Bak and Sneppen model
and
inhomogeneous branching processes

Thesis submitted for the degree of
Magister Philosophiae

Condensed Matter Sector

CANDIDATE: SUPERVISOR:
Claudio Tebaldi   Prof. A.Stella

OCTOBER 1995

Contents

1 Description of the Bak and Sneppen model.

Bak and Sneppen (BS) [1, 2] introduced perhaps the simplest model describing mutations of an ecology of interacting species. This model is intended to provide an illustration of the basic mechanism leading to intermittency and scale invariance in natural evolution. In a coarse grained simplified description, to the i-th species of the ecology a single fitness parameter, x_i, $(0 < x_i < 1)$ is associated. The fitness x_i represents the ability of species i to survive: the higher x_i, the higher the barrier to overcome in order to switch a mutation in the species. A genetic mutation changes the barrier of the species and modifies also the barriers of the other species interacting directly with it. This interaction should represent the fact that two species, e. g., take part in the same food chain. Sites of a lattice can be used to represent the species: in this case nearest neighbor (n. n.) sites can be assumed as directly related biologically, and thus interacting.

In the BS model the various mechanisms are governed by the following dynamical evolution rules. Starting from an initial fitness landscape, the i with lowest x, i_{min}, is selected and its fitness $x_{i_{min}}$ is modified (as a consequence of mutation) into a new one, chosen at random in the interval $(0, 1)$. Due to the interaction, also some neighbors of i_{min} undergo mutation and get modified x's. In the case of a linear chain with n. n. interactions this implies that, e. g., $x_{i_{min}-1}$ and $x_{i_{min}+1}$ are replaced by new randomly chosen x's.

This simple model shows interesting features that can give an explanation of similar observations obtained from the analysis of paleontological data such i.e.:

- Fitness probability distribution function has the shape shown in FIG.1; the system self-organizes in such a way that the probability of survival for species with fitness under a critical value is 0, while it is uniform over this value;

- When the system has arrived at the stationary condition (it is self-organized) distribution of activity is intermittent (FIG.2 shows intermittency patterns measured in an ecology consisting of 512 species);
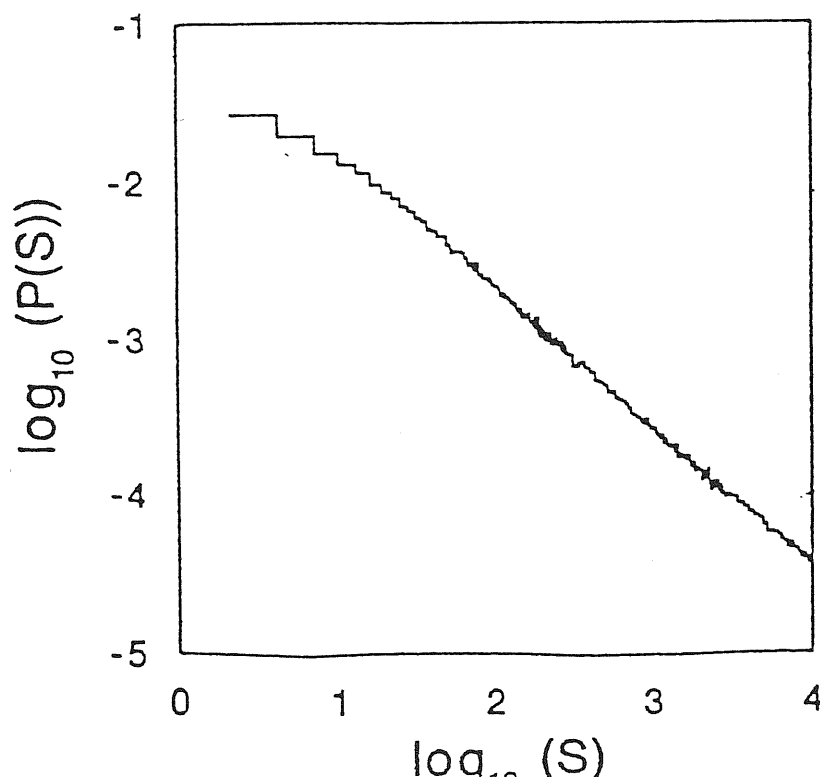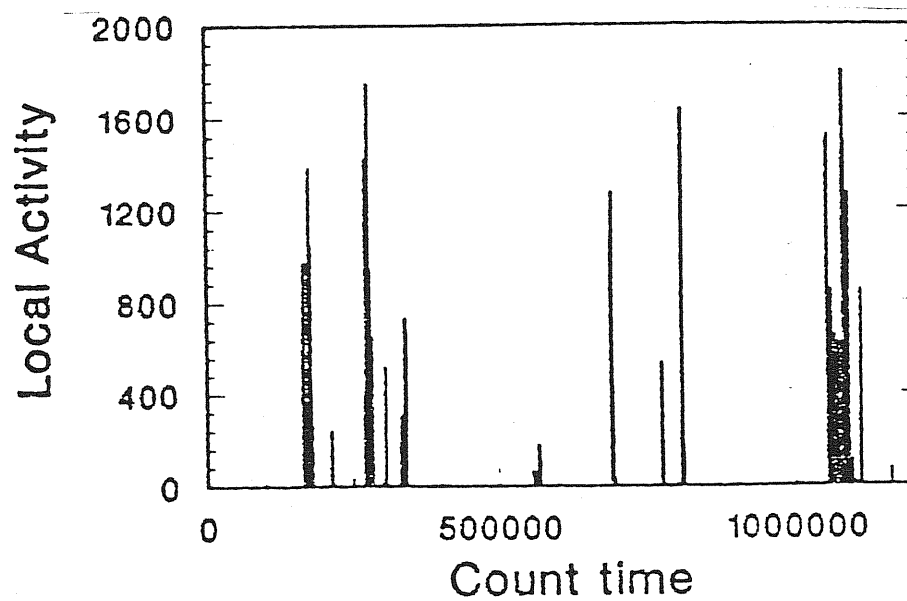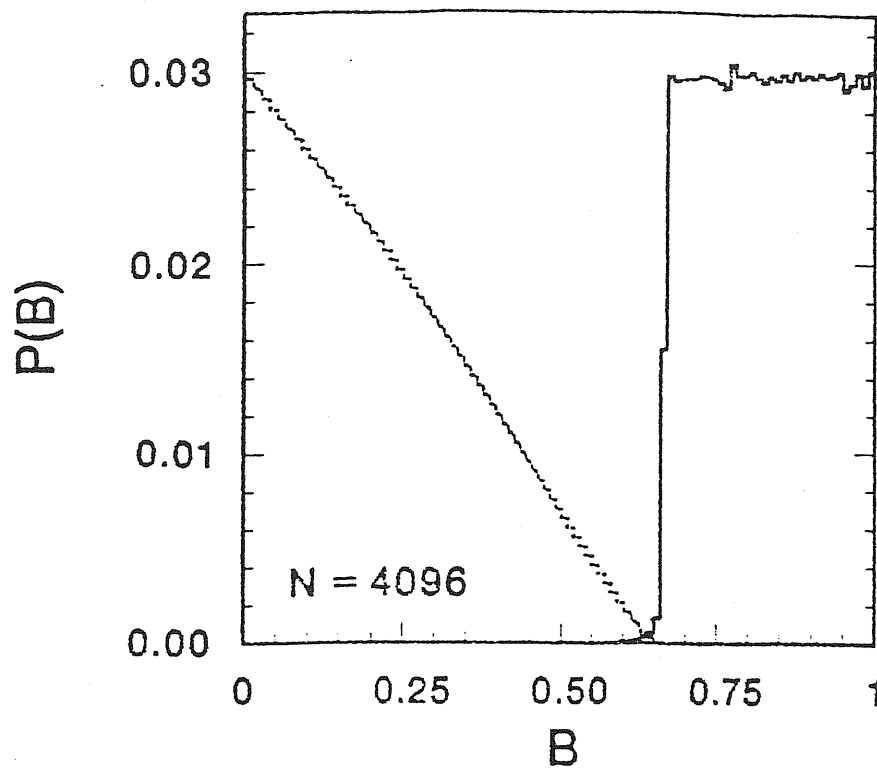
- Distribution of avalanche size is critical with power law behaviour as shown in FIG.3.

2 Biological reasons for considering Bak and Sneppen Model

The concept of punctuated equilibrium was introduced by Gouldberg and Eldredge [3][4] as a possible model for biological evolution. The main goal of this new explanation of evolution patterns is to explain efficiency of natural selection in adaptation of life to biosphere evolution. A simple Darwinian mechanism, in fact, would require, to explain observed morphological differences, an evolution time longer for some orders of magnitude with respect to the observations (340 mys instead of 3 mys).

The new model can be essentially characterized as follows:

- evolution and morphological adaptation concerns with species as spatio-temporally discrete and independent reproductive lineages while genetic mutation concerns with single organisms;

- evolution takes place through bursts that give evidence to "punctuation " in the evolutionary patterns, that is presence of long periods of stasis alternated with sudden, quantal morphological changes.

Bak and Sneppen model is asked to show in a simple model how interplay between interaction of species in the same ecology and genetic mutation can give rise to punctuated equilibrium as claimed by Gould and Eldredge.

The main idea is that if we can consider every species through a single value that represents the average probability for survival in the population called barrier, the "weakest " species (that for which adaptation is at a minimum value) will modify its adaptation barrier in a random way (mutation); this will generate random modifications also in the barriers of species "biologically" interacting with them; this will imply an avalanche like process in the evolution until a sort of stationarity is restored; distribution of evolution bursts becomes then critical; scale invariance of extinction events is in fact shown by paleontological data.

3

3 Theoretical motivations to investigate biological evolution models.

Numerical Evidence of Self-Organization in a large number of dynamical critical systems has been pointed out by Bak, Tang and Wiesenfeld [5]. In their papers it has been argued that a common underlying mechanism could relate criticality with stationarity in these spatio-temporal non-equilibrium systems; this relation would explain the presence of critical correlations in a large number of natural systems such sandpiles, biological equilibrium, markets etc.

The analytic approach to the description of these systems has shown some difficulties due essentially to the necessity of considering the close relationship between spatial and temporal correlations [7]. Most of the literature on this subject, in fact, considers alternatively only the temporal [6] or the spatial critical behaviour of the self-organizing systems (in the sandpile models a general conservation argument predicts critical spatial behaviour) and in general there's no uniformity in the analytic treatment of different models even at the mean field level.

Here we intend to suggest a common framework at the mean field level valid for all the self-organizing critical systems.

A common requirement in all S.O.C. systems is a local dynamics, that is: propagation of activity from site to site depends on the values of the site variables only in a local way; self-organization in these models consists on the fact that the bursts of activity due to a single, local excitation give raise to a long range, power law slowing down. The final goal of a S.O.C. theory is to describe how these bursts, so to speak, tune themselves toward a critical relaxation. The problem involves together spatio-temporal degrees of freedom in such a way that usual approaches fail (in general if one considers a Lagrangian approach to these problems non local, quite intractable terms appear). Bak et al. [5] argue that the right probabilistic description of bursts can be suggested from the following fact: the stationary state of these non-equilibrium systems can be characterized from a probabilistic point of view as the critical state of an equilibrium statistical system (the state of a system that is suffering a second order phase transition).

4

4 Mean Field and Branching Processes

Following [5] way of thinking we have to identify an equilibrium system that, at the critical point, describes propagation of activity from site to site. We will consider the system in an equilibrium mean field approximation; that is we neglect spatial fluctuations and we consider the propagation independent of the site where it happens. Moreover, since we want to describe stationary state properties we suppose that parameters describing propagation are time independent (we neglect temporal fluctuations: mean field in a dynamical sense).

These two approximations allow for defining a common phase space for such processes: the behaviour of a site is completely determined specifying the probabilities p_i for the number of directions of propagation when the site has already been activated. The problem stated in this way is from a mathematical point of view equivalent to a classical probabilistic topic [8][9] called branching process. In terms of random variables the problem can be stated in this form:

Let $X_1, ..., X_N$ N integer random variables equally distributed with a generating distribution function of the form:

$$Q\left(x\right) = \sum_{i=1}^{\infty} p_i x^i \tag{1}$$

and N itself a random variable; let us consider what happens if we generate a succession $N^{(n)}$ in the following way:

$$N^{(1)} = 1$$
$$N^{(n)} = X_1 + ... + X_{N^{(n-1)}} \tag{2}$$

Then we can interpret the random variable $N^{(n)}$ as the number of branches generated from a common ancestor if in every step each branch ramifies independently and the number of branches is the random variable X.

If we take the correspondence between sites activated and branches then we obtain some useful results:

* The branching process undergoes a phase transition:

The extinction of the activity is the event that starting from an arbitrary n $N^{(n)} = 0$; the total probability of extinction q is given as a solution of the fixed point equation:

5

$$q = Q(q) \tag{3}$$

this result is easily obtained observing that the probability of extinction at the n-th generation $q^{(n)}$ is given by the recursive formula:

$$q^{(1)} = p_0$$
$$q^{(n)} = Q\left(q^{(n-1)}\right) \tag{4}$$

Since for P is monotone, for $0 < p_0 < 1$ the succession of $q^{(n)}$ is growing and the limiting value is given by the solution of 3. We can discuss the solution observing that the graph of $Q(x)$ is a convex curve starting at the point $(0, p_0)$ and ending at the point $(1, 1)$ on the bisector. Only two situations are possible:

i. The graph is entirely above the bisector and the only solution of the equation 3 is given by $q = 1$; in this case it is easily to observe that $Q'(1) \leq 1$

ii. The graph intersects the bisector at some point $\sigma < 1$ so $q^{(n)} \to \sigma$ as $n \to \infty$ and $Q'(1) > 1$.

Let us observe that $A = Q'(1)$ is the expected number of directly activated sites; we obtain therefore that the extinction probability is 1 if $A \leq 1$; while it is σ if $A > 1$.

- The total number of sites activated has a critical distribution when $A = 1$: asymptotically the probability of having s sites activated in a single process has a power law distribution:

$$P(s) \sim s^{-\tau} \tag{5}$$

Let us define the random variable $Y_n = 1 + Z_1 + \cdots + Z_n$ it will equal the number of descendants up to generation n; the generating function for the distribution of Y_n that we will call $P_{(n)}(x)$ obeys a recursion formula of the form:

$$P_{(n)}(x) = xQ\left(P_{(n-1)}(x)\right) \tag{6}$$

The sequence of $P_{(n)}(x)$ decreases monotonically and is limited from below so it converges to a function $P(x)$ that is the root of the algebraic (Watson's equation):

$$t = xQ(t) \tag{7}$$

It can be proved [8] that the $P(x)$ is the generating function of the random variable total number of activated sites Y during a relaxation process.

Solving for t in 7 we obtain an explicit form for the probabilities $P(s)$ of having s descendants; the asymptotic estimate of these coefficients is given to be [9]:

$$P(s) \simeq A^s s^{-\frac{3}{2}} + higher orders \tag{8}$$

When $A = 1$ the process becomes critical and the moments of this distribution diverge.

Let us give an explicit example of how the calculation can be carried out: Let $Q(t)$ be the function :

$$Q(t) = p_0 + p_1 t + p_2 t^2 \tag{9}$$

The extinction probability for $A < 1$ is the solution of:

$$q = p_0 + p_1 q + p_2 q^2$$
$$q = \frac{1 - p_1 - \sqrt{(1-p_1)^2 - 4p_0 p_2}}{2p_2} \tag{10}$$

It is easy to note that when $A = p_1 + 2p_2 \to 1$, $q \to 1$.

Let us now calculate the generating probability distribution function $P(x)$:

$$P(x) = \frac{(1 - p_1 x) - \sqrt{(1 - p_1 x)^2 - 4x^2 p_0 p_2}}{2p_2 x} \tag{11}$$

Now the asymptotic estimate of the coefficients of the series expansion of $P(x)$ can be obtained observing that $P(x)$ has a singularity for $x \to 1^-$ such that expanding around 1 we obtain:

$$P(x) = 1 + a(1 - x)^{\frac{1}{2}} + a_1(1 - x) + \dots \tag{12}$$

and, as usually, we can relate the behaviour of $P(s)$ with the leading singularity [10] term:

$$P(x) = 1 + a(1 - x)^{\tau - 1} + \dots \tag{13}$$

so we obtain the estimate for $A = 1$, $P(s) \sim s^{-\frac{3}{2}}$.

7

5 Dynamical mean field.

We have thus obtained a minimal description of a critical process that can describe the stationary state of a Self Organized system in a mean field approximation. Obviously determination of the stationary state is only the first part of the task, since we have to identify very general physical conditions that give rise to a 'reasonable dynamics' attractive toward this critical condition.

Strictly speaking, since the branches are non-interacting, we have lost spatial correlations and this fact has important consequences when we analyze non-stationary behaviour because approximations on temporal correlations have to be consistent with this approximation (SMFA).

What "consistently" means is still an open problem that we want here briefly describe.

The most significant system in which the dynamical mean field problem has been treated analytically is the Bak and Sneppen model [6][2].

The main approximation in these treatments assumes that propagation of activity goes through a randomness annealed in the system in the sense that the neighbours to update are chosen in a random way between all the sites at every time step.

This approximation has the main advantage that in this way we can consider the probability distribution of the barriers independent of the site, this is obviously consistent with the SMFA

We can now define the quantity of interest as the probability distribution function of the barrier heights $p(x,t)$ and correspondingly the distribution of the smallest barrier:

$$p_1(x,t) = Np(x,t)Q^{N-1}(x)$$
$$Q(x) = \int_x^1 dx' p(x') \tag{14}$$

at the stationarity.

We simply quote the results from [2] where it is deduced from the master equation given by:

$$p(x,t+1) = p(x,t) - \frac{1}{N}p_1(x,t) - \frac{K-1}{N-1}\left(p(x,t) - \frac{1}{N}p_1(x,t)\right) + \frac{K}{N} \tag{15}$$

p(x) and p₁(x)

Barrier x

a)
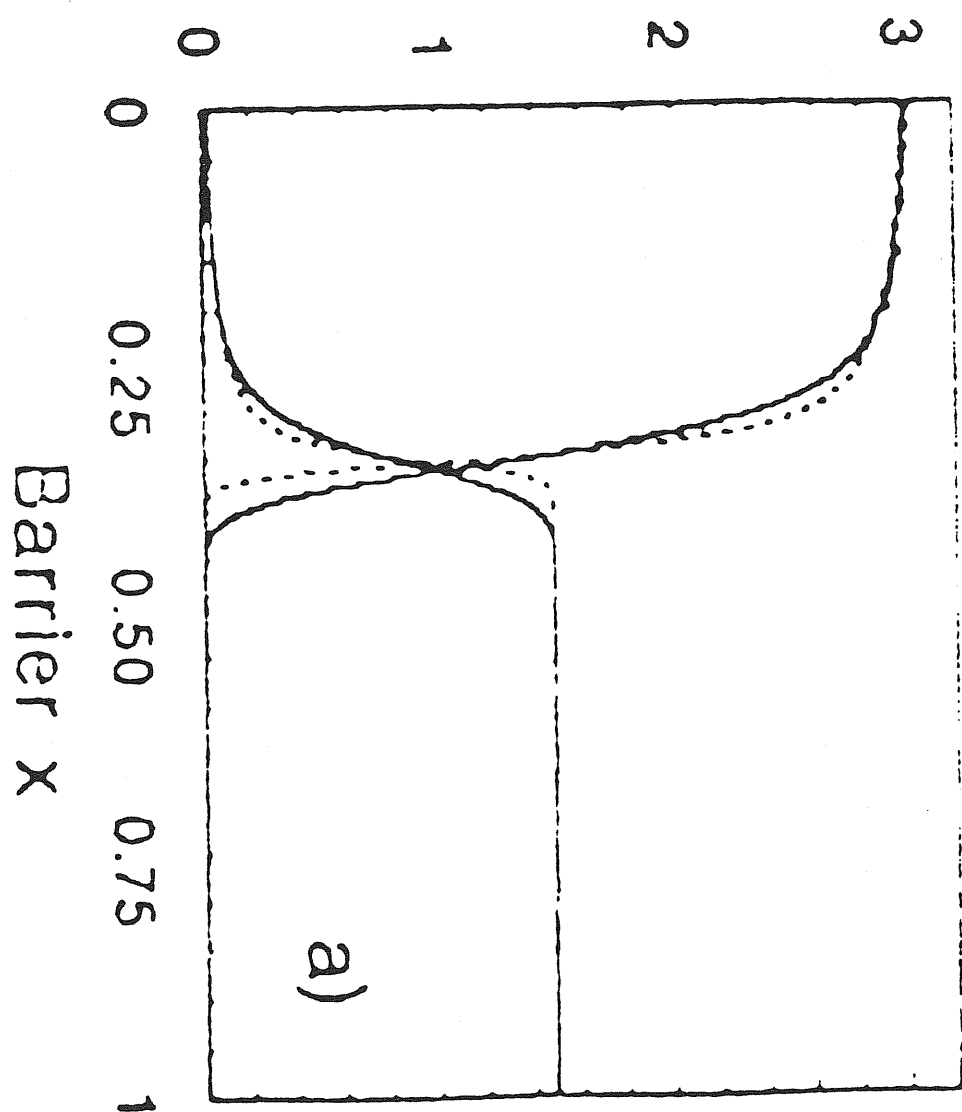
that the stationary distribution function $p(x)$ is of the form:

$$p(x) \simeq \frac{K}{N} \text{ for } 1/K - x >> O(1/N)$$
$$p(x) \simeq \frac{K}{K-1} \text{ for } 1/K - x << O(1/N) \qquad (16)$$

Increasing N (going toward an infinite system) $p(x)$ converges toward a step function with the step at $1/K$ .This implies that with probability 1 barriers are higher than $1/K$. This fact corresponds to the assumption that only a finite fraction (not growing with N) of sites will have at every time step height smaller than $1/K$.

In FIG.4 it is shown the equilibrium distribution barrier values $p(x)$ for simulated random neighbour model (full curves) and mean field values (dashed curves).

Moreover let us analyze what happens at every time step when the system is stationary: the probability of having a site with a barrier smaller than $1/K$ is 1, so we almost certainly take away a value smaller than $1/K$ and other $K-1$ values distributed as $p(x)$ and substitute them with K values chosen from a uniform distribution. Stationarity requires that $p(x)$ remains invariant under time evolution; this imposes that new values will be distributed as follows(as stated in [2]): the mean number of values chosen smaller or equal to $1/K$ will be 1 while $K-1$ values on the average will be chosen greater than $1/K$. We are now able to establish a connection between dynamical and static mean field: propagation of activity (updating of the site when the barrier is the lowest) will "usually" involve sites recently refreshed; we can make the assumption that, given K the number of sites updated at the last step, there exists a time independent probability p_i that i sites ($0 < i < K$) will take part to the same burst of activity (they will be updated before inert sites, roughly speaking are the sites with a new barrier lower than $1/K$; this fact is under current investigation).

Since the mean number of sites with height smaller than $1/K$ is exactly 1, this imposes the condition on the p_i that:

$$\sum_{i=1}^{K} p_i i = 1 \qquad (17)$$

but this is the condition of criticality for the SMFA. We have been able to show (at least from a physical point of view) how Bak and Sneppen dynamics provides an explanation of the link between stationarity and criticality; still

9

remains unsolved the problem of relating spatial criticality with mean field criticality, since spatial correlations are neglected in the SMFA.

6   Boundary effects at a critical point.

It is a well known fact that geometrical properties influence the universal behaviour of critical systems (critical exponents are different in different geometries).

It is a difficult task to describe this critical behaviour on particular geometries; there was a great debate if a mean field approach can be used to describe these critical properties, a result due to [?] shows as in an Ising model the boundary mean field exponent can be obtained in a seminfinite model and a diverging correlation length appears showing spatial critical properties. Our purpose is to show that in Self-Organized Critical systems presence of boundary effects (modification of the critical behaviour of $P(x)$) and some type of diverging correlation length appears due to the same dynamical reason that gives bulk criticality.

Let us consider now a semi-infinite sequence of species on a chain. To each species is associated an integer coordinate $k = 0, 1, 2, \ldots\ldots$ In a n.n. model the presence of the boundary clearly requires to allow for a k dependence of the avalanche probability distribution: some $P_k(s)$ or $\tilde{P}_k(z)$ will describe avalanches starting at site k along the chain. This situation can still be described within what we call inhomogeneous Branching Process. Since, as a consequence of a mutation at $k \geq 1$, at most 3 species can be further involved in the avalanche($K = 3$), probabilities $p_0, p_1, p_2,$ and p_3 will describe the possible outcomes of such a mutation. The model can be further specified assuming that with probability p_0, no further mutation takes place in the avalanche; with probabilities $p_1/3$ and $p_2/3$ the avalanche propagates, respectively, in any one and any two of the species in the set $k - 1, k, k + 1$; finally, p_3 is the probability that the avalanche involves all three species. Our assignments assume full equivalence between the central species, k, and the lateral ones, $k - 1$ and $k + 1$. This choice is only dictated by convenience and it is not strictly necessary for obtaining the final results. Notice that it is sensible to assume k-independence for the p_i's as long as $k \geq 1$. These probabilities, in a MF description, are naturally treated as local interaction parameters. Of course, there should be different probabilities for $k = 0$,

where the boundaries imposes i.e. $p'_3 = 0$.

A possible choice made below is to assign $p'_0 = p_0 + p_1/3$, $p'_1 = p_1/3 + p_2/3$ and $p'_2 = p_2/3 + p_3$ at $k = 0$, again implying equivalence of $k = 0$ and $k = 1$ with respect to single branch outcomes.

With the above assumptions, Watson's equation is replaced by a full hierarchy of equations:

$$\tilde{P}_0\left(z\right) = z\left(p'_0 + p'_1\left(\tilde{P}_0\left(z\right) + \tilde{P}_1\left(z\right)\right) + p'_2\left(\tilde{P}_0\left(z\right)\tilde{P}_1\left(z\right)\right)\right)$$

$$\cdots$$

$$\tilde{P}_k\left(z\right) = zG(p_0 + (\tfrac{p_1}{3}\tilde{P}_{k-1}\left(z\right) + \tilde{P}_k\left(z\right) + \tilde{P}_{k+1}\left(z\right))$$

$$+\tfrac{p_2}{3}\left(\tilde{P}_{k-1}\left(z\right)\tilde{P}_{k+1}\left(z\right) + \tilde{P}_k\left(z\right)\tilde{P}_{k+1}\left(z\right) + \tilde{P}_{k-1}\left(z\right)\tilde{P}_k\left(z\right)\right)$$

$$+p_3\left(\tilde{P}_{k-1}\left(z\right)\tilde{P}_k\left(z\right)\tilde{P}_{k+1}\left(z\right)\right) \tag{18}$$

The solution of eqs.(18) should converge to the bulk solution of eq.(7), for k approaching infinity. It is useful to adopt the following ansatz for the solution of eqs.(18):

$$\tilde{P}_k(z) = \tilde{P}(z) + \Delta(z)e^{-q(z)k} + l.s.t. \tag{19}$$

where q is an inverse length and $\tilde{P}$ is the solution of the eq.(7). As shown below, the assumed k-independence of Δ and q is consistent, as corrections to it would only involve subleading singular terms for $z \to 1^-$. By substituting eq.(19) into eqs.(18) one can deduce the singularity behaviour of $\tilde{P}_0$ and q. We are looking for such a solution that, for $z \to 1^-$, $\Delta(z) \simeq (1 - z)^\alpha$ and $q(z) \simeq (1 - z)^\beta$, with α and β suitable exponents. After substitution in eqs(18) for $k \geq 1$ one gets

$$1 = \frac{z}{3}(1 + 2\cosh q(z))\left(G'\left(P\left(z\right)\right) + \frac{\Delta\left(z\right)}{2}G''\left(P\left(z\right)\right)\right) + l.s.t.. \tag{20}$$

Taking into account that $\tilde{P}$ has the form (12) with $\tau = 3/2$, the leading singular terms in eq.(20) lead to:

$$\frac{q(z)^2}{3G''\left(1\right)} + \frac{1}{2}\Delta(z) \simeq a(1 - z)^{1/2}; \tag{21}$$

where $a = (2/G''\left(1\right))^{\frac{1}{2}}$. The same kind of substitution in the first of eqs.(18) leads to:

$$\Delta(z) \simeq a(1 - z)^{1/2} + b\Delta(z)q(z) + l.s.t. \tag{22}$$

11

with $b = \frac{p_1' + p_2'}{1 - 2(p_1' + p_2')}$ if only leading terms are taken into account.

Eq.(21) and eq.(22) are such to determine both α and β. In particular the singular behaviour of $\tilde{P}_0(x)$ takes the form

$$\tilde{P}_0(z) \simeq \tilde{P}(z) + \Delta(z) \simeq 1 + b(1 - z)^{3/4} \qquad (23)$$

According to eq.(13) this singularity is clearly consistent with

$$\tilde{P}_0(s) \simeq s^{-7/4}. \qquad (24)$$

Thus in present MF description the BS self organized critical state is characterized by a boundary scaling with an exponent different from the bulk one. Boundary avalanches of course suffer more rapid extinction and their distribution decreases more rapidly for large s. It is interesting to notice that $\tau_b = 7/4$ has been obtained recently within a MF approach to Abelian Sandpile Models (ASM) with Dirichelet boundary conditions, exploiting the analogies with magnetic systems [11].

Thus our result leads further support to the idea that also in ASM a BP description underlies the statistics of avalanches in the MF limit,for which also $\tau = 3/2$ holds.

A further consequence of eqs.(21),(22) is the singularity of q:

$$q(z) \simeq (1 - z)^{1/4}. \qquad (25)$$

Thus, the penetration length of the border disturbance, q^{-1}, diverges for $z \to 1^-$. In standard MF treatments of inhomogeneous equilibrium models, lengths of this kind show the same divergence with temperature as classical correlation lengths. Here it is not obvious how to associate an avalanche exponent to β, because z has no autonomous physical meaning for the avalanche problem. By interpreting z as a standard fugacity in a polymer, one should deduce from eq.(25) a correlation length diverging like $(1 - z)^{-1/4}$, which is indeed the classical ν exponent of branched polymers. Of course definition of a ν exponent for a SOC system requires to identify physically meaningful parameters describing approach or departure from criticality.

We have thus obtained meaningful results on the nature of the stationary critical state; we can summarize them as follows:

- The Mean Field Critical state predicts also some sort of spatial criticality;

Numerical data

$D(x)$

Barrier x

Numerical data
$f(s) = As^{-0.25} + C$

$P'_{sur}(s)$
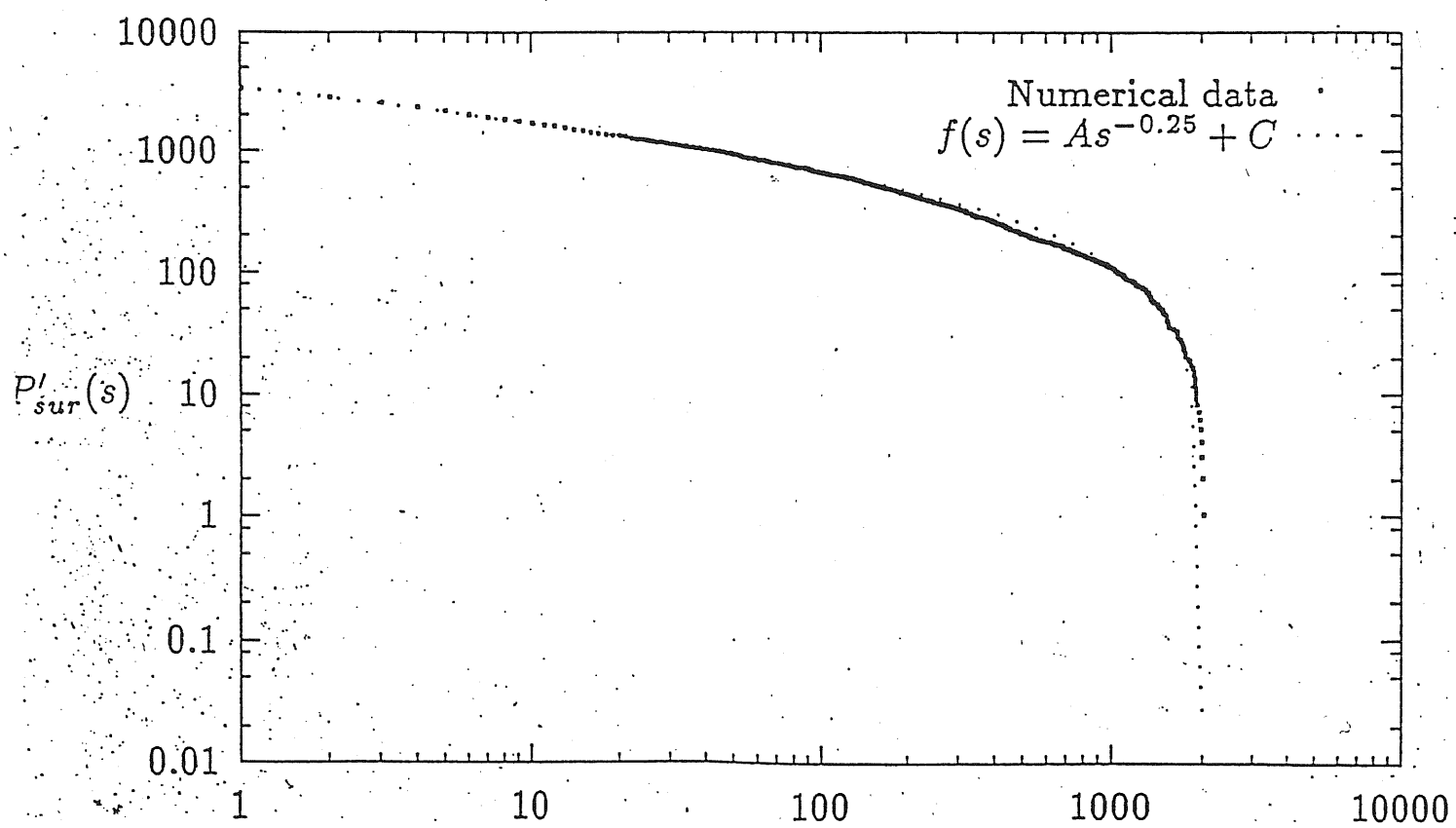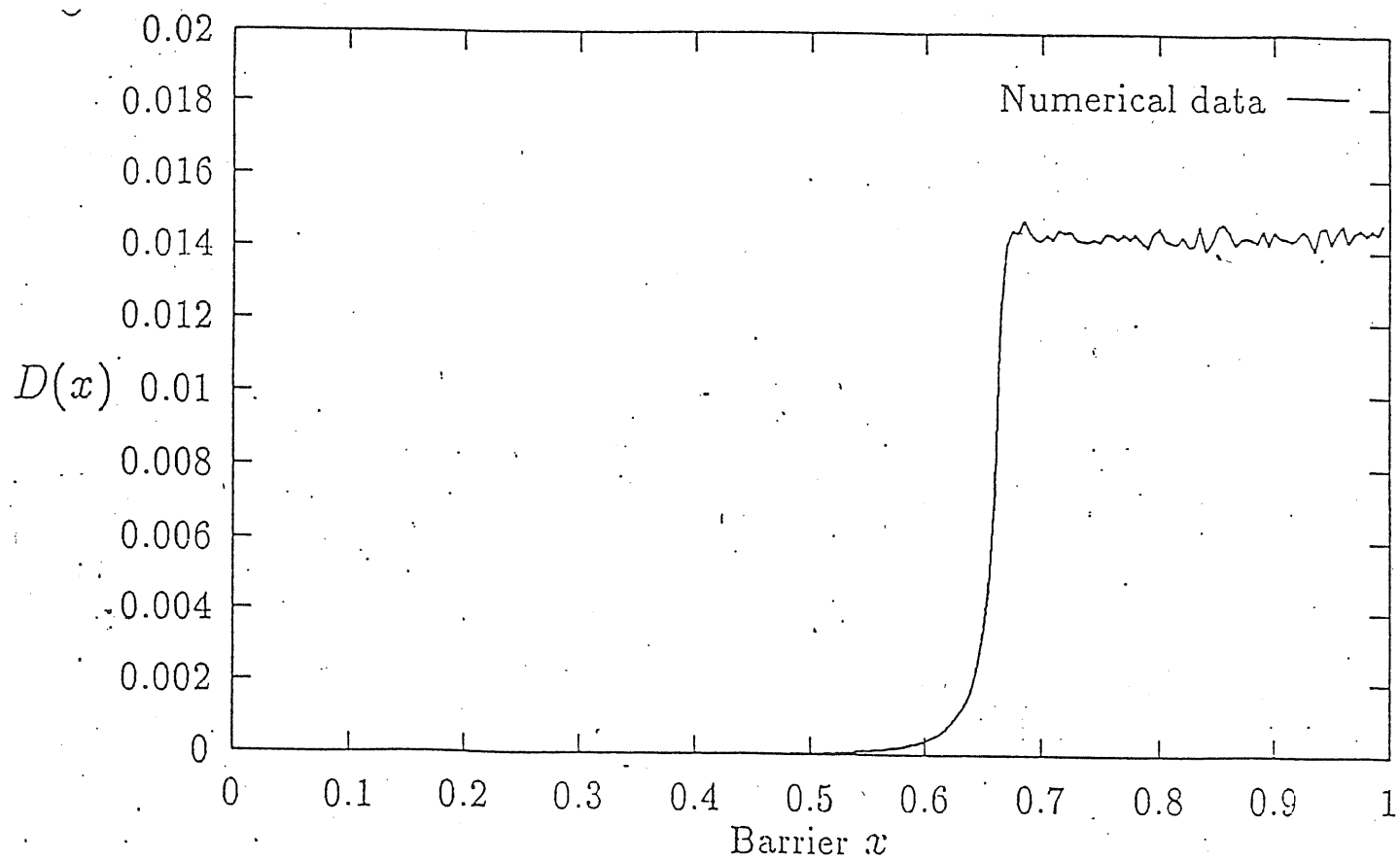
- Presence of a boundary modifies critical behaviour of the system.

We will see that the mean field prediction is qualitatively correct for the real Bak and Sneppen model.

7 Numerical investigations and experimental data.

To verify whether boundary scaling occurs beyond the MF domain, we also performed numerical simulations on open chains of different lengths ($N = 100, 500$ and 1000) with n.n. interactions. Even if the system dimensions are small compared with the works present in the literature, one has to consider that in this case we are interested in the avalanche starting from boundary. This means that only a vanishing small fraction of the starting processes has to be taken into account. The global properties of the model are not affected by the introduction of the two boundaries. Also in this case for the BS the distribution of barriers self organized itself over the critical value $B_c \simeq 0.67$ as indicated in Fig.5.

The avalanche instead displays a different behaviour with respect to the distance of the boundary. An avalanche is the ensemble of subsequent mutations through barrier below a certain threshold. The avalanche is characterized by its size s, that is the number of mutations under the threshold and the seed, that is the site of the first species that start the reaction.

The distribution of avalanches sizes displays the power law. If one considers the bulk sites, the exponent is $\tau = 1.05 \pm 0.02$ in agreement with the value obtained in the original definition of the model with periodic boundary conditions [1]. The addition of boundaries introduces a different behaviour as known for other model of SOC[11]. In fact, for sites belonging to the external surface we find a value of $\tau_{sur} = 1.25 \pm 0.02$. This behaviour is indicated in Fig.6, where, to avoid statistical noise, we have used the integrated distribution probability $P'_{sur}(s) = \int_0^{s_{max}} P_{sur}(s)ds$.

It has also to be noticed that the behaviour of the sites in the bulk of system is quite different and we find a value of $\tau = 1.05 \pm 0.01$. Furthermore different sizes of the system do not change remarkably this behaviour, but avoids the finite size effects in the bulk avalanches.

13

References

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

[2] H. Flyvbjerg, K. Sneppen and P. Bak, Phys. Rev. Lett. **71**, 4087 (1993).

[3] S. J. Gould, Eldredge, Nature 332 (1988)

[4] Paleobiology **3**, 135 (1977).

[5] P. Bak, C. Tang and K. Weisenfeld, Phys. Rev. Lett. **59**, 381, (1987).

[6] J. de Boer, B. Derrida, H. Flyvberg, A. D. Jackson, T. Wettig Phys. Rev. Lett. **73**, 906 (1994).

[7] J. de Boer, A. D. Jackson and T. Wettig Phys. Rev. E **51**, 1059 (1995)

[8] W. Feller An Introduction to Probability Theory and its Applications, Wiley & Sons, New York . (1971)

[9] T. Harris Theory of Branching Processes, Springer, Berlin (1963).

[10] H.Stanley; introduction to phase transitions and critical phenomena.

[11] A. Stella, C. Tebaldi and G. Caldarelli, Phys. Rev. E **52**, 72 (1995).