Computer Simulations of Statistical Models and Dynamic Complex Systems

J.S. Sá Martins and P.M.C. de Oliveira

Instituto de Física, Universidade Federal Fluminense, Av. Litorânea s/n, Boa Viagem, Niterói, Brazil, 24210-340

These notes concern the material covered by the authors during 4 classes on the Escola Brasileira de Mecânica Estatística, University of São Paulo at São Carlos, February 2004. They are divided in almost independent sections, each one with a small introduction to the subject and emphasis on the computational strategy adopted.

1 Introduction

This article is an expanded version of the notes used by the authors in the course taught in the 2004 edition of the Brazilian School of Statistical Mechanics (EBME), held in February of 2004 at the São Carlos campus of the State University of São Paulo (USP). The various sections are mostly independent of each other, and can be read in any order. Different styles and notations, independently adopted by each author while writing their first versions, were kept.

The subject of all sections can be interpreted as the same, in a broad sense, namely the consequences of lacking multiple characteristic scales (time, size, etc) in both equilibrium statistical models (two next sections), and complex dynamic systems (remainder sections).

2 Percolation

Mathematicians prefer to define percolation starting from an infinite lattice. For simplicity, let's consider a square lattice. Each site is randomly tossed to be present or absent, according to a fixed concentration p. By considering links between nearest neighbour present sites, one studies the structure of clusters (or islands) which appears. Fig. 1 shows a piece of such a lattice, as an example.

This system presents a phase transition, when one varies the concentration p. For small enough values of p, all islands are finite, including the largest one (highlighted by black circles in Fig. 1). Starting from any present site, it is impossible to go too far away from this point by walking only through nearest neighbouring present sites: eventually one returns back to the starting point. On the other hand, by increasing p, suddenly the largest cluster becomes infinite, at a very precise critical value p_c, and then one can cover infinite distances starting the walk from any present site belonging to this cluster. The parameter of order for this transition can be defined as follows. One takes a site at random, and measures the probability $P(p)$ that this site belongs to the largest cluster. For $p < p_c$, this largest cluster being finite on an infinite lattice, this probability vanishes. For $p > p_c$, however, this

probability monotonically increases. Fig. 2 shows the qualitative behaviour of this parameter. Near p_c (above), this order parameter is given by

$$P(p) \sim (p - p_c)^\beta \quad \text{for} \ \ 0 < p - p_c <<< 1 \ ,$$

where β is a universal critical exponent, which depends only on the lattice dimension, not on the local lattice geometry (its value $\beta = 5/36$, for instance, is the same for square, triangular or any other two-dimensional lattice). The threshold $p_c = 0.59274621(13)$ [1] is valid for the square lattice. However, this is not a universal quantity: the triangular lattice, for instance, has $p_c = 1/2$. See [2] for this and other interesting issues.
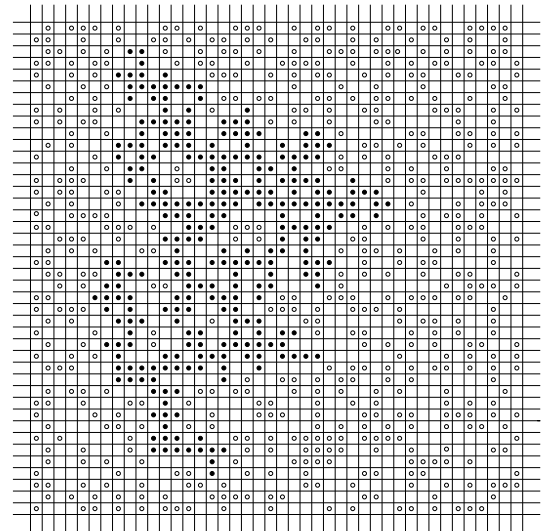

Figure 1. Finite piece of an infinite square lattice. Small black circles define the largest cluster. All other smaller clusters are denoted by open circles.
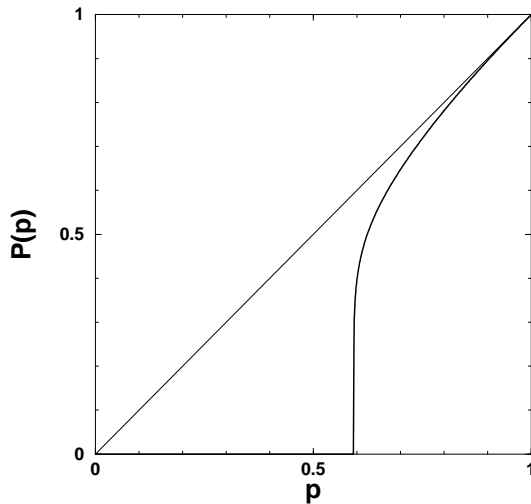
Figure 2. Probability for a random site to belong to the largest cluster.

Many further different quantities can be considered. One of then is the correlation function $G(x)$, defined as follows. For a given configuration like that on Fig. 1, one takes two sites distant x from each other, and counts $g = 1$ if they belong to the same cluster **excluded the largest one**, otherwise one counts $g = 0$. Then, one takes many other pairs of sites distant the same value x, and perform the average of g. Finally, one considers many other configurations tossed under the same concentration p, and perform the configuration average. For fixed p and large values of x, this function exponentially decays as a function of x,

$$G(x) \sim e^{-x/\xi(p)} \quad \text{for large } x \text{ and } p \neq p_c ,$$

where the so-called correlation length $\xi(p)$ measures the range of correlations, or, alternatively, the typical diameter of an island. Even the lattice being infinite, one does not need to consider distances larger than $\xi(p)$: a finite piece of the lattice, larger than this size, presents the same behaviour as the whole infinite lattice.
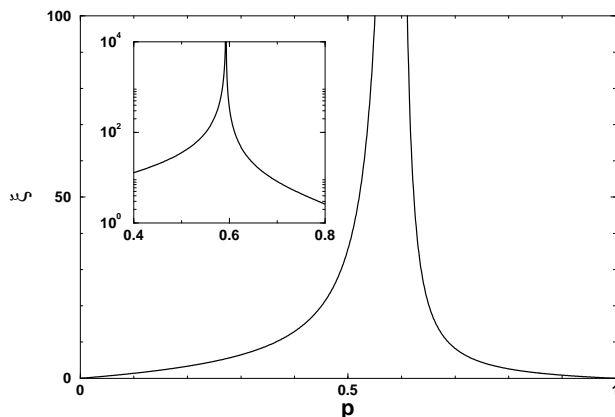

Figure 3. Correlation length (typical diameter of an island).

The exponential form above is only the asymptotic leading factor of $G(x)$. Exactly at $p = p_c$, however, long-range correlations appear, i.e. $\xi(p)$ diverges, Fig. 3, and another leading factor emerges, namely the power-law

$$G(x) \sim x^{-\eta} \quad \text{for large } x \text{ and } p = p_c .$$

For lattice dimensions other than $d = 2$, the critical exponent is $d - 2 + \eta$, instead. Again, the critical exponent η is universal (in two dimensions, $\eta = 5/24$). Contrary to the exponential decay which defines the typical length ξ, this power-law lacks any characteristic length scale. At p_c, any finite piece of the infinite lattice is not enough to describe its complete critical behaviour: correlations overflow the boundaries of this piece, whatever is its finite size. For this and any other critical system, the classical perturbation approach of determining first the behaviour of a small piece of the system, including later the influences of all the rest (the "perturbation"), does not work at all: the "rest" is never negligible.

Near but **not** exactly **at** the threshold p_c, the correlation length itself incorporates this criticality as

$$\xi(p) \sim |p - p_c|^{-\nu} \quad \text{for } |p - p_c| <<< 1 ,$$

where ν is another universal critical exponent (in two dimensions, $\nu = 4/3$). Note the importance of excluding the largest cluster from the above definition of correlation function: the correlation length remains finite below as well as above the threshold p_c. By further increasing p above p_c, some finite islands glue on the already existent infinite cluster, and the average diameter of the remainder islands decreases. Fig. 3 shows the qualitative aspect of $\xi(p)$. The exponent ν is the same on both sides of p_c. The multiplicative factors of $|p - p_c|^{-\nu}$ omitted from the last equation, however, are two different constants.

Another interesting quantity is the mean cluster size $\chi(p)$, the average number of sites belonging to each island, again excluding the largest cluster. Assigning a unit "mass" to each present site, this quantity can be interpreted as the typical mass of an island. It reads

$$\chi(p) \sim |p - p_c|^{-\gamma} \quad \text{for } |p - p_c| <<< 1 ,$$

where γ is the corresponding critical exponent, also universal (in two dimensions, $\gamma = 43/18$). The qualitative plot for $\chi(p)$ is shown in Fig. 4. Again, the universal exponent is the same below and above p_c, but the amplitudes multiplying $|p - p_c|^{-\gamma}$ are different.
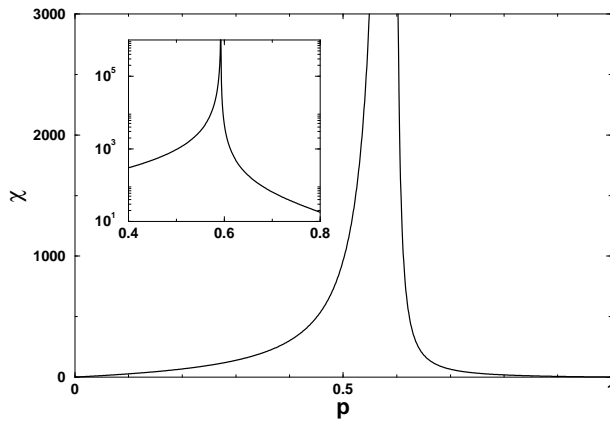
Figure 4. Mean cluster size (typical mass of an island).

All these quantities are related to the average number $n_s(p)$ of s-clusters per site, where s denotes the cluster size, or mass. The order parameter, the correlation length, the mean cluster size and others can be obtained from $n_s(p)$, which is then called the **potential** or **generating** function. (The term "per site" means to consider an enormous lattice piece, and divide the counting by the number of sites inside it, i.e. $n_s(p)$ is indeed the density of s-clusters.) In particular, at p_c this size distribution follows a power-law,

$$n_s(p_c) \sim s^{-\tau} \quad ,$$

where τ is another universal exponent (in two dimensions, $\tau = 187/91$). Again, the system lacks any characteristic size scale: at p_c, one cannot neglect clusters larger than any predefined cut off.

Near but not exactly at p_c and for large sizes s, $n_s(p)$ is a generalised homogeneous function [3] of $p - p_c$ and $1/s$, i.e. it obeys the property

$$n[\Lambda^\sigma (p - p_c), \Lambda s^{-1}] = \Lambda^\tau \, n(p - p_c, s^{-1}) \quad ,$$

where Λ is an arbitrary number, and σ is a further critical, also universal exponent (in two dimensions, $\sigma = 36/91$). The critical point corresponds to both variables $p - p_c$ and $1/s$ vanishing. By choosing $\Lambda = s$, one obtains

$$n_s(p) = s^{-\tau} \, f[(p - p_c)s^\sigma] \quad ,$$

which allows an interesting interpretation. By measuring $n_s(p)$ in units of $n_s(p_c)$, i.e. by taking $n_s(p)/n_s(p_c)$, one does not need to consider it as a function of two variables p and s: it depends only on the single combination $z = (p - p_c)s^\sigma$.

The order parameter $P(p)$, Fig. 2, for instance, is related to $n_s(p)$ through

$$p = P(p) + \sum_s s\, n_s(p) \quad .$$

The reasoning behind this formula is simple. For a fixed size s, one can determine the probability $s\, n_s(p)$ of picking a random site belonging to some s-island. (Why? First, one

counts the total number of s-islands and multiply with s, getting the total number of sites belonging to all s-islands. Then, one divides the result by the total number of lattice sites, getting the quoted probability. Remember that $n_s(p)$ is the number of s-islands *per site*.) Thus, the sum over s which appears in the above equation corresponds to the probability of picking a random site belonging to any **finite** island (of course, only finite values of s are scanned by this sum.) Finally, the concentration p of present sites is the probability for a random site to belong either to the infinite cluster (if any) or to any other island, obtained by summing up the probabilities $P(p)$ and $\sum_s s\, n_s(p)$ of each case, respectively. In Fig. 2, the region above the curve for $P(p)$ and below the straight $45°$-diagonal corresponds just to all sites belonging to **finite** islands.

The mean cluster size $\chi(p)$ is also related to $n_s(p)$ through

$$\chi(p) = \sum_s s^2 n_s(p) \quad ,$$

which rid off further explanations.

Both $P(p)$ and $\chi(p)$ are one-site averages, obtained by scanning the lattice site by site. Thus, for each site, the probability $s\, n_s(p)$ to belong to some s-island is considered in the two last equations. The correlation length $\xi(p)$ is different, it is a two-site average. In this case, one needs to scan all pairs of lattice sites which belong to the same island. The class of s-islands contribute, then, with a weight proportional to $s^2 n_s(p)$. (Why? The probability $s\, n_s(p)$ to pick the first site inside this class is multiplied with the number $s - 1$ of remainder sites of the same island: any of them could be the second site, in order to form a pair. For large islands, s and $s - 1$ can be confounded.) Instead of determining $\xi(p)$ directly from the correlation function $G(x) \sim e^{-x/\xi(p)}$, one can compute a related quantity denoted here with the same symbol $\xi(p)$ by

$$\xi^2(p) = \frac{\sum_s R_s^{\,2}\, s^2\, n_s(p)}{\sum_s s^2\, n_s(p)} \quad ,$$

where R_s measures the average radius: for each s-island, one first computes its gyration radius defined by $R^2 = \sum_i r_i^{\,2}/s$, where r_i is the distance between site i and the center of mass, and then averages over all s-islands.

At a first glance, one is tempted to relate the average mass to the average radius through the Euclidean relation $mass \sim radius^d$. However, percolation clusters are not compact objects. Instead, they are **fractals** with dimension D a little bit smaller then the Euclidean dimension d. Thus, the correct relation is $mass \sim radius^D$. Near p_c, a typical fractal island with radius ξ has mass ξ^D smaller than the volume ξ^d, due to the holes which characterise the fractal. The volume fraction ξ^D/ξ^d actually occupied by this island equals the probability $P(p)$. From this equality, namely $P = \xi^D/\xi^d$, one finds the scaling relation

$$D = d - \beta/\nu \quad .$$

The fractal dimension D is also a universal critical exponent (in two dimensions, $D = 91/48$). By transforming

the s-sums into integrals, a procedure valid near the critical point, and by profiting from the generalised homogeneous form of $n_s(p)$ in order to change the variable from s to $z = (p - p_c)s^\sigma$, one can find various other scaling relations between critical exponents, for instance

$$\alpha = 2 - \frac{\tau - 1}{\sigma} \quad,$$

$$\beta = \frac{\tau - 2}{\sigma} \quad,$$

$$\gamma = \frac{3 - \tau}{\sigma} \quad,$$

$$\delta = 1 + \frac{\gamma}{\beta} \quad,$$

$$D = \frac{d}{\tau - 1} \quad,$$

$$D\sigma = \frac{1}{\nu} \quad,$$

$$\nu d = \frac{\tau - 1}{\sigma} \quad,$$

$$\gamma = (2 - \eta)\nu \quad,$$

etc. Exponents α and δ concern respectively the temperature dependence of the specific heath, and the magnetic field dependence of the magnetisation. They were not directly treated in the present text. The corresponding equations were included above for completeness. Indeed, these various critical exponents are not independent from each other: knowing two of them, for instance τ and σ, one automatically knows all others. Equivalently, the many power-law dependences which characterise the critical behaviour are related to each other.

Contrary to mathematicians, physicists prefer finite lattices, for instance a $L \times L$ square, a lattice piece like that shown in Fig. 1. It is treated by choosing some artificial rule to define the neighbourhood along its boundary. A particularly convenient rule is to take the bottom-row spins to replace the missing up-neighbours of the top-row. Symmetrically, one chooses the top-row spins as down-neighbours of the bottom-row. This rule is usually called periodic boundary condition, the lattice becoming a cylinder. One advantage is the preserved translation symmetry, all rows are rigorously equivalent. Considering also periodic boundary condition along the vertical direction, all columns equivalent, the cylinder turns into a torus.

The infinite lattice limit is considered by including the lattice size as a further parameter $1/L$ into the generalised homogeneous relation, the so-called finite-size-scaling hypothesis

$$n[\lambda^{1/\nu}(p-p_c), \lambda^D s^{-1}, \lambda L^{-1}] = \lambda^{d+D} \; n(p-p_c, s^{-1}, L^{-1}) \quad,$$

where λ is an arbitrary number related to the previous one by $\Lambda = \lambda^D$. From this, one can express the density $n_{s,L}(p)$ of s-clusters as

$$n_{s,L}(p) = s^{-\tau} f[(p - p_c)s^\sigma, s/L^D] \quad,$$

a function of only two variables $z = (p - p_c)s^\sigma$ and $w = s/L^D$, instead of p, s and L. In particular, exactly at p_c, one can perform integrals in w instead of s, getting a series of useful finite-size-scaling relations

$$\chi_L(p_c) \sim L^{\phi_\chi} = L^{\gamma/\nu} \quad,$$

$$P_L(p_c) \sim L^{\phi_P} = L^{-\beta/\nu} \quad,$$

$$\xi_L(p_c) \sim L^{\phi_\xi} = L \quad,$$

etc. The exponent $\phi_\chi = \gamma/\nu$ is called the **anomalous dimension** of the corresponding quantity $\chi(p)$, $\phi_P = -\beta/\nu$ is the same for the order parameter $P(p)$, etc. Not surprisingly, the correlation length shares the same dimension $\phi_\xi = 1$ of the length finite size L, last relation. It has an interesting interpretation, normally used as an argument in favour of the finite-size-scaling hypothesis: instead of diverging at p_c, the plot of $\xi(p)$ for a finite lattice presents a peak near p_c, the height of which is proportional to the lattice length L.

Figure 5 gives the qualitative behaviour of some generic quantity $Q(p)$ which diverges at p_c, for an infinite lattice, according to its characteristic critical exponent q, i.e. $Q(p) \sim |p - p_c|^{-q}$. The plot, however, corresponds to a finite lattice of length L. The divergence is replaced by a cut off peak, with $Q_L(p) \approx L^{q/\nu}$ for values of p inside an interval of width $L^{-1/\nu}$ around p_c. If the threshold p_c is known, one can profit from this finite-size behaviour in order to determine the anomalous dimension $\phi_Q = q/\nu$, by measuring $Q_L(p_c)$ for different lattice sizes. An estimate for p_c itself can be also obtained, by tuning the precise concentration for which the log-log plot of Q against L is a straight line.
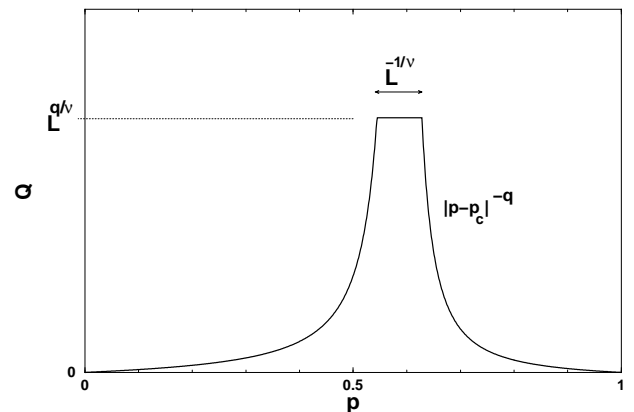

Figure 5. Finite size cut off of a divergence.

The finite-size-scaling form

$$Q[\lambda^{1/\nu}(p - p_c), \lambda L^{-1}] = \lambda^{-\phi_Q} \, Q(p - p_c, L^{-1})$$

exhibits the anomalous dimension ϕ_Q for this quantity Q. By choosing $\lambda = L$, one reveals an interesting property: the plots of $Q_L(p)/Q_L(p_c)$ against $(p - p_c)L^{1/\nu}$ will collapse on the same curve for different lattice sizes L. This equation is obtained by performing the proper $s-$integration on the finite-size-scaling relation.

Particularly useful are quantities with vanishing anomalous dimension, $\phi_Q = 0$. Plotted against L, they are asymptotically constant for large values of L, when the correct concentration $p = p_c$ is chosen. The so-called Binder cumulant [4] and the correlation between opposite surfaces [5] are general examples of these quantities. For percolation, one has the spanning probability $S_L(p)$ around, say, the horizontal direction. For a fixed $L \times L$ finite lattice, it varies with the concentration p as a sigmoid curve schematically shown in Fig. 6.
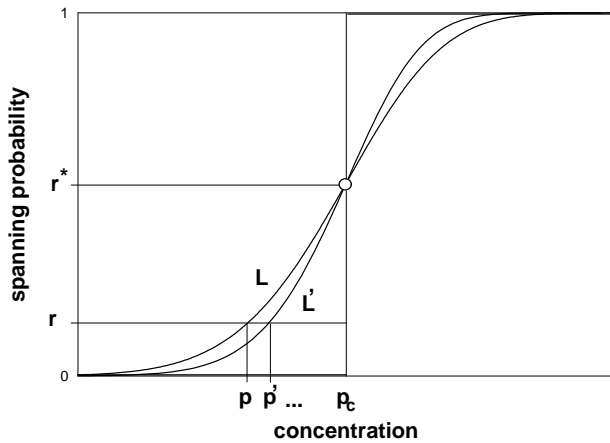

Figure 6. Spanning probability for two lattice sizes L and L' ($L' > L$). Due to its anomalous dimension $\phi_S = 0$, for larger and larger lattices this function approaches the step $S(p < p_c) = 0$; $S(p_c) = r^*$; $S(p > p_c) = 1$, also highlighted by two heavy straight lines plus the open dot. By fixing an arbitrary value r at the vertical axis, one finds the sequence $p, p' \ldots$ for increasing lattice sizes, which converges to p_c.

In the limit of infinite lattice size, this curve approaches a step function, suddenly jumping from 0 (valid for all $p < p_c$) to an isolated universal value r^* exactly at $p = p_c$. Then, it immediately jumps again from r^* to 1 (which remains for all $p > p_c$). It is a kind of order parameter, vanishing along all the disordered phase. However, instead of gradually increasing according to the smooth curve $(p - p_c)^\beta$ when the threshold is surpassed, it becomes immediately constant along all the ordered phase, i.e. it behaves as $(p - p_c)^0$, with a zero exponent instead of β. Thus the anomalous dimension of $S(p)$ is $\phi_S = 0$, whereas $\phi_P = -\beta/\nu$ for the order parameter $P(p)$.

A simple an efficient way to get the value of p_c is to extrapolate $S_L(p)$ for larger and larger lattices. By fixing some value r on the vertical axis of Fig. 6, one measures the corresponding values p_{L_1}, p_{L_2}, p_{L_3} $\ldots$ for increasing lattice sizes $L_1, L_2, L_3 \ldots$ In [6], the formula

$$p_L(r) = p_c + \frac{1}{L^{1/\nu}} \left[A_0(r) + \frac{A_1(r)}{L} + \frac{A_2(r)}{L^2} + \ldots \right] \, ,$$

is proposed for the L-dependence of this series. The fitting of this form to real Monte Carlo data is excellent, giving numerical estimates for the threshold p_c within an accuracy of one part in 10^6, for an acceptable computational effort [6]. One can fix any value of r at the vertical axis, Fig. 6. The rate of convergence is dictated by $1/L^{1/\nu}$, for increasing values of L.

A faster rate of convergence, and hence a better accuracy, can be obtained by choosing $r = r^*$, the critical universal value which in some cases is exactly known through conformal invariance arguments. Within this convenient choice, one has $A_0(r^*) = 0$ in the above equation, accelerating the convergence rate to $1/L^{1+1/\nu}$ [6]. Better yet is to consider the probability of wrapping along the torus, instead of spanning. In this case, the convergence rate is dictated by $1/L^{2+1/\nu}$, and the already quoted best known estimate $p_c = 0.59274621(13)$ was reported [1].

In [6], the spanning between two parallel lines distant $L/2$ from each other is verified: one counts $s = 1$ if there is some island linking these two lines, $s = 0$ otherwise. Then, $S_L(p)$ is the configuration average of s. This definition multiplies the statistics by a factor of L: for the same configuration, one considers rows 1 and $1 + L/2$, also 2 and $2 + L/2$, etc, and repeats the same for columns. Larger lattices can be monitored in this way, allowing for instance a much better definition of the threshold distribution tails [7].

The canonical way to simulate the percolation problem on computers corresponds to a fixed concentration p. According to this value, one tosses a random distribution, then measures the quantity Q of interest for this particular configuration, and finally tosses many other configurations and repeats the whole process. At the end, one has a single average value $Q(p)$ for the fixed concentration p. In order to get the whole function $Q(p)$, one needs to repeat the process for many other values of p. Even so, the result is not a continuous function of p.

Here, we will show a way to get the continuous dependence between Q and p [1]. One starts with an empty lattice. Then, the lattice is filled-up, one random site at each step. After each new site, one determines the value Q_n of the quantity of interest, where n is the number of present sites so far, and accumulates it on a n-histogram. When the lattice is completely filled with all its N sites present, one starts the same process again, from the empty lattice. After a sufficient large number M of repetitions, one divides all $N + 1$ entries of the histogram by M: now it stores the "microcanonical" averages Q_n, i.e. the averages of Q for fixed numbers $n = 0, 1, 2 \ldots N$ of present sites.

Once the histogram Q_n is complete, the simulation step is over, we miss now simply to treat the data already stored on it. The "canonical" average can be obtained from

$$Q(p) = \sum_{n=1}^{N} C_N^n \ p^n (1-p)^{N-n} \ Q_n \quad ,$$

```
    F

D       G       L

        K
```

```
F       D       G       L

            K
```
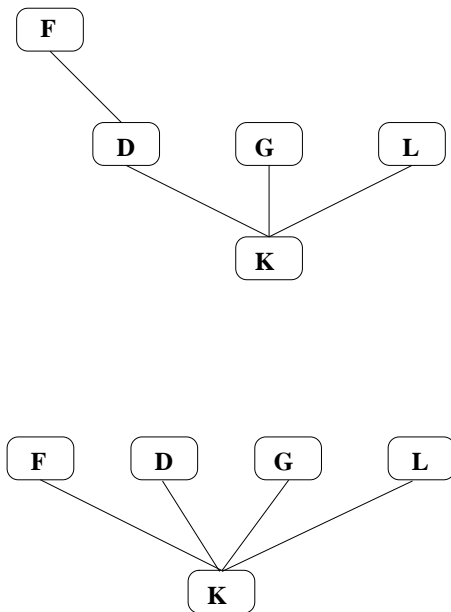
Figure 7. Tree structure of cluster labels. The root for this particular cluster is site **K**. In order to find the ultimate label for site **F**, one needs a two-step search, because it does not point directly to the root (up). Once this process is performed for the first time, however, the pointer becomes direct (down), saving computer time from now on.

where C_N^n is the combinatorial factor. Note that p was introduced as a continuous parameter, only after the simulation step is over. $Q(p)$ is simply a polynomial in p, the coefficients of which are stored in Q_n. The continuous character of $Q(p)$ may be of fundamental importance. To find the roots p, $p' \dots$ shown in Fig. 6, for instance, one needs to solve the equations $S_L(p) = r$, $S_{L'}(p') = r$, etc, which would be a hard work without knowing S_L as a continuous function of p.

During the simulation process, the determination of Q_n for the current configuration is not necessarily an easy task, if the quantity Q does not depend only on the local situation where the last site was occupied. However, the filling-up process makes the structure of clusters easy to study: only three situations should be analysed for each new site included. First, if this new site is isolated from all others, it forms a new one-site cluster and receives a new cluster-label, namely its own number n according to the chronological order it appeared. Second, this new site n could simply aggregate itself to an already existent cluster, being the next-neighbour of another previous site n' belonging to this cluster: in this case, the new site receives n' as its cluster label. The third possibility, for which non-locality should be considered, corresponds to join two or more old clusters into a single larger one. The structure of cluster labels is exemplified in Fig. 7. Each site points to another previous site belonging to the same cluster. Only the root of each cluster points to itself.

In order to join a cluster into another, one needs to find its root, and change its status: now it points to some site of the other cluster. To find the root of some cluster could be a multi-step search procedure, going from site to site along a tree branch, until finding the only one which points to itself, the root. In doing so, one can return back through the same path, updating all pointers to the real root. This saves computer time for future searches. The following recursive C-language routine performs this task.

```
int root(S) int S; {
   /* finds the cluster root of site S; the array Cluster[S]
      points to some site belonging to the same cluster as S */
 int              s;
 if((s=Cluster[S])==S) return(S);
 return(Cluster[S]=root(s));
}
```

3   Broad Histogram Method

The main purpose of equilibrium statistical physics is to obtain the canonical average

$$< Q >_T = \frac{\sum_s Q_s \mathrm{e}^{-E_s/T}}{\sum_s \mathrm{e}^{-E_s/T}}$$

of some quantity Q of interest, when the system under study is kept at fixed temperature T. The sums run over all possible states s. For each of them, Q_s and E_s represent the values of the quantity Q and the energy, respectively. For simplicity, the Boltzmann constant is taken to be unitary. For macroscopic systems, the number of states s is enormous. Consider a system with N binary units, for instance a set of N Ising spins which can point either *up* or *down*. One state s of the whole system is a fixed distribution of spins *up* and *down*. The total number of such states is 2^N, which is unimaginably huge for a macroscopic N. Thus, for most cases the sums appearing in the above equation cannot be analytically performed.

The so-called Monte Carlo method performs these sums for a restricted sub-set of states, instead of scanning all of

them. In doing so, one can control the degree of approximation by choosing the sampling set of states s_1, s_2 ... s_M to be representative enough, according to the desired accuracy.

For instance, in order to construct a completely random state, one chooses each Ising spin to point *up* or *down* according to 50% chance. The random sampling approach corresponds to perform the sums within M states tossed in this way. However, it would work properly only in the limit $T \to \infty$, where differences in energy are irrelevant. Normally, the number $g(E)$ of states sharing the same energy E is a fast increasing function of E. (For finite systems, besides a lower bound, there is also an energy upper bound. Thus, $g(E)$ normally presents a peak in between these bounds. However, only the energy region where $g(E)$ increases, from low to high energies, is important. One can forget the ultra-hight-energy region of the spectrum, where $g(E)$ decreases back, because equilibrium is impossible there: for that, one would need a negative temperature.) Within this random sampling approach, high-energy states near the peak of $g(E)$ are more likely to belong to the sampling sub-set. In the averaging process, on the contrary, the Boltzmann factor $\mathrm{e}^{-E_s/T}$ prescribes large weights just for low-energy states, which are seldom sampled under a completely random choice. In short, a sub-set of completely random states would be a good sample for high energies, which are not so important for the averages, but presents a poor sampling performance within the more important low energies.

In order to solve this drawback, the importance sampling Monte Carlo method incorporates the Boltzmann factor into the choice of M random states. In order to construct the sampling set, one chooses states according to a probability proportional to $\mathrm{e}^{-E_s/T}$. The pioneering recipe to perform this task is the half-century old Metropolis algorithm [8]. One starts to construct the sample set from a single random state s_1. Then, some variant of s_1 leads to a new state s, for instance a random spin is flipped in s_1, leading to s which is then a candidate to be s_2. If the energy difference $\Delta E = E_s - E_{s_1}$ is negative, then s is accepted, and $s_2 = s$ is included into the sampling set. Otherwise, one accepts s only within a probability $\mathrm{e}^{-\Delta E/T}$. In order to implement this conditional acceptance, one tosses a random number r between 0 and 1, and compares it with the quoted probability: if $r \leq \mathrm{e}^{-\Delta E/T}$, state s is accepted, i.e. $s_2 = s$ is included into the sampling set. However, if $r > \mathrm{e}^{-\Delta E/T}$, s is discarded and $s_2 = s_1$ is repeated in the sampling set. Following the same rule, the third element s_3 is constructed from s_2, s_4 from s_3, and so on, up to s_M. This sequence where each state is constructed from the previous one is called a Markov chain. At the end, one simply determines the unweighted average

$$< Q >_T \approx \sum_{i=1}^{M} Q_{s_i}$$

among these M states. Note that the Boltzmann weights $\mathrm{e}^{-E_{s_i}/T}$ are already taken into account during the construc-

tion of the sampling set. That is why they are absent from the above equation.

The important feature of this Metropolis recipe is to provide a Markov chain of states obeying the Gibbs canonical distribution, i.e. each state appears along the chain according to a probability proportional to $\mathrm{e}^{-E/T}$. Alternatively, each energy E appears along the Markov chain according to a probability proportional to $g(E)\,\mathrm{e}^{-E/T}$. In reality, the argument works in the reverse sense: given this canonical distribution, one can construct a recipe to toss random states obeying it. Normally, this construction is based on some arguments like detailed balance, and others. The above described Metropolis algorithm is one of these recipes.

If the protocol of allowed movements is ergodic, i.e. if any state s is reachable starting from any other, the degree of approximation depends only on M. Namely, the error is proportional to $1/\sqrt{M}$. The protocol defined by performing only single spin flips is obviously ergodic. Thus, in principle, one can surpass any predefined error tolerance, simply by improving the statistics with larger and larger values of M (also more and more computer time, of course). Other issues are indeed important, as the quality of the random number generator one adopts. This canonical importance sampling Monte Carlo method is by far the most popular in computer simulations of statistical models. However, it presents a strong limitation: one needs to fix the temperature T in order to construct the Markov chain of sampling states. At the end, for each complete computer run (which can represent days, months or even years), one has the average $< Q >_T$ only for that particular value of T. One needs to repeat the whole process again for each other value of T one wants to record.

The same canonical average can be re-written as

$$< Q >_T = \frac{\sum_E < Q(E) > g(E)\,\mathrm{e}^{-E/T}}{\sum_E g(E)\,\mathrm{e}^{-E/T}}$$

where

$$< Q(E) > = \frac{1}{g(E)} \sum_{s(E)} Q_s$$

is the microcanonical, fixed E average. The sum runs over all $g(E)$ states $s(E)$ sharing the same energy E. Both $g(E)$ and $< Q(E) >$ do not depend on the temperature T, they are quantities dependent only on the energy spectrum of the system, nothing to do with thermodynamics which describes its interactions with the environment. The temperature T appears only in the Boltzmann factors $\mathrm{e}^{-E/T}$, which presents a known mathematical form independent of the particular system under study. For a fixed temperature, each energy contributes to the average according to the probability

$$P_T(E) = \frac{g(E)\mathrm{e}^{-E/T}}{\sum_{E'} g(E')\,\mathrm{e}^{-E'/T}}$$

which can be related to the probability $P_{T'}(E)$ corresponding to another temperature T', by

$$P_{T'}(E) = \frac{\mathrm{e}^{E(\frac{1}{T} - \frac{1}{T'})} P_T(E)}{\sum_{E'} \mathrm{e}^{E'(\frac{1}{T} - \frac{1}{T'})} P_T(E')} \quad .$$

Thus, in principle, one can re-use data obtained from simulations performed at the fixed value T, in order to obtain the averages corresponding to the other temperature T', instead of simulating the same system again. This is the so-called *reweighting* method [9], which became famous after reinvented 3 decades later [10].

As an example, consider the Ising model on a $L \times L$ square lattice, with periodic boundary conditions. Only nearest neighbouring spins interact with each other through a coupling constant J. Normally, one counts energy $-J$ for each pair of parallel spins (*up-up* or *down-down*), and $+J$ otherwise. Here, we prefer an alternative counting: only pairs of anti-parallel spins (*up-down* or *down-up*) contribute with an energy $+2J$, while pairs of parallel spins do not contribute. For $L = 32$, the probability $P_T(E)$ is displayed in Fig. 8 for two slightly different temperatures $T = 2.247$ and $T' = 2.315$, one below and the other above the critical temperature $T_c = 2.269$ (all values measured in units of J). This narrow temperature range corresponds to 3% of T_c. Energy density means $E/2L^2$, the average energy per nearest neighbouring pair of spins (lattice bonds). Instead of using reweighting through the last equation, both curves were independently constructed.
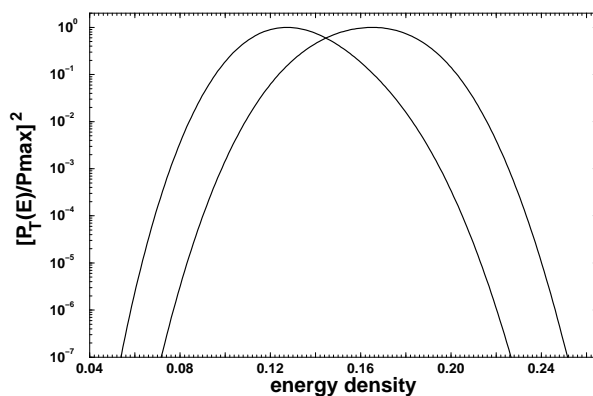

Figure 8. Squared energy probability distribution for the Ising model on a 32×32 lattice, for $T = 2.246$ (1% below $T_c = 2.269$) and $T' = 2.315$ (2% above T_c). For the sake of comparisons, both peaks were normalised to unit.

Note the vertical axis of Fig. 8, displaying the squared probability. On the other hand, the error is inversely proportional to the square root of the statitics, i.e. $\sqrt{M}$. Hence, this plot can be interpreted as the statistics one needs inside each energy level, in order to fulfill a previously given error tolerance. In order to obtain the right curve (T') from the left one (T), by using the reweighting equation, one would pay a price in what concerns accuracy. Near the maximum corresponding to T', the statistics for T is 10 times poorer. For this tiny 32×32 system size, the price to pay is not so bad. Consider now Fig. 9, showing the same probabilities

with the same temperatures, for a larger 90×90 lattice, a still small but 8-fold larger size. Instead of only 10, the loss in statistics corresponds now to a factor of 10 millions! The situation would be even worse for larger lattices, because the canonical distributions $P_T(E)$ are narrower. Reweighting does not help very much in these cases.
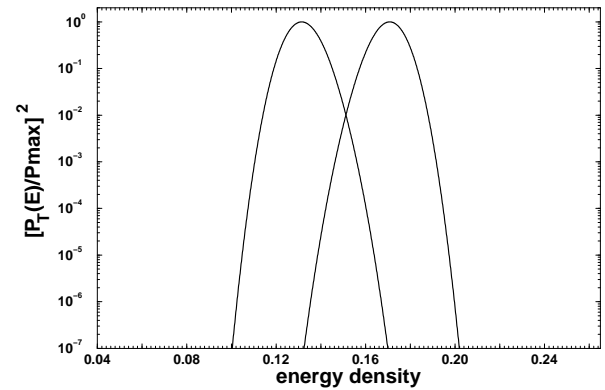

Figure 9. The same as Fig. 8, with the same temperatures, now for a 8-fold larger 90×90 lattice.

Reweighting methods based on the last equation do not solve the problem they are supposed to solve. One cannot extract the thermal averages for a temperature T' from Monte Carlo data obtained with another fixed temperature T, unless T and T' are very near to each other. One still needs to run the whole computer simulation again and again for a series of temperatures within the range of interest. However, as a solace, it solves another important problem: it allows one to get the thermal average $< Q >_T$ as a **continuous** function of T. After determining $< Q >_{T_1}$, $< Q >_{T_2}$, $< Q >_{T_3}$ ... by repeating the simulation for a discrete series of fixed temperatures T_1, T_2, T_3 ..., the above reweighting equation allows the interpolation in between these values [11].

The first successful attempt to obtain thermal averages over a broad temperature range from a single simulation was the multicanonical sampling [12]. In order to reach this goal, first one needs to abandon the Gibbs distribution $P_T(E)$, because it covers a very narrow energy range, hence a very narrow temperature range. Instead of the canonical dependence $P_T(E) \propto g(E)\,\mathrm{e}^{-E/T}$ which produces the undesired very narrow curves in Fig. 9, the simplest choice is a completely flat probability distribution, i.e. a constant $P_M(E)$ within the whole energy range of interest, where the label M stands for multicanonical. How to get such a flat distribution? The idea is to gradually tune the acceptance of candidates to belong to the Markov chain, in real time during the computer run: one rejects more often states corresponding to already over-populated energies, while less populated energies are accepted. In order to get a completely flat distribution, the acceptance rate corresponding to each energy E would be proportional to $1/g(E)$. However, $g(E)$ as well as $< Q(E) >$ are not known, otherwise the thermal average $< Q >_T$ could be directly obtained. Thus, one needs to measure $g(E)$ from the computer simulations. This can

be done during the run, by booking the number $A(E)$ of attempts to populate each energy E. If the final distribution is really flat, these numbers are proportional to $g(E)$, because only a fraction proportional to $1/g(E)$ of these attempts were really implemented, i.e. $A(E) = Kg(E)$. In this way, one determines $g(E)$ apart from an irrelevant multiplicative factor K which cancels out in the previous formula for $< Q >_T$. During the same run, the microcanonical average $< Q(E) >$ can also be measured, by determining the value of Q for each visited state, and accumulating these values on E-histograms. Again, different recipes could be invented in order to gradually tune the acceptance rate leading to a final flat histogram. Perhaps the most effective of them is proposed in [13], where a multiplicative tolerance factor f controls the distribution flatness. One starts with a large tolerance, and gradually decreases the value of f towards unity.

The broad histogram method [14] follows a different reasoning. Instead of defining a dynamic recipe, or rule, like Metropolis or multicanonical approaches, the focus is the direct determination of $g(E)$ by relating this function to two other **macroscopic** quantities $< N^{\mathrm{up}}(E) >$ and $< N^{\mathrm{dn}}(E) >$ defined as follows. First, one defines a protocol of allowed movements which potentially could be performed on the current state s. For instance, the single spin flips already mentioned. They are virtual movements, which **are not** implemented: they are not necessarily the same movements actually performed in order to define the Markov chain of states, from which one measures the desired quantities. They have nothing to do with acceptance rates or probabilities, each virtual movement can be only allowed or not. In order to avoid confusion, let's call **dynamic rule** the sequence of movements actually implemented in order to construct the Markov chain, with its characteristic acceptance probabilities, rejections, detailed balance arguments and so on. This rule is not part of the broad histogram method, any such a rule able to determine microcanonical averages is good.

The only restriction to the protocol of allowed virtual movements is the reversibility, i.e. if some movement $s \rightarrow s'$ is allowed, then $s' \rightarrow s$ is also allowed. Furthermore, one chooses a fixed energy jump ΔE: any choice is valid, thus the user can choose the more convenient value. For each state s corresponding to energy E, one counts the total number N_s^{up} of allowed movements one could potentially implement, increasing the energy to $E + \Delta E$: $< N^{\mathrm{up}}(E) >$ is the microcanonical average of this counting, over all $g(E)$ states, i.e.

$$< N^{\mathrm{up}}(E) > = \frac{1}{g(E)} \sum_{s(E)} N_s^{\mathrm{up}} \quad ,$$

the sum running over all states $s(E)$ with energy E. Analogously, for each such state, one counts the total number N_s^{dn} of allowed potential movements which result in an energy decrement of ΔE, i.e. the new energy would be $E - \Delta E$, and computes the microcanonical average

$$< N^{\mathrm{dn}}(E) > = \frac{1}{g(E)} \sum_{s(E)} N_s^{\mathrm{dn}} \quad .$$

These two microcanonical averages are related to $g(E)$ through the broad histogram equation [14]

$$g(E) < N^{\mathrm{up}}(E) > = g(E + \Delta E) < N^{\mathrm{dn}}(E + \Delta E) > \quad .$$

This equation is exact, completely general, valid for any model [15]. It follows from the required reversibility of the virtual movements one chooses as the protocol: the total number $\sum_{s(E)} N_s^{\mathrm{up}}$ of allowed movements which transform E into $E + \Delta E$, considering all $g(E)$ possible starting states, is the same number $\sum_{s(E+\Delta E)} N_s^{\mathrm{dn}}$ of reverse movements which transform back $E + \Delta E$ into E, now considering all $g(E + \Delta E)$ possible starting states.

The microcanonical averages $< N^{\mathrm{up}}(E) >$ and $< N^{\mathrm{dn}}(E) >$ can be determined by any means. If one knows how to get their exact values, the exact $g(E)$ can be obtained from the above equation. Otherwise, some approximate method should be adopted, for instance some Monte Carlo recipe. In this case, the only requirement the dynamic rule needs to fulfill is the uniform visitation among the states sharing the same energy, separately. The relative visitation rates to different energy levels is unimportant, one does not need to care about detailed balance and other complicated issues. Thus, the choice of the particular dynamic rule can be made within much more freedom than would be necessary for canonical averages. In particular, any dynamic rule which is good to determine canonical averages (Metropolis, multicanonical, etc) is equally good for microcanonical, but the reverse is not true.

Once the microcanonical averages $< N^{\mathrm{up}}(E) >$ and $< N^{\mathrm{dn}}(E) >$ are known as functions of E, one uses the above equation in order to determine $g(E)$. For instance, starting from the lowest energy E_0 of interest, and prescribing some arbitrary value for $g(E_0)$, this equation gives $g(E_1)$ where $E_1 = E_0 + \Delta E$. Now, from $g(E_1)$, one determines $g(E_2)$ where $E_2 = E_1 + \Delta E$, and so on. In this way, $g(E)$ is obtained along the whole energy range of interest, apart from an irrelevant multiplicative factor which cancels out when $< Q >_T$ is calculated. For that, the microcanonical average $< Q(E) >$ was previously determined in the same way as $< N^{\mathrm{up}}(E) >$ and $< N^{\mathrm{dn}}(E) >$, during the same computer run.

In order to implement this broad histogram method, when treating some generic problem, the only tool the user needs is a good calculator for microcanonical averages $< N^{\mathrm{up}}(E) >$, $< N^{\mathrm{dn}}(E) >$ and $< Q(E) >$. The accuracy obtained at the end depends only on the quality of this tool, because the broad histogram relation itself is exact, and does not introduce any bias.

The advantages of this approach over importance sampling, reweighting or multicanonical are three, at least. First, the already commented freedom to choose the microcanonical simulator, for which only a uniform visitation inside each energy level is required, independent of other levels. No need of a precise relation between visits to different levels (detailed balance, etc).

Second, and more important advantage, each visited state s of the Markov chain contributes for the averages with **macroscopic** quantities N_s^{up} and N_s^{dn}, instead of the mere counting of one more state accumulated on E-histograms.

This behaviour improves very much the accuracy. Moreover for larger and larger lattices, these macroscopic quantities provide a deeper probe to explore the internal details of each state: the computer coast to construct each new Markov state increases with system size, then it is not a profitable approach to extract only the information concerning how many states were visited, or how many attempts were booked, as importance sampling or multicanonical approaches do. Within broad histogram, on the other hand, the number $V(E)$ of visits to each energy level plays no role at all: it needs only to be large enough to provide a good statistics, nothing more. The real information extracted from each Markov state s resides on the values of N_s^{up} and N_s^{dn}.

Third advantage, following the previous one, the profile of visits to the various energy levels are not restricted to some particular form, canonical or flat. The user can choose it according to the desired accuracy. For instance, if one is interested in getting canonical averages over the temperature range between $T = 2.246$ and $T' = 2.315$, for the Ising model exemplified in Fig. 8, the ideal profile is shown in Fig. 10, where the number of visits $V(E)$ is normalised by its maximum value. This profile was obtained by superimposing the canonical distributions corresponding to the extreme temperatures within the desired interval, Fig. 8, forming an umbrella. Only inside a narrow energy range, between densities 0.124 and 0.170, one needs a large number $V_{\max}$ of visits. Outside this range, one can sample a smaller number of Markov states, saving computer time without compromising the overall accuracy. These advantages and others are exhibited with details in [16].

A simple dynamic rule [17] seems to be a good microcanonical calculator. Combined with the broad histogram, it gives good results. It was inspired by an earlier work related to a distinct subject [18]. Although one can use it for general situations, an easy generalisation, let's consider, for simplicity, the nearest neighbour Ising model on a square lattice, within the single spin flip dynamic rule. The energy spectrum consists of equally spaced levels, separated by a gap of $4J$, considered here as the energy unit. By flipping just one spin, the energy E can jump to $E+2$ or $E+1$, stay at E, or decrease to $E-1$ or $E-2$, depending on its four neighbours. In order to measure canonical averages at fixed E, the quoted dynamic rule is: 1) start from some random state with energy in between $E-2$ and $E+2$, an interval covering 5 adjacent levels; 2) accept any tossed movement which keeps the energy inside this interval, rejecting all others; 3) accumulate statistics only for the central level E, whenever the current state falls there, by chance, along its five-level-bounded Markov walk. One predicate in favour of this rule is the complete absence of rejections within the single averaged level E: any movement *to it* or *out it* is accepted. Although this property alone is not a guarantee of uniform visitation within E (because rejections occur for visit attempts to other neighbouring levels, ancestors of each visit to E along the Markov chain), we believe the sampling performance of this dynamic rule is good. Let's see the results.
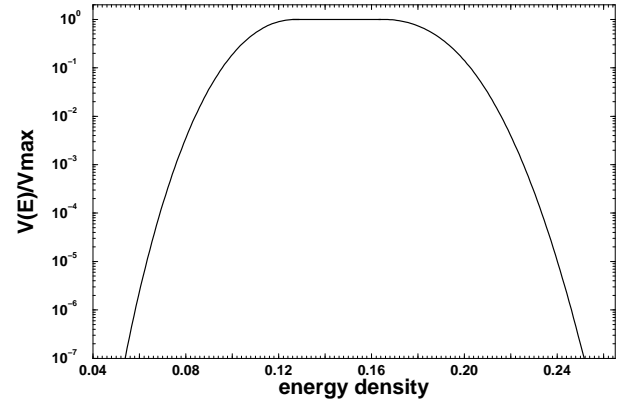

Figure 10. Optimum profile of visits along the energy axis, for the Ising model on a 32×32 lattice, in order to obtain equally accurate canonical averages between $T = 2.246$ and $T' = 2.315$. Compare with Fig. 8.

This dynamic rule was applied according to a previously planned visit profile, equivalent to Fig. 10, covering **continuously** an entire interval of temperatures within a uniform accuracy. For a 32×32 lattice, the exact partition function is available [19], and we can compute the deviations obtained for the specific heat, for instance. At the peak, in particular, these deviations are the largest, shown in Fig. 11 as a function of the statistics. Different symbols correspond to the possible values $\Delta E = 1$ or 2, fixed for the energy jump in the broad histogram equation.
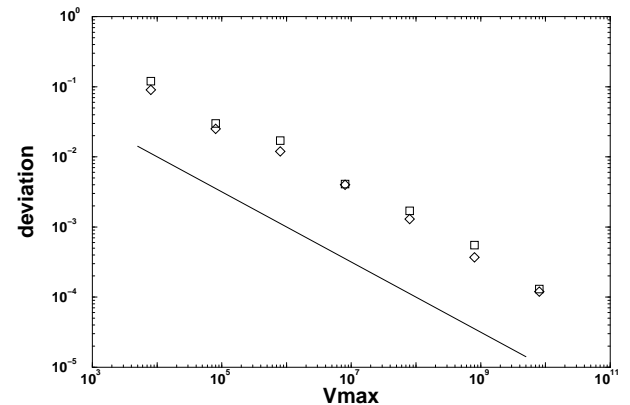

Figure 11. Deviations relative to exact values of the specific heat peak, 32×32 Ising model. The line shows $1/\sqrt{V_{\max}}$, in order to compare the slopes. Diamonds correspond to $\Delta E = 1$, and squares to $\Delta E = 2$.

This plot shows in practice that broad histogram method does not introduce any further bias, besides the normal statistical fluctuations. Another interesting quantity is the correlation between opposite surfaces of the finite lattice [5]. For the $L \times L$ square lattice Ising model, with periodic boundary conditions and even L, it is defined as follows. For a given state, one verifies the spins along a fixed horizontal row i, and books $S_i = +1$ or -1 if their majority points *up* or *down*, respectively. On ties, one books $S_i = 0$. Then, one repeats the same for another parallel row j, distant $L/2$ from i, and books $S_j = +1$, -1 or 0. Finally, one

determines the product $r = S_i S_j$, and perform the average of this quantity over all L possible pairs of parallel lines distant $L/2$ from each other (rows and columns). The quoted surface correlation is the canonical average $< r >_T$ of this quantity.

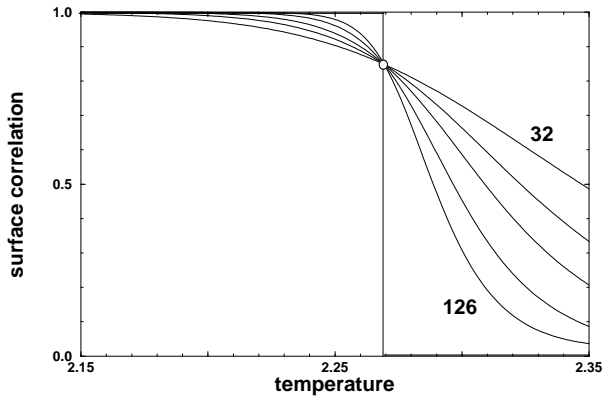
Figure 12. Surface correlation function for the $L \times L$ square lattice Ising model, $L = 32, 46, 62, 90$ and 126. For larger and larger lattices, these plots approach the step function defined by the bold straight lines plus the circle.

In the thermodynamic limit $L \to \infty$, this quantity is a kind of order parameter, Fig. 12, vanishing above the critical temperature T_c. For increasing lattice sizes, it approaches a step function: $< r >_T = 1$ below T_c; $< r >_T = r^*$ at T_c, where r^* is a universal value; and $< r >_T = 0$ above T_c. The distinguishing feature of $< r >_T$, compared with the usual order parameter (the magnetisation $< m >_T$) is its constant behaviour **below** the critical temperature. On one hand, $< m >_T \sim (T_c - T)^\beta$ near and below T_c. On the other hand, also below T_c one has $< r >_T = 1$, hence, instead of β, the corresponding exponent **vanishes** [5]. For finite systems, considering the finite size scaling hypothesis, this comparison can be better appreciated within the generalised homogeneous forms

$$m[\lambda^{1/\nu}(T - T_c), \lambda L^{-1}] = \lambda^{-\phi_m} \, m(T - T_c, L^{-1}) \;\;,$$

where the alternative notation $< m >_T = m(T - T_c, L^{-1})$ is adopted, and

$$r[\lambda^{1/\nu}(T - T_c), \lambda L^{-1}] = \lambda^{-\phi_r} \, r(T - T_c, L^{-1}) \;\;,$$
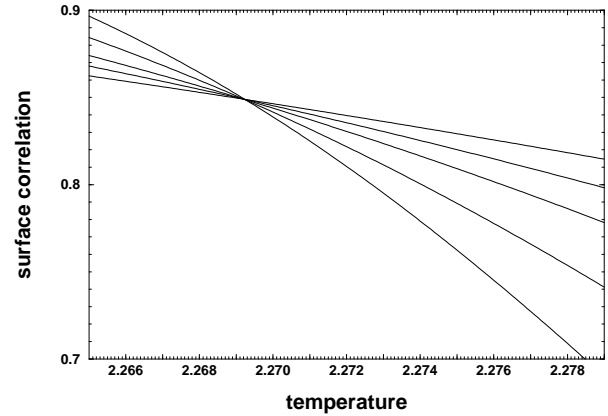
where $< r >_T = r(T - T_c, L^{-1})$.

Figure 13. A zoom performed on the previous plot.

These forms are asymptotically valid for large L and near T_c, where λ is an arbitrary length. By choosing $\lambda = |T - T_c|^{-\nu}$ and taking the limit $L \to \infty$, one finds $\phi_m = -\beta/\nu$ and $\phi_r = 0$. On the other hand, by choosing $\lambda = L$ and tuning the system exactly at T_c, one finds $< m >_T \sim L^{\phi_m} = L^{-\beta/\nu}$ and $< r >_T \sim L^{\phi_r} = L^0$. At T_c and for large L, $< r >_T$ becomes size independent, i.e. $< r >_T = r^*$. A possible approach to find the critical temperature T_c is by plotting $< r >_T$ against T for two different lattice sizes, and finding the crossing of both curves. Fig. 12 shows $< r >_T$ for 5 different lattice sizes, obtained by the broad histogram method. The maximum number $V_{\max}$ of sampled states per energy level was 8×10^9 for $L = 32$ and 46, and 8×10^8 for $L = 62, 90$ and 126.
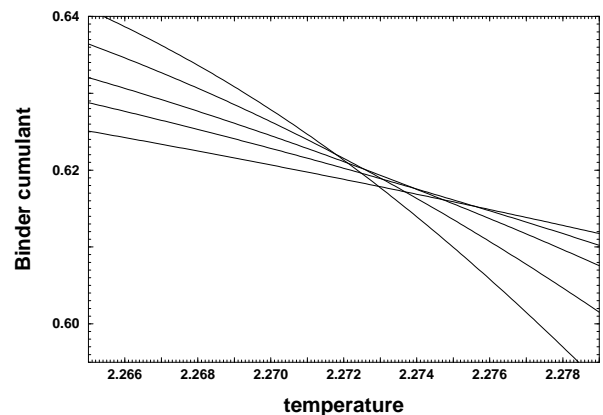
Figure 14. The Binder cumulant within the same zoom.

The so-called Binder cumulant $< u >_T$ [4] also presents zero anomalous dimension. It was very often used in order to find critical temperatures of many systems, following the same idea of finding the crossing between two curves $< u >_T$ against T, obtained for different lattice sizes. This 23-years-old method is generally considered one of the most accurate ways to study critical systems. Fig. 13 shows a zoom of the last Fig. 12, while Fig. 14 shows the Binder

cumulant obtained from the same computer runs, thus with the same accuracy. The surface correlation function seems to work better than the Binder cumulant. The reason for that may be traced from the scaling behaviour on three distinct points, namely $T = 0$, $T \to \infty$ and, of course, the critical point $T = T_c$. The system is invariant under scaling transformations not only at the critical point, but also for $T = 0$ and $T \to \infty$. Thus, besides the thermal exponent ν and the magnetic exponent δ, both describing the critical behaviour near T_c, there are other (non-critical) exponents related to the lattice dimensionality which describes the behaviour of the system under scaling transformation near $T = 0$ and near $T \to \infty$. By construction, both the Binder cumulant $< u >_T$ and the surface correlation function $< r >_T$ respect the correct critical behaviour near T_c. However, only $< r >_T$, not $< u >_T$, does the same near $T = 0$ and $T \to \infty$ [5]. Thus, being forced to obey the correct behaviours in three different points along the T axis, $< r >_T$ has less freedom to deviate from the correct $L \to \infty$ behaviour, even for small system sizes.

4 Self-Organized Criticality

4.1 Introduction

Self-organized criticality (abbreviated as SOC from here on) describes a large and varied body of phenomenological data and theoretical work. Because its current meaning has evolved far from any initial precise meaning the term may have had, we will in these notes adhere to a broad characterization, if only for pedagogical purposes.

The original motivation for SOC came from recursive mathematics. Iterated maps have been show to generate fascinating images - *fractals* - which are now in common usage, both in the sciences and in popular culture. Some of these patterns looked very similar to naturally occurring ones, such as river erosion patterns, plant and leaf structure, and geological landscapes, to name but a few. The question about which kinds of dynamical mechanisms would be required to create such patterns in nature is, thus, unavoidable. One would hope, in addition, this mechanisms to be fairly simple: very little would be gained in our understanding if they were as complex as the patterns they generate.

The breakthrough in this field can undoubtedly be attributed to a seminal paper by Back, Tang, and Wiesenfeld (BTW) in 1987 [20], that opened the path to a variety of computational models that show SOC behavior. These models are also successful in demonstrating the possibility of generating complex and coherent structures from very simple dynamical rules. A key feature of these models is that they are capable of generating spatially distributed patterns of a fractal nature without requiring long-range interactions in their dynamics: the long-range correlations "emerge" from short-range interaction rules. Another key feature of these models is the fact that they converge to "absorbing" configurations exhibiting long-range coherence even having not started at one. This convergence does not require the tuning of any external parameter, and thus these models are said to be "self-organized".

Criticality is a notion that comes from thermodynamics. In general, thermodynamic systems become more ordered and decrease their symmetry as the temperature is lowered. Cohesion overcomes entropy and thermal motion, and structured patterns begin to emerge. Thermodynamic systems can exist in a number of phases, corresponding to different orderings and symmetries. Phase transitions that take place under the condition of thermal equilibrium, for which temperature is a well-defined quantity and the free energy is minimized, are well understood. These equilibrium systems are ergodic and have had time to settle into their most probable configuration, allowing simplifying statistical assumptions to be made. Although most phase transitions in nature are discontinuous, in the sense that at least one response function - a first-order derivative of the free energy - suffers a discontinuous jump at the transition point, some systems may show a critical point at which this discontinuity does not happens. As a system approaches a critical point from high temperatures, it begins to organize itself at a microscopic level. Near but not exactly at the critical point, correlations have an exponential decay - as an example for the Ising model, where $S(\vec{r})$ is the spin, the spin-spin correlation has the typical behavior $< S(\vec{r})S(0) >_\tau - < S(0) >_\tau^2 \sim exp(-r/\xi)\ r^{-\mathcal{A}}$, where ξ is called the correlation length. Close to the critical point, large structural fluctuations appear, yet they come from local interactions. These fluctuations result in the disappearance of a characteristic scale in the system at the critical point - in the example above, the same spin-spin correlation function is now $< S(\vec{r})S(0) >_\tau - < S(0) >_\tau^2 \sim r^{-\mathcal{A}}$, which amounts to having $\xi \to \infty$ in the former equation. It is fair to say that the lack of a characteristic scale is the hallmark of a critical process. This behavior is remarkably general and independent of the system's dynamics.

The kind of structures that SOC aspires to explain look like equilibrium thermodynamic systems near their critical point. Typical SOC systems, however, are not in equilibrium with their surroundings, but have instead non-trivial interactions with them. In addition, SOC systems do not require the tuning of an external control parameter, such as temperature, to exhibit critical behavior. They tune themselves to a point at which critical-type behavior occurs. This critical behavior is the bridge between fractal structures and thermodynamics. Fractal structures appear the same at all size scales, hence they have no characteristic scale, just as the fluctuations near an equilibrium phase transition at a critical point. It is then logically compelling to seek a dynamical explanation of fractal structures in terms of thermodynamic criticality. A proper understanding of SOC requires an extension of the notion of critical behavior to non-equilibrium thermodynamical systems and hence an extension of the tools required to describe such systems.

The absence of a characteristic scale is closely related to the idea of scale invariance, as can be exemplified by the way they are connected in the study of geometric fractals. There, the absence of a scale stems from self-similarity, in the sense that any subset of the curve magnified by an appropriate amount cannot be distinguished from the full curve. Building on this same example, self-similarity in a fractal

generated from an initial element by an iterated map implies that the number of elements $N(s)$ has a power-law dependence on their size s. This is precisely what we mean by scale invariance, since $f(x) = x^{\mathcal{A}} \rightarrow f(kx)/f(x) = k^{\mathcal{A}}$.

The main hypothesis contained in the work of BTW is that a dynamical system with many interacting elements, under very general conditions, may self-organize, or self-tune, into a statistically stationary state with a complex structure. In this state there are no characteristic scales of time and space controlling the time evolution, and the dynamical responses are complex, with statistical properties described by power-laws. Systems with very different microscopic dynamics can present power-law responses with the same exponents, in a non-equilibrium version of Wilson's idea of universality: These self-organized states have properties similar to those exhibited by equilibrium systems at their critical point.

This mechanism is a candidate to be the long sought dynamical explanation for the ubiquity in nature of complex space-time structures. A list of phenomena that could exhibit SOC behavior include the avalanche-response of sand piles, earthquakes, landslides, forest fires, the motion of magnetic field lines in superconductors and stellar atmospheres, the dynamics of magnetic domains, of interface growth, and punctuated equilibrium in biological evolution.

What are the ingredients that lead a driven dynamical system to a SOC state? From the analysis of the computational models mentioned at the beginning of this section, a key feature is a large separation between the time scales of the driving and the relaxation processes. This can be ensured by the existence of a threshold that separates them. As an example, let us consider the physics of driven rough interfaces, which is believed to be one of the main earthquake source processes. This connection will be analyzed in more detail in the next section, but for our present purposes it is enough to mention that the separation of time scales is in this case provided by static friction, which has to be overcome by the increasing stress between the two sides of a geological fault.

A SOC state is thus a statistically stationary state, said to be "marginally stable", sharing with thermodynamical critical states the property of space-time scale invariance, with algebraically decaying correlation functions. It is perhaps best seen as a collection of metastable states, in the thermodynamical sense. The system is driven across phase space, spending some time at the neighborhood of each of its metastable states, and moving between these states after some large avalanche of relaxation events.

The original goal of BTW was to provide an explanation for the frequent occurrence in nature of fractal space and time patterns. The latter are usually referred to as one-over-f ($1/f$) noise, indicating the scale invariance of the power spectrum. The SOC dynamical answer to this quest goes as follows: In driven extended threshold systems, the response (*signal*) evolves along a connected path of regions above the threshold. Noise, generated either by the initial configuration or built in the dynamics, creates random connected networks, which are modified and correlated by the intrinsic dynamics of relaxation. The result is a complex patchwork of dynamically connected regions, with a sparse geometry resembling that of percolation clusters. If the activated region consists of fractals of different sizes, the energy release and the time duration of the induced relaxation events that travel through this network can vary enormously.

4.2 Characterization of the SOC state

The nature of the state is best described by the system's response to an external perturbation. Non-critical systems present a simple behavior: The response time and the decay length have characteristic sizes, their distribution is narrow and well described by their first moment (average response). In critical systems, on the other hand, the same perturbation can generate wildly varying responses, depending on where and when it is applied. The statistical distributions that describe the responses have the typical functional form $P(s) \sim s^{-\tau}$ and $Q(t) \sim t^{-\alpha}$. These distributions have a lower cutoff s_1 and t_1, defined by the scale of the microscopic constituents of the system. For finite systems of linear size L, the distributions present a crossover, from a certain scale up, to a functional form such as $P(s) \sim exp(-s/s_2)$, $s > s_2$. For a genuine critical state, one must have $s_2 \sim L^w$, $w > 0$. If the exponent of the distribution is $\omega < 2$, it has no mean in the thermodynamic limit; if it is < 3, its second moment and width diverge.

The temporal fluctuations are characterized by a $1/f$ noise. This is a rather imprecise label used to describe the nature of certain types of temporal correlations. If the response of a physical system is a time-dependent signal $N(\tau)$, the temporal correlation function $G(\tau)$, defined by

$$G(\tau) = < N(\tau_0)N(\tau_0 + \tau) >_{\tau_0} - < N(\tau_0) >_{\tau_0}{}^2, \quad (1)$$

describes how the value of the signal at some instant in time $N(\tau_0)$ has a statistical significance in its value at some later time $N(\tau_0 + \tau)$. If there is no statistical correlation, $G(\tau) = 0$. The rate of decrease of $G(\tau)$, from $G(0)$ to 0, measures the duration of the correlation, and is related to memory effects in the signal.

The power spectrum of the signal $N(\tau)$ is related to its Fourier transform

$$S(f) = \lim_{T \to \infty} \frac{1}{2T} \left| \int_{-T}^{T} N(\tau) exp(2i\pi f\tau) d\tau \right|^2 \quad (2)$$

For a stationary process,

$$S(f) = 2 \int_{0}^{\infty} G(\tau) cos(2\pi f\tau) d\tau \quad (3)$$

and temporal correlations may be discussed in terms of the power spectrum.

The special nature of $1/f$ fluctuations can be seen by a simple example. Suppose that, for some temporal signal, we have $S(f) \sim 1/f^{\beta}$ and $G(\tau) \sim 1/\tau^{\alpha}$; then, $\rightarrow 1/f^{\beta} \sim 1/f^{1-\alpha}$: if $\beta \sim 1$, $\alpha \sim 0$. If $\beta = 1$, $G(\tau)$ has a logarithmic decay, showing that temporal correlations have a very long range if $\beta \simeq 1$.

It is interesting to note that the sandpile model proposed as an archetypical SOC system by BTW, even though showing long range time correlations, does *not* present $1/f$ noise, despite the claims to the contrary originally made by its proponents.

Let us now discuss the spatial correlation function. For a system described by a dynamical variable which is a field $n(\mathbf{r}, \mathbf{t})$, such as the local density (or the local magnetization) for fluid (or magnetic) systems this function is

$$G(r) = \; < n(\mathbf{r_0} + \mathbf{r}) \, n(\mathbf{r_0}) >_{\mathbf{r_0}} - < n(\mathbf{r_0}) >_{\mathbf{r_0}}{}^2 \quad (4)$$

where the brackets indicate thermal and positional averages.

If $T \neq T_c$, $G(r) \sim exp(-r/\xi)$, where ξ is the correlation length. If $T \to T_c$, $\xi \sim \mid T - T_c \mid^{-\nu}$. At $T = T_c$, $G(r) \sim r^{-\eta}$. The divergence of ξ indicates the absence of a characteristic length scale and leads to spatial scale invariance.

4.3   Systems that exhibit SOC

Model systems have been shown to fulfill the SOC characterization. But this is not enough for the SOC program: one should examine real physical systems and identify experimentally SOC behavior, if this concept is to be at all significant to our understanding of nature. In the experiments, one usually collects data about the statistical size distribution of the dynamical responses, arising from the relaxation avalanches. But what should we look for in this data? One possibility, raised by our previous comments, would be to measure time-dependent dynamical quantities and compute their power spectrum. Unfortunately, it turns out that a power spectrum of the form $\sim 1/f^{\beta}$, $\beta \simeq 1$, is a necessary, but not sufficient condition for critical behavior, thus for SOC. One has also to identify the presence of spatial fractals in order to be able to characterize a real-life SOC system.

Experiments measuring the avalanche signals of sandpile-like systems, which mimic the original BTW model, failed in all but one case to show evidence of SOC behavior. Typically, in these systems, sand is slowly added to a pile. In one experiment, for example, this pile had a circular base, and was set on top of a scale, which could monitor fluctuations of the total mass. Small piles did show scale invariance. As the diameter of the base increases, a crossover to an oscillatory behavior is observed: Small avalanches in real sandpiles do show a behavior consistent with SOC, which disappears when the mean avalanche size increases. In order to investigate inertial effects that could be responsible for the destruction of the incipient SOC state, experiments were performed with one-dimensional rice piles, with grains differing by their aspect ratio. In this case, the signal observed was the potential energy release. For small and nearly spherical grains (aspect ratio ~ 1), the distribution of event sizes is a stretched exponential of the form $P(E) \sim exp[(-E/E_0)^{\gamma}]$. For elongated grains, with a larger aspect ratio, this distribution, obtained through a finite-size-scaling of the data, is $P(E) \sim E^{-\alpha}$, with no fi-

nite size cutoff [21]. This scale invariance is consistent with SOC behavior.

Another strong candidate for SOC behavior in the real world is the avalanche response of geological plate tectonics, the earthquakes. Earthquakes result from the tectonic motion of the plates on the lithosfere of our planet. This motion, driven by convection currents in the deep mantle, generates increasing stresses along plate interfaces, since sliding is avoided by static friction. Plate speeds being typically in the range of a few cm/year, which amounts to a slow drive, and the threshold dynamics implicit in friction, are the two ingredients that can generate SOC for this planetary system. In fact, the distribution of earthquake occurrence has been shown to obey Gutenberg-Richter's law: it is a broad distribution, with $P(E) \sim E^{-B}$, $B \sim 1.8 - 2.2$, and some geographical dependence on the exponent. There is some controversy in the geophysical literature about the validity of this statistical description, and some claim that a characteristic, periodic regime can be observed for some faults, for events at the far end (large size) of the distribution.

The cascades of species extinctions that result from biological evolution are other phenomena for which claims of SOC behavior have been stated. Paleontological evidence, though sparse and debatable, seems to point towards a description of a system in "punctuated equilibrium", a term that invokes long periods of relative quiescence ("equilibrium") interrupted ("punctuated") by short, in the time scale of evolution, periods of frenetic activity, in which large numbers of species disappear. In contrast to a gradual process of species extinction and creation, such a view is underlined by an understanding of extinction as a result of fluctuation in environment pressure, rather than to species obsolescence. As a species disappear because of its poor fitness to the changing environment, the effects will propagate through a variable-length network of neighboring species in the food chain, in an avalanche of extinctions. The distribution of avalanche sizes appear, from the known evidence, to exhibit scale invariance, thus the already mentioned claim for SOC behavior.

5   Computational Models

Computational models of dynamical systems have been the main tool used for the inquiries about the existence and characterization of the SOC state. In a very general setting, these models have some common features. Their dynamics are described by the rules for the time evolution of one or more dynamical variables, or fields, defined on an extended region of space. These variables may be associated with the local slope or height, for sandpiles, strain and/or stress for earthquake models, and fitness of a species, in evolutionary models. This field is updated at each location and each discrete time step following some rule or algorithm. The choice of the updating algorithm is the key to the success of the model, requiring ingenuity, physical intuition and possibly some luck - or thorough testing! - to come up with an interesting idea. Successful models have usually very simple rules, but complex behavior, deriving from the large number of individual degrees of freedom of the system, rather than

from complicated dynamics. The emergence of complexity out of simple dynamics is the outcome of these models. They play a role much similar to the Ising model in the theory of magnetism and critical phenomena: they are "toys" that illuminate the main features required for the appearance of complex patterns and behavior.

The updating algorithm is a rule F by which the value of the dynamical variable $\sigma(\vec{r}, t)$ at some location $\vec{r}$ changes as discrete time t flows:

$$\sigma(\vec{r}, t + 1) = F(\sigma(\vec{r_i}, t)) \tag{5}$$

It is a function of the value of the variable at that particular location $\sigma(\vec{r}, t)$ at time t and of its value in a neighborhood. In this sense, we may call it a cellular automaton, except that now the value of the dynamical variable may be either an integer or a real number - we will use the adjective "continuous" for the latter case.

5.1 Sandpiles: a conservative model

The prototypic SOC model is the sandpile model (BTW). It is a conservative model, in the sense that the dynamical variable added over all sites does not change during relaxation, except possibly when sites at the border also relax. It is a metaphor for the behavior of a real sandpile, where stability is controlled by a local gradient (slope): If $z(\mathbf{r})$ is the slope at site $\mathbf{r}$ and z_c is a critical value of this slope, than if $z(\mathbf{r}) > z_c$ the sandpile becomes unstable. The model is slowly driven by the addition of grain and relaxation takes over when the sandpile becomes unstable as a consequence. There are, thus two dynamical rules, one for the driving and another one for the relaxation.

For a one-dimensional sandpile, this gradient model may be cast as follows: Let h_i be the (integer) height of a $N + 1$-long pile at site i. The local slope is $z_i = h_{i+1} - h_i$. The addition of a single grain at site i, resulting from a random choice, increases the local height by 1 and changes the slope according to the rule $z_{i-1} \rightarrow z_{i-1} + 1$, $z_i \rightarrow z_i - 1$. Threshold dynamics of relaxation is triggered by the condition $z_i > z_c$, where z_c is a pre-assigned critical slope. Site i becomes overcritical and relaxes via $z_i \rightarrow z_i - 2$, which causes an update of the neighboring slopes $z_{i-1} \rightarrow z_{i-1} + 1$ and $z_{i+1} \rightarrow z_{i+1} + 1$. These may in turn exceed the critical value, and relaxation proceeds until all sites become undercritical. The sequence of relaxations that follows after a site becomes overcritical is the avalanche response of the system. The border at $i = 0$ is closed, and $z_0 = 0$. At $i = N$ the border is open, and when $z_N > z_c$, then $z_{N-1} \rightarrow z_{N-1} + 1$, $z_N \rightarrow z_N - 1$. The relaxation dynamics do not change the total value of z, and the model is called conservative because of this feature.

A simpler and more popular, though less intuitive, version of the sandpile model is the so-called height model. Now, the dynamical variable is z_i and the addition of a grain at site i implies simply that $z_i \rightarrow z_i + 1$. The threshold condition and relaxation dynamics are still the same as for the previous version.

The one-dimensional sandpile has a rather trivial behavior. The configuration $z_i = z_c \ \forall i$ is the unique global attractor, or absorbing state, of the dynamics. It is a minimally stable configuration: the addition of a single grain leads to relaxation. It is also a critical state, in the sense that a small perturbation will cause an avalanche with any size, limited only by the size of the system. The probability distribution of avalanche sizes is trivial: if $P(s)$ ($P(t)$) is the probability that s sites (t time steps) are involved in the relaxation process, than $P(s) = P(t) = 1/(N+1)$ if $1 \leq s(t) \leq (N+1)$, 0 otherwise.

A much more interesting version is the multi-dimensional sandpile. The higher dimensionality avoids the unique absorbing state, and the probability distributions are no longer trivial. The dimensional extension of the model is straightforward: Let $z(\mathbf{r})$ be an integer-valued field defined on a d-dimensional hyper-cubic lattice, and $\mathbf{e_i}$ the lattice vectors. We may distinguish between a conservative perturbation, in which the system is driven by $z(\mathbf{r}) \rightarrow z(\mathbf{r}) + \mathbf{d}$, $z(\mathbf{r} - \mathbf{e_i}) \rightarrow z(\mathbf{r} - \mathbf{e_i}) - \mathbf{1}$, and a non-conservative one, for which $z(\mathbf{r}) \rightarrow z(\mathbf{r}) + \mathbf{1}$. In both cases, relaxation of an over-critical site, for which $z(\mathbf{r}) > z_c$, proceeds through $z(\mathbf{r}) \rightarrow z(\mathbf{r}) - \mathbf{2d}$, $z(\mathbf{r} \pm \mathbf{e_i}) \rightarrow z(\mathbf{r} \pm \mathbf{e_i}) + \mathbf{1}$.

This version of the sandpile model is usually studied for closed boundaries, for which $z(\mathbf{r}) = \mathbf{0}$ if any component $r_i = 0$, and for open boundaries, in which case one has the rule that if a toppling site has any component $r_i = L$, relaxation proceeds through

· $z(\mathbf{r}) \rightarrow z(\mathbf{r}) - \mathbf{2d} + \mathbf{n}$, $n = (\sharp$ components which are $= L)$
· $z(\mathbf{r} + \mathbf{e_i}) \rightarrow z(\mathbf{r} + \mathbf{e_i}) + \mathbf{1}$ if $r_i \neq L$
· $z(\mathbf{r} - \mathbf{e_i}) \rightarrow z(\mathbf{r} - \mathbf{e_i}) + \mathbf{1}$
· $z(\mathbf{r}) = \mathbf{0}$ if $r_i = 0$ for some i.

The update algorithm processes all sites in parallel, and one time step is counted after every site in the lattice is examined. After an initial transient, the statistical properties of the model become time independent and do not depend on the initial configuration (e.g. $< z > (t) \rightarrow < z > = \lim_{T \rightarrow \infty}(1/T)\int_0^T < z > (t)dt$).

The attractor of the model is a non-trivial set of configurations, which contain some that are not marginally stable. For the characterization of the critical response of the model, let A be a connected region stricken by an avalanche, $\sharp A = |A|$, and
• $\mathbf{R}_{CM} = (1/|A|)\sum_{\mathbf{r} \in A} \mathbf{r}$
• $l = (1/|A|)\sum_{\mathbf{r} \in A} |\mathbf{r} - \mathbf{R}_{CM}|$ (linear size);
• s = total number of toppled sites, which is a proxy for the total energy dissipated in the relaxation process;
• t = total number of time steps, or the duration of an avalanche.

Then, the model has a critical behavior when lying on the attractor, in the sense that the distributions of sizes and durations of the avalanches are power-laws: $P(l) \sim l^{-\gamma}$, $P(s) \sim s^{-\tau}$, $P(t) \sim t^{-\alpha}$. These three exponents are not independent, and a scaling law may be established among them. The upper critical dimension of the model is 4, and the mean-field (MF) Fisher exponent is $\tau_{MF} = 3/2$. These results were firmly established by an exact solution for the Abelian sandpile in two dimensions [22]. Computer simulations have shown that in three dimensions this same exponent is $\tau_{3D} \simeq 1.3$ [23].

Since most studies of this and similar models are per-

formed through computer simulations, let us discuss one such algorithm. Reference [24] has a simple example, which works well at an introductory level, but is not particularly efficient. Some text books on computer simulations, such as Ref. [25], present similar algorithms, with pedagogical emphasis. We will present here a generic version, that uses a circular stack to store indexes of sites that will topple and relax in this and the next time step of the simulation. The structure of the algorithm, in a self-explanatory metalanguage, is:

- initialization (e.g. $z_i = 0 \ \forall i$)
- main loop: *do*
- selection of the site i to which a grain will be added

• if $z_i > z_c \rightarrow$ avalanche
• if time $>$ transient, collects statistical data until time is up.

    - avalanche:
• $i \rightarrow$ stack
• while stack is not empty, *do*:
· unstacks next toppling site
· redistributes sand to neighbors (n); if $z_n > z_c$ for a neighbor n, stacks it $n \rightarrow$ stack

We present below a code that implements this algorithm, written in plain C.

```c
//adds grain to site (m,n)
void stack(m,n) {
  z[m][n] = z[m][n] + 1; ztot = ztot + 1;
  if (z[m][n]==(zcrit+1)) {
    x[next] = m; y[next] = n;
    next = next + 1;
    if (next==first) {
      file=fopen("err.dat","w");
      fprintf(file,"\n   STACK OVERFLOW!!!!");
      fclose(file);
    }
//  circular stack
    if (next==stacksize) next = 0;
  }
}


// avalanche algorithm
void avalanche() {
  size = 0;
  do {
//  stack is FIFO
    i = x[first]; j = y[first];
        /* unstacks site to be examined */
    first = first + 1;
        /* points to next site to be unstuck */
    if (first==stacksize) first = 0;
        /* circular stack */
    z[i][j] = z[i][j] - 4;
    ztot = ztot - 4;
    size = size + 1;
//  open boundary conditions
    if (i<(L-1))  stack(i+1,j);
    if (j<(L-1))  stack(i,j+1);
    if (i>0) stack(i-1,j);
    if (j>0) stack(i,j-1);
  } while (!(next==first));
//  resets stack pointers
  first = 0; next = 0;
}

main() {
  initialization();
//  main loop
```

```
    for (time=1; time<=TOTIME; time++) {

//  chooses random site
        i = random(L); j = random(L)

//  adds grain of sand
        stack(i,j);
        if (!(next==first)) avalanche();

        if (time>transient) collects_statistics();
    }
    files_results();
}
```

5.2   Earthquakes: a dissipative model

The frequency distribution of earthquakes appears to be the most well documented evidence of critical behavior in a natural system, as synthesized by the Gutenberg-Richter law mentioned above. The first model that succeeded in representing the basic physics involved in tectonic processes and was able to obtain a frequency distribution consistent with that law was due to Burridge and Knopoff [26]. In its first version, the model was a one-dimensional chain of massive blocks, representing the asperities in the boundary between two moving tectonic plates, connected to each other by elastic forces (springs). The driving mechanism was modeled as a set of additional springs, connecting each block to a moving rod, and the threshold dynamics was ensured by friction between blocks and the ground. The set of coupled second-order differential equations representing this physics was then solved numerically, with a friction force that decreases with velocity.

Computational models are able to increase greatly, both in number of elements and dimensionality, the ability to investigate the above physics. The trade-off is to eliminate the massive terms of the equations, by considering that these are related to seismic waves, that carry typically less than 10% of the energy released in an earthquake. By so doing, the differential equations can be easily discretized, and the resulting model is a real-valued - or continuous - cellular automaton (CCA). Several models with these characteristics appeared in the literature in recent years - Refs. [[27, 28, 29]] are a few of those - and the most successful of those, at least in the physics community, was that of Olami-Federsen-Christensen (OFC) [30], which were the first to have claimed the observation of SOC in a non-conservative model. Using the language of this latter model, the physics is contained in the dynamics of a single field, a real-valued dynamical variable E_i, a metaphor for the stress, defined on the $N = L^d$ sites of a d-dimensional cubic lattice. The dynamics of this field is completely deterministic, except for the initial configuration, which sets the initial value of the field at a site i by a random choice $E_i \in [0, E_c)$, where the critical value E_c is usually set to one. The driving is uniform and homogeneous, and at each step of the driving time scale the field evolves with $E_i \rightarrow E_i + \nu$, $\forall i$, where ν is

an indirect measure of this time scale. Relaxation, which happens when a site i becomes critical $E_i \geq E_c$ and topples, follows the rule $E_i \rightarrow 0$ for this event initiator site and $E_{nn} \rightarrow E_{nn} + \alpha E_i$ for its q_i neighbors. These sites may, in turn, become critical, and the process proceeds until $E_i < E_c$, $\forall i$. This cascade, or avalanche, of topplings is the proxy for an earthquake in the model, and data can than be collected on the statistical distribution of such events. The parameter α has an important role, since it measures the intrinsic dissipation ratio of these dynamics: the amount of stress that is lost by the system after site i topples is $E_{dis} = (q_i\alpha - 1)E_i$, and the system would be called conservative if $\alpha = \max q_i{}^{-1}$. One is usually interested in the zero-velocity limit $\nu \rightarrow 0$, since there is an intrinsically very large time scale separation between the driving and relaxation mechanism in the real earthquake process. This is easily implemented in the model by driving it to relaxation in a single step of the simulation: since the driving is homogeneous, the site with the largest stress at the completion of an avalanche will be the event initiator of the next. So, the new driving rule for the zero-velocity limit is to compute $E^* = \max E_i$ and then perform $E_i \rightarrow E_i + (E_c - E^*)$, $\forall i$. Care must be taken when working with the model, for its approach to a statistically stationary state proceeds rather slowly, and transients are very long. One has to wait typically for $\sim 10^9$ avalanches in a $2D$, $L \sim 10^2$ model before collecting meaningful data.

The OFC model with nearest neighbors has been extensively studied lately, and a plethora of information gathered about the nature of its (quasi-)critical attractor state. The signatures of this state become first visible near the borders, and it spreads through the lattice from there on. The model has a strong tendency to synchronization, which generates spatial correlations and is partially responsible for critical behavior. This behavior is lost when periodical boundary conditions are used, and synchronization forces the system to periodical non-critical behavior. The inhomogeneity induced by an open boundary is enough to destroy the periodic state while leaving intact correlations, thus allowing for the establishment of the (quasi-)critical attractor.

The reader should have noticed the awkwardness, or care if the reader wants to be nice to the authors, with which we

refer to the nature of the attractor. In fact, the issue of criticality in the OFC model is still a matter of intense debate among experts. There is general agreement on the fact that the model is indeed critical in $2D$ if $\alpha > \alpha_c$, but there is no such agreement on the critical value itself. Estimates for α_c vary widely in the literature, ranging all the way from $\alpha_c = 0$ [31], to $\alpha_c \simeq 0.18$ [32, 33], while the Brazilian expert C. Prado and collaborators developed strong arguments in favor of $\alpha_c = 0.25$, which corresponds to the conservative limit mentioned above, and propose the notion of a quasi-critical state to describe the nature of the attractor when α is close to α_c [34], to which we adhere. The reader is invited to participate in this discussion by trying to determine α_c through her or his own simulations of the model. On the other hand, there is no question about the fact that the OFC shares with a few other statistical models, such as the eight-vertices model, the unusual dependence of the Fisher critical exponent τ on the value of the conservation parameter α.

Another version of the OFC model that has been studied recently is the one with random annealed neighbors. Here, the extra noise induced by a random choice of neighbors at every relaxation, made anew each time a site topples, destroys the spatial correlations and avoids synchronization of the model. The borders are open, and each time a site relaxes, the dynamics chooses $q_i = 2d$ random neighbors with probability $p_{bulk} = (L-1)^d/L^d$, or $q_i = 2d-1$ with probability $p_{sup} = 1 - p_{bulk}$. After a first claim for an analytical solution of this mean-field-like model with $\alpha_c = 2/9$ [35], further investigation showed that it is only critical at the conservative limit [34].

Other flavors of CCA models have been studied that involve long-range interactions, in particular by the geophysics community [29, 36]. These versions are supposedly more "realistic", since viscoelastic interactions in the earth's crust are known to be long-range, presumably decaying as the third power of distance. Although some of the excitement involved in the emergence of long-range order from short-range interactions is lost in these models, they still present interesting features and deserve the attention of the physics community. The infinite-range interaction model, where each site interacts with all other sites in the lattice, is one of those, and its mean-field character has served well the purpose of establishing a test ground for the exploration of new ideas [37].

As mentioned above, CCA models have usually very long transients, forcing the researcher in the field to be very demanding on the efficiency of the computer code used in the simulations. Useful comments on this subject for the beginner can be found in Refs. [32] and [28]. As a rule of thumb, one has to try to deal only with one-dimensional data structures, and draw the code aiming at its efficient use. As an example, the address of a site in a 2-D lattice, usually taken as an ordered pair (i,j), should be transformed into a single integer as in $index = i * L + j$. The integer operations of division and remainder are then used to get the 2-D address from the 1-D version, when needed. The dynamical field is then a 1-D real vector, and loops sweeping

the lattice are controlled by a single integer variable. Another extremely useful strategy is to have moving failure and residual stresses: instead of sweeping the whole lattice each time the driving mechanism brings the site closest to failure - the initiator of the next avalanche - to its critical stress, the value of the latter is updated to the value of the stress of this initiator, and the residual stress is also moved to keep the difference between these two constant. The usage of last in-first out (LIFO) stacks is also good advice, in particular when running long-range models, for it allows a simple way of keeping track of only a few sites to examine for failure at each step of the relaxation avalanche. In the same meta-language used previously for the sandpile, here is a short description of an algorithm using these ideas for a 2-D model:

- site address structure: $(i,j) \rightarrow index = i * L + j$
- initialization: (e.g. $E_i = \text{random}(0, E_c) \ \forall i$)
- main loop: *do*
- selects site i with maximum E_i
- moves failure and residual stresses: $E_r \rightarrow E_r + E_c - E_i, E_c \rightarrow E_i$
- avalanche
- if time $>$ transient, collects statistical data
until time is up.

The key routine is the relaxation algorithm, the avalanche procedure. In this example, we use two separate stacks (1 and 2), where stack1 stores the sites that will topple in the present relaxation step, while stack2 keeps the sites that will topple in the next step.

- avalanche:
- set duration to one, $i \rightarrow$ stack1 (this is the initiator)
- while stack1 NOT empty *do*:
· unstacks from stack1;
· dumps stress on neighbors (v): if $E_v > E_c$, $v \rightarrow$ stack2;
· if stack1 is empty, but stack2 is not, exchanges stack1 and stack2, increments duration.

Codes written in plain C for several versions of earthquake CCA's can be obtained on request through the e-mail jssm@if.uff.br.

6   Bak-Sneppen-like Models

Slightly modifying an earlier model for surface growth [38], Per Bak and Kim Sneppen [39, 40] introduced their now-famous model for biological evolution, based on an extreme value dynamics. A population is linearly displayed around a circle, two neighbours per individual. The characteristic of each individual is measured by a single numerical value, interpreted here as its survival fitness. At beginning, all fitnesses are uniformly tossed at random between 0 and 1. The population evolves according to the following dynamic rule: 1) The individual with the smallest fitness is found; 2) The fitness values of this individual and its two neighbours are replaced by three new, freshly tossed random values, uniformly distributed between 0 and 1. The total number N of

individuals is supposed to be very large.

One time step consists in iteratively applying this rule N successive times (an N-cycle), where N is the same number which measures the population size. Thus, the time $t = 0$, 1, 2, 3 ... is a discrete variable. $1/N$-fractions of the time unit can also be measured, by considering incomplete N-cycles. Thus, for larger and larger values of N, one obtains the *continuous* time limit.

The band of distributed fitnesses starts in between 0 and 1, but shrinks as time goes by, Fig. 15. For large enough times, all fitnesses fall above a non-trivial critical value $F_c \approx 0.6670$ [41, 40], except for a null-measure set of recently replaced values, which fluctuates in avalanches.

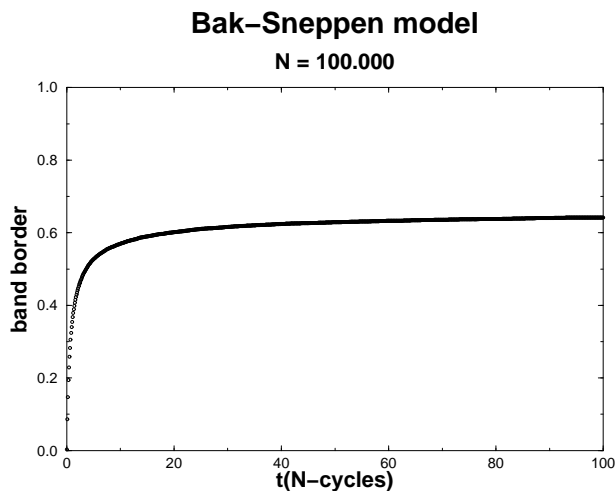## Bak–Sneppen model
N = 100.000

Figure 15. Bak-Sneppen model: lower bound for the band of fitnesses.

The first important feature of this model is its **long-term memory**. Instead of the fast exponential decay characteristic of a regular system, with a finite lifetime (short-term memory), here the evolution is very slow, following a power-law without any characteristic lifetime (long-term memory). This feature gives rise to many by-products which preserves forever the critical, complex behaviour of this system. For instance, the distribution of avalanche sizes is also dictated by a power-law, thus one lacks also a characteristic avalanche size (or lifetime).

This subject deserves a two-paragraphs comment. Regular exponential decays impose the appearance of a typical time scale for the system under study. The radioactivity lifetime characteristic of each element can be found in 100-years-old tables. For $_{37}\mathrm{Rb}^{82}$, for instance, one reads 2 minutes, without mention to the mass of the sample: after this time, radioactivity almost vanishes, independent of the sample size. For these regular systems, there is no relation at all between its size and its lifetime. The decay of a particular nucleus at a given moment has no influence at all on the later moment when another far nucleus will also decay. The absence of long-range spatial correlations (length) implies the corresponding absence of long-term memory (time).

Contrary to this, in non-regular systems, the long-term memory (time power-law decay) is normally associated with long-range spatial correlations (length power-law decays). Those complex systems lack both time and length typical scales. The lifetime is limited only by the finite size of the system. The larger the system, the longer it lives. This is the precise meaning of the sentence "For large enough times, all fitnesses fall above a non-trivial critical value $F_c = 0.6670$ ..." written three paragraphs above. The transient time T one needs to wait before this situation is of the same order of magnitude as the system size, i.e. $T \sim N$. In the limit of larger and larger sizes, $N \to \infty$, the system would evolve in eternal transient, a very profitable feature for evolutionary systems [42].

The second important feature is the **diversity**, preserved by the non-vanishing band-width, which also remains forever. This is the main difference between an evolutionary system and a simple optimisation process. In the latter, one searches for the single best situation among many possibilities, discarding all other options. Evolution, on the contrary, preserves many alternative options, not only a single "best", in order to keep the system able to adapt itself to future environment changes.

Another two-paragraphs comment follows, along the lines of [42]. Science is a modest enterprise, a theory of everything does not exist. The subject of any scientific work is necessarily bounded. For the evolution of species, nobody tries to describe the behaviour of all living beings in the Universe. On the contrary, the researcher considers a certain number of species, living on a limited geographic region, during a limited time. Thus the "system" under study is bounded, although not closed. Through its boundaries, this system interacts with the rest of the Universe, generically called "the environment", exchanging mass, energy, food, debris, information, etc. An otherwise closed system, according to the second law of thermodynamics, exponentially fast would reach the situation of maximum entropy, rendering impossible any kind of organisation such life: evolution would stop. In short, evolutionary systems are necessarily bounded but open.

The dynamic evolution of such a system is dissipative in what concerns its entropy: organised forms are selected, leading to the extinction of other forms. In other words, as time goes by, the space of possibilities shrinks. After a transient, the system eventually becomes trapped into a low-dimension tiny set of possibilities, the final attractor. There, lacking diversity, it looses the chance of exploring new forms outside this tiny set. Hence, the arrival to that final situation must be avoided, or postponed forever. This goal is accomplished if the dynamic evolution is very slow, i.e. a critical dynamics without any finite characteristic lifetime, an eternal transient in which the system approaches the attractor but never becomes completely trapped there. Not by coincidence, this is the case of Bak-Sneppen model.
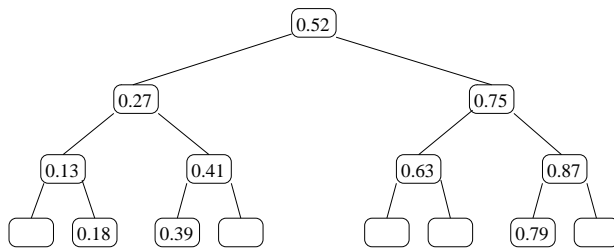
Figure 16. Binary tree where the sequence {0.52, 0.27, 0.41, 0.75, 0.13, 0.63, 0.39, 0.18, 0.87, 0.79} is stored in this order.

In order to implement this model on computers, the crucial point is to find the minimum fitness among all individuals. The naïve approach is to scan all of them, sequentially, registering at each step the minimum value so far. This requires N comparisons, and forbids the simulation of large systems. A good alternative is to construct a **binary tree** with the N fitnesses. This tree has a root, below it there are two other sites, on the left and on the right. Below each new site, another pair, on the left and on the right, and so on. See Fig. 16. The sequence of N fitnesses is stored on this tree as follows. The first entry is located at the root. If the second entry is larger than the first, it is stored on the right side below the root; otherwise, on its left. For each new entry along the sequence, one compares its value with the root, deciding to go downwards to left or right. Then, one repeats the comparison at this new place, deciding again to go downwards to left or right, and so on, until a vacant site is reached.

This construction turns trivial the job of getting the minimum value stored on the tree: one simply goes downwards always to left, until the last occupied site. The advantage over the naïve procedure of scanning all N entries is to follow only a single branch along the tree. On a binary tree, the average branch length is of the order of $\log_2(N)$, much smaller than N, moreover for large systems. The following C-language routine does the job.

```
unsigned minimum() {
/* finds the minimum on the binary tree */
 unsigned              i,j;
 j = root;
 do i = j; while(j=left[i]);
 return(i);
}
```

Another routine puts a new entry on the tree.

```
void put(new) unsigned new; {
/* includes a new entry into the binary tree */
 unsigned            i,j,lf,f;
 j = root; f = F[new];
 do {i = j;
  if(f<F[i]) {j = left[i]; lf = 1;}
  else {j = right[i]; lf = 0;}
 } while(j);
 top[new] = i;
 if(lf) left[i] = new; else right[i] = new;
}
```

Instead of using the pointer facilities of C-language, we explicitly define vectors `top[]`, `left[]` and `right[]`. This implies a coast in further computer memory, but makes clear what is done. Moreover, nowadays the memory available on most computers is far enough for our needs.

Other more efficient ways to implement the tree certainly exist, in particular by using recursivity and pointers. However, we prefer this old-fashioned way of programming, because the resulting code is much easier to understand. The interested reader can learn much more in [43], where efficient programming techniques are presented. Also in [44], the reader can find specific numerical approaches for many scientific problems. In order to deal with bitwise parallel operations, useful for Ising-like models, see [45].

A third routine designed to remove an entry from the tree deserves an explanation. In order to remove entry K from the tree, one considers entries L and R below it (left and right, respectively), as well as entry A below L (right). If L is empty, K is replaced by R, otherwise by L. Furthermore, if L, R and A are all three occupied (generating 2 right branches below L, instead of only 1), then A is transferred to the first empty position along the leftmost branch below R.

```
void remove(K) unsigned K; {
/* removes an entry from the binary tree */
 unsigned           t,L,R,A,i,j;
 t = top[K]; L = left[K]; R = right[K];
 top[K] = left[K] = right[K] = 0;
 A = right[L];
 if(t) {
  if(L) {
   if(K==left[t]) left[t] = L; else right[t] = L;
   top[L] = t;
   if(R) {top[R] = L; right[L] = R;
    if(A) { /* re-position of A */
     j = R;
     do i = j; while(j=left[i]);
     top[A] = i; left[i] = A;
    }
   }
  }
  else {
   top[R] = t;
   if(K==left[t]) left[t] = R; else right[t] = R;
  }
 }
 else {     /* remove root */
  if(L) {root = L; top[root] = 0;
   if(R) {top[R] = root; right[root] = R;
    if(A) {
     j = R;
     do i = j; while(j=left[i]);
     top[A] = i; left[i] = A;
    }
   }
  }
  else {root = R; top[root] = 0;}
 }
}
```

Finally, the main program to simulate the Bak-Sneppen model follows.

```
        /* population size */
#define      N      100000
        /* initial random integer seed */
#define     seed     4785671
        /* time limit */
#define      T      10000000
#include <stdio.h>
#include <math.h>
 unsigned            r,t,i,root,imin,neigL,neigR,f1,f2,f3,
                     mt,ft,tprint = T/1000,
                     F[N+1],top[N+1],left[N+1],right[N+1];
 double             sum,invN = 1.0/N,i2,invMAX = 1/4294967296.0;
 void               put(),remove();
 unsigned           minimum();

main() {
 r = seed|1; i2 = 2*invMAX/N;
 for(i=0; i<=N; i++) {r *= 16807; top[i] = left[i] = right[i] = 0;}
```

```
 imin = 1; F[imin] = 0;
 root = 2; r *= 16807; F[root] = r; sum = r*i2;
 for(i=3; i<=N; i++) {r *= 16807; F[i] = r; put(i); sum += r*i2;}
 printf("\n\n         time              band");
 for(t=0; t<=T; t++) {
  if((t%tprint)==0) printf("\n  %10.6lf    %10.6lf",t*invN,sum-1.0);
  sum -= F[imin]*i2; r *= 16807; f1 = F[imin] = r; sum += r*i2;
  neigL = (imin==1) ? N : imin-1; remove(neigL);
  sum -= F[neigL]*i2; r *= 16807; f2 = F[neigL] = r; sum += r*i2;
  neigR = (imin==N) ? 1 : imin+1; remove(neigR);
  sum -= F[neigR]*i2; r *= 16807; f3 = F[neigR] = r; sum += r*i2;
  ft = F[mt=minimum()];
  if(f1<f2) put(neigL); else {put(imin); imin = neigL; f1 = f2;}
  if(f1<f3) put(neigR); else {put(imin); imin = neigR; f1 = f3;}
  if(f1>ft) {remove(mt); put(imin); imin = mt;}
 }
 printf("\n\n");
}
```

Two comments concerning this program follow.

First, the random number generator we adopt is the simple multiplication with 16807, applied to 32-bit-integer odd numbers, starting from a given seed. Any alternative random number generator can be used. Instead of reducing these random odd integers to the interval between 0 and 1, we prefer to use directly the integer version, saving computer time. If one needs more than, say, 10^9 random numbers, then this simple generator does not work: repetitions of the same sequence start to appear again and again.

Second, after the transient time, when almost all fitnesses are already above the final band-border $F_c \approx 0.6670$, one can introduce a further smart trick, in order to save computer memory and (more important) time. It is very simple: to store only fitnesses below a certain limit a little bit larger than F_c, say, only values of $F < 0.668$ remain stored on the tree. Of course, in this case, one needs also to introduce a further array to keep the information concerning which individuals are currently on the tree. All others can be forgot. In the present case, however, we are interested just on the transient, thus this trick was not used. Nevertheless, its implementation is simple.

So far, nobody was able to present an analytical solution for this model, i.e. a closed mathematical form for the function displayed on Fig. 15. Even the threshold $F_c \approx 0.6670$ is not exactly known. On the other hand, there are at least two simplified versions with known analytical solutions. The Yee model [46, 47, 48] simply forgets the neighbours, replacing only the current smallest fitness. Fig. 17 shows the result, to be compared with Fig. 15. Diversity is lost, one does not remain with a finite band-width at the end, on the contrary all fitnesses are pushed towards the maximum possible value, $F_c = 1$. According to our previous interpretation, this is no longer an evolutionary system, only an optimisation process. However, the first important feature of Bak-Sneppen model remains: the approach to the limit

$F_c = 1$ is very slow, following a power-law. Criticality is not lost. It is very easy to determine the analytical form of the function displayed on Fig. 17. The original solution is presented is [47].
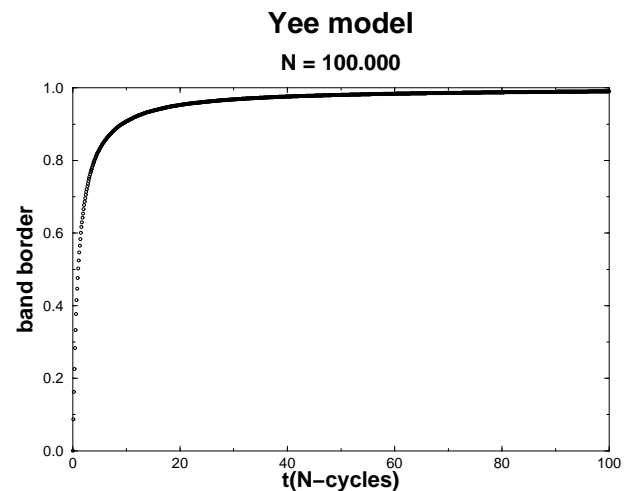

Figure 17. Yee model.

Here, we prefer a simpler reasoning based on the concept of a "virgin individual" at step τ: its fitness was never replaced, since $\tau = 0$. We use the symbol $\tau = 0, 1, 2, 3 \ldots$ to count the number of replacements performed so far. This counting variable is related with the time t through $\tau = Nt$. Let $x(\tau)$ be the minimum virgin fitness at step τ. Above $x(\tau)$, the distribution of fitnesses is uniform, because the minimum fitnesses replaced so far were never there, and all new fitnesses which appeared above $x(\tau)$ were randomly tossed. The function $x(\tau)$ evolves monotonically, always upwards, by successive plateaux. Let's call τ_n the

step when the nth virgin fitness becomes the global minimum, and then will be the next replaced value: at this moment, $x(\tau)$ will suffer a jump. The next smallest virgin value $x(\tau_n + 1) = x(\tau_{n+1})$, the next plateau, is a little bit above the previous one. The average jump is

$$\Delta x = 1/N \quad ,$$

the mean separation between the N original values at $\tau = 0$. However, in general, $x(\tau_n + 1)$ is not the global minimum fitness to be replaced at $\tau_n + 1$, because some other non-virgin values could remain below it.

How many? As the distribution above $x(\tau_n)$ is still uniform at τ_n, with all N fitnesses distributed in between $x(\tau_n)$ and 1, the number of fitnesses still below $x(\tau_n + 1)$ is $N[x(\tau_n+1)-x(\tau_n)]/[1-x(\tau_n)]$, i.e. $1/(1-x)$ on average. How many attempts one needs to transfer all these values to above $x(\tau_n + 1)$? For each attempt, the probability of success is $1 - x$, then, in average, one needs $1/(1-x)$ attempts to transfer each one of the $1/(1 - x)$ values. The total time interval to complete this task, in average, is

$$\Delta \tau = 1/(1 - x)^2 \quad .$$

Thus one can write the relation

$$\frac{\Delta x}{\Delta \tau} = \frac{1/N}{1/(1-x)^2} = \frac{(1-x)^2}{N} \quad .$$

Now, taking the continuous time limit and expressing x as a function of time $t = \tau/N$, this relation becomes the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (1 - x)^2 \quad .$$

Taking the initial condition $x(0) = 0$, the solution is

$$x(t) = \frac{t}{t + 1} \quad ,$$

which perfectly fits the plot on Fig. 17. Now, the power-law dependence becomes evident, asymptotically: $1 - x(t) \sim t^{-1}$ for $t \to \infty$, with the critical exponent -1.

Random neighbour BS model

N = 100.000

Figure 18. Bak-Sneppen random model.

Another simplification is the so-called random neighbours Bak-Sneppen model [49, 40]. Now, one always replaces the smallest fitness and also another fitness randomly tossed. The result is shown in Fig. 18. Diversity is restored: at the end, almost all fitnesses remain above the limit $F_c = 0.5$, defining a band-width. However, the long-term memory is lost: the decay to $F_c = 0.5$ follows now a fast exponential rate, no longer a power-law. Surprisingly, the lifetime distribution of avalanches is still described by a power-law, although following a classical, mean-field critical exponent of $-3/2$, different from the original Bak-Sneppen model [49, 40]. Somehow, criticality is not completely lost, in spite of the fast convergence to the final band-width. A related issue is the so-called $1/f$ noise claimed to be present within this model [50]. This would mean a lack of typical frequency scale within the power spectrum, equivalent to the lack of time or length scales characteristic of complex systems. However, this noise was later proposed to be in fact $1/f^2$ [51], which corresponds to the regular exponential decay for the correlations. This matter is not settled so far, being a good problem for future investigation, a challenge for the reader.

It is also possible to obtain an analytical expression for the function displayed on Fig. 18, originally presented in [49].

Once more, here we prefer our simpler reasoning based on virgin individuals. Let's divide each step τ in two substeps: first, the global minimum fitness is replaced, then another randomly tossed fitness is also replaced. The distribution of fitnesses above $x(\tau)$, the smallest virgin fitness so far, is also uniform as in the case of Yee model. At τ_n, $x(\tau)$ is also the global minimum. How many values remain virgin? For the simple Yee model, the answer would be $N[1 - x(\tau_n)]$, in average. However, now some formerly virgin values were also replaced during the above quoted second sub-steps performed up to now. Thus, the average number of still virgin values at step τ is $V(\tau) = N[1 - x(\tau)]f(\tau)$, where $f(\tau)$ is less than unity. The average x-jump is then

$$\Delta x = \frac{1 - x}{V} = \frac{1}{Nf} \quad .$$

Again due to uniformity, inside this interval Δx, the average number of fitnesses is $n(\tau_n) = N\Delta x/[1 - x(\tau_n)] = 1/\{f(\tau_n)[1 - x(\tau_n)]\}$ at step τ_n. This number will decay until vanishing at τ_{n+1}. Let's determine how it decays, by analysing a single step from τ to $\tau + 1$. On average, one has

$$n(\tau+1) = [n(\tau)-1](1-x)^2 + n(\tau)2x(1-x) + [n(\tau)+1]x^2 \quad .$$

The first term corresponds to get new values above x in both the first and the second sub-steps. The second term counts the cases where only one new value is above x, and the third
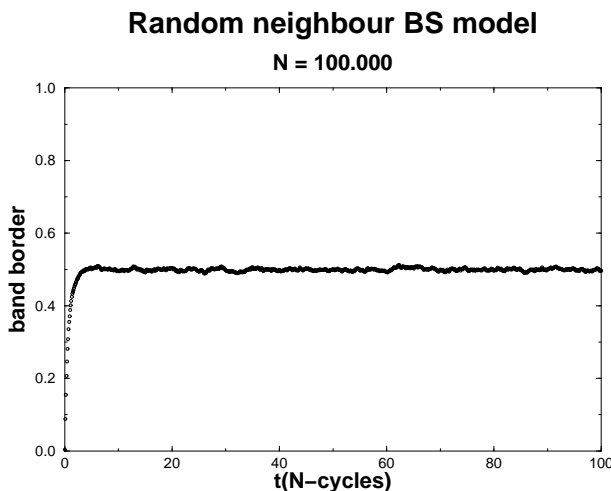
corresponds to none. Developing the above equation, one has $dn/d\tau = 2x - 1$, a constant rate decay. Thus, the total time one needs to wait until the complete exhaustion of the $1/[f(1-x)]$ fitnesses at τ_{n+1} is

$$\Delta\tau = \frac{1}{(1-x)(1-2x)f} \qquad \text{for} \ \ x < 1/2 \ \ .$$

For $x > 1/2$, the interval Δx will remain populated forever. Thus, as a first conclusion, different from the Yee model where the band of populated values always shrinks, all values accumulating near 1, now one remains forever with a uniformly populated band in between $F_c = 1/2$ and 1.

Finally, the evolution of $x(\tau)$, while below $F_c = 1/2$, is given by

$$\frac{\Delta x}{\Delta\tau} = \frac{(1-x)(1-2x)}{N} \quad ,$$

or, taking the continuous time limit,

$$\frac{dx}{dt} = (1-x)(1-2x) \ \ .$$

The solution for that differential equation, with initial condition $x(t) = 0$ is

$$x(t) = \frac{1 - e^{-t}}{2 - e^{-t}} \qquad \text{for} \ \ x < F_c = 1/2 \ \ ,$$

which fits perfectly the plot of Fig. 18. The asymptotic limit, $t \to \infty$, is now $F_c - x(t) \sim e^{-t}/4$, a fast exponential decay, no longer a slow power-law.

A good exercise to the reader is to solve the same model when, at each step, one replaces the minimum fitness and K other randomly chosen values.

Comparing these two simplified models, one can formulate the following interpretation. The extreme value dynamics, i.e. to replace the smallest fitness among all individuals, is the key ingredient to get criticality. However, this ingredient alone is not able to preserve diversity, other fitnesses besides the smallest one must be also continuously replaced. If these further replaced fitnesses are completely uncorrelated with the smallest one, then criticality is lost. In short, in order to get both criticality and diversity, one needs: 1) to replace the smallest fitness; 2) to replace also other fitnesses **somewhat correlated** to it. Within the original Bak-Sneppen recipe, this correlation is provided by the neighbourhood along the linear chain of individuals. The usual, naïve interpretation in favour of this particular choice is a food chain, considering each individual as a whole species: it gets food from its left neighbour, and provides food for its right neighbour. As soon as one species becomes extinct, its two neighbours also follow the same destiny, all three being replaced by new species. Because of the geometrical character of this rule, the analytical solution for the problem becomes hard. However, one can imagine many other non-geometrical origins for this kind of correlation we need.
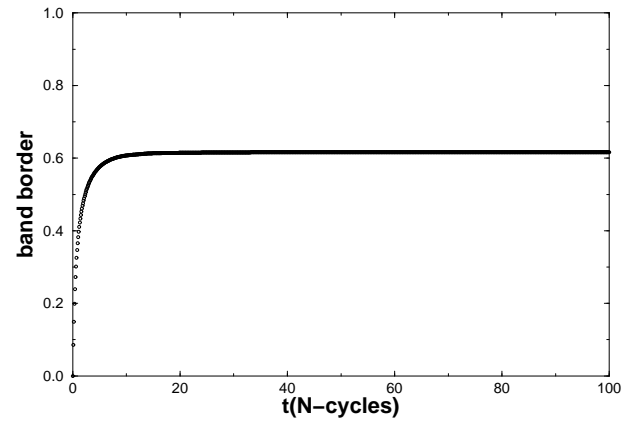

Figure 19. P-bag model.

One possible definition follows. Individuals are classified as active or quiet. At each step, the list of $P + 1$ active individuals have their fitnesses replaced by new, random values. This list (the bag) includes always the current minimum-fitness individual. At beginning, $P = 0$ with only the minimum-fitness individual being active. After replacements, if the same individual carries again the smallest fitness, then a new, random quiet individual is included into the list ($P \to P + 1$). Otherwise, a random active individual is excluded from the list, becoming quiet. In this case, the list could shrink ($P \to P - 1$) if the new minimum-fitness individual was already active before. P fluctuates always much smaller than N. The result for this dynamic rule is shown in Fig. 19, for a single run. The saturated final value as well as the time to reach it fluctuate for different runs. One can observe that diversity and criticality are present. In particular, the tail of the curve is well fitted by a power-law. Without geometrical constraints, perhaps somebody could provide an analytical formula for the function displayed in Fig. 19: so far, the not-so-smart authors were not able to perform this job, thus it is a challenge to the smart reader.

7   Conclusions

The above sections treat different situations where time and length scales were lost. They are examples of statistical models at criticality, as well as complex dynamic systems. The overall behaviour is the same in all cases, driven by the universality imposed by the quoted absense of characteristic scales.

It is not possible to treat such systems by classical approaches as perturbation theory and alike. Analytical results are rarely available, thus special attention to computational aspects were taken.

References

[1] M.E.J. Newman and R.M. Ziff, *Phys. Rev.* **E64**, 016706 (2001); *Phys. Rev. Lett.* **85**, 4104 (2000).

[2] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd edition, Taylor and Francis, London (1992).

[3] H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, Oxford (1971).

[4] K. Binder, *Z. Phys.* **B43**, 119 (1981).

[5] J.M.F. Neto, S. Moss de Oliveira and P.M.C. de Oliveira, *Physica* **A206**, 463 (1994); P.M.C. de Oliveira, *Europhys. Lett.* **20**, 621 (1992).

[6] P.M.C. de Oliveira, R.A. Nobrega and D. Stauffer, *Braz. J. Phys.* **33**, 616 (2003), also in cond-mat 0308525.

[7] P.M.C. de Oliveira, R.A. Nobrega and D. Stauffer, *J. Phys.* **A** (2004), also in cond-mat 0308566).

[8] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

[9] Z.W. Salzburg, J.D. Jacobson, W. Fickett and W.W. Wood, *J. Chem. Phys.* **30**, 65 (1959).

[10] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).

[11] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).

[12] B.A. Berg and T. Neuhaus, *Phys. Lett.* **B267**, 149 (1991); *Int. J. Mod. Phys.* **C3**, 1083 (1992).

[13] F.G. Wang and D.P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).

[14] P.M.C. de Oliveira, T.J.P. Penna and H.J. Herrmann, *Braz. J. Phys.* **26**, 677 (1996), also in cond-mat 9610041.

[15] P.M.C. de Oliveira, T.J.P. Penna and H.J. Herrmann, *Eur. Phys. J.* **B6**, 111 (1998), also in cond-mat 9807354.

[16] A.R. Lima, P.M.C. de Oliveira and T.J.P. Penna, *J. Stat. Phys.* **99**, 691 (2000).

[17] P.M.C. de Oliveira, *Comp. Phys. Comm.* **147**, 410 (2002), also in cond-mat 0204332.

[18] K.-C. Lee, *J. Phys.* **A23**, 2087 (1990).

[19] P.D. Beale, *Phys. Rev. Lett.* **76**, 78 (1996).

[20] P. Bak, C. Tang and K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987); *Phys. Rev.* **A38**, 364 (1988).

[21] V. Frette, K. Christensen, A. MaltheSorenssen, J. Feder, T. Jossang and P. Meakin, *Nature* **379**, 49 (1996).

[22] D. Dhar, *Phys. Rev. Lett.* **64**, 1613 (1990).

[23] D.L. Turcotte, *Rep. Prog. Phys.* **62**, 1377 (1999).

[24] H.J. Jensen, *Self-organized Criticality*, Cambridge University Press, Cambridge (1998).

[25] H. Gould and J. Tobochnik, *An Introduction to Computer Simulation Methods*, 2nd edition, Addison-Wesley (1996).

[26] R. Burridge and L. Knopoff, *Bull. Seism. Soc. Am.* **57**, 341 (1967).

[27] H. Nakanishi, *Phys. Rev.* **A43**, 6613 (1990).

[28] E.F. Preston, J.S. Sá Martins, J.B. Rundle, M. Anghel and W. Klein, *Comput. Sci. Eng.* **2**, 34 (2000).

[29] D.S. Fisher, K. Dahmen, S. Ramanathan and Y. Ben-Zion, *Phys. Rev. Lett.* **78**, 4885 (1997); K. Dahmen, D. Ertaş and Y. Ben-Zion, *Phys. Rev.* **E58**, 1494 (1998).

[30] Z. Olami, H.J.S. Feder and K. Christensen, *Phys. Rev. Lett.* **68**, 1244 (1992).

[31] A.A. Middleton and C. Tang, *Phys. Rev. Lett.* **74**, 742 (1995).

[32] P. Grassberger, *Phys. Rev.* **E49**, 2436 (1994).

[33] A. Corral, C.J. Perez, A. Diaz-Guilera and A. Arenas, *Phys. Rev. Lett.* **74**, 118 (1995).

[34] C.P.C. Prado and J.X. de Carvalho, *Phys. Rev. Lett.* 84, 4006 (2000); O. Kinouchi, S.T.R. Pinho and C.P.C. Prado, *Phys. Rev.* **E58**, 3997 (1998).

[35] S. Lise and H.J. Jensen, *Phys. Rev. Lett.* **76**, 2326 (1996).

[36] J.B. Rundle and D.D. Jackson, *Bull. Seism. Soc. Am.* **67**, 1363 (1977).

[37] J.S. Sá Martins, J.B. Rundle, M. Anghel and W. Klein, *Phys. Rev.* **E65**, 056117 (2002).

[38] K. Sneppen, *Phys. Rev. Lett.* **69**, 3539 (1992); K. Sneppen and M.H. Jansen, *Phys. Rev. Lett.* **70**, 3833 (1992); L.H. Tang and H. Leschhorn, *Phys. Rev. Lett.* **70**, 3822 (1993).

[39] P. Bak and K. Sneppen, *Phys. Rev. Lett.* **71**, 4083 (1993).

[40] P. Bak, *How Nature Works: the Science of Self-Organized Criticality*, Oxford University Press, Oxford/Melbourne/Tokyo (1997).

[41] M. Paczuski, S. Maslov and P. Bak, *Phys. Rev.* **E53**, 414 (1995).

[42] P.M.C. de Oliveira, xxx.lanl.gov cond-mat 0101170, short version in *Theory in Biosciences* **120**, 1 (2001); P.M.C. de Oliveira, *Physica* **A306**, 351 (2002), also in cond-mat 0108234.

[43] D.E. Knuth, *The Art of Computer Programming*, vols 1, 2 and 3, Addison-Wesley, Boston/London (1997).

[44] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C, the Art of Scientific Computing*, Cambridge University Press, Cambridge/New York (1988).

[45] P.M.C. de Oliveira, *Computing Boolean Statistical Models*, World Scientific, New York/London/Singapore (1991).

[46] K.K. Yee, xxx.lanl.gov nlin.A0/0106028 (2001).

[47] D. Stauffer and M.E.J. Newman, *Int. J. Mod. Phys.* **C12**, 1375 (2001).

[48] A. Aharony and D. Stauffer, *Int. J. Mod. Phys.* **C13**, 602 (2002).

[49] J. de Boer, B. Derrida, H. Flyvbjerg, H. Jackson and T. Wettig, *Phys. Rev. Lett.* **73**, 906 (1994).

[50] F. Daerden and C. Vanderzande, *Phys. Rev.* **E53**, 4723 (1996).

[51] J. Davidsen and N. Lüthje, *Phys. Rev.* **E63**, 063101 (2001).