford Univ. Press, Oxford, 1997); *Self-organized Criticality*, H. J. Jensen, Cambridge Lecture Notes in Physics 10, Eds. P. Goddard, J. Yeomans, (Camb. Univ. Press, 1998).
[3] D. Dhar, Physica **A 369**, (2006), 29.
[4] Dodds P. S. and Rothman D. H., Phys. Rev. **E 59** (1999) 4865. See also Colaiori F., Flammini A., Maritan A., and Banavar J. R., Phys. Rev. **E 55** (1997) 1298.
[5] Scheidegger A. E., Bull. Assoc. Sci. Hydrol. **12** (1967) 15; Somfai E. and Sander L. M., Phys. Rev. **E 56** (1997) R5; Huang J. and Turcotte D. L., J. Geophys. Res., **94** (1989) 7491.
[6] R. N. Mantegna and H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance*, (Camb. Univ. press, Cambridge, 2000).
[7] A. Chatterjee and B. K. Chaktrabarti (Eds.), *Econophysics of Markets and Business Networks*, (Springer, Berlin, 2007).
[8] Basu U. and Mohanty P. K., Euro. Phys. J. B, **65** (2008) 585.