Model of correlated evolution

Michele Vendruscolo,[1,2] Paolo De Los Rios,[1,2] and Laura Bonesi[3]
[1]*Istituto Nazionale di Fisica della Materia (INFM), Via Beirut 2-4, 34014 Trieste, Italy*
[2]*International School for Advanced Studies (SISSA), Via Beirut 2-4, 34014 Trieste, Italy*
[3]*Durham University, Department of Biological Sciences, Durham DH1 3LE, England*
(Received 10 June 1996)

We introduce a model of biological evolution inspired by the Bak and Sneppen (BS) model. Microscopic dynamical rules are modified with respect to the BS model in order to account for predator-prey and competitor correlations. We perform numerical simulations of the system and compare them with both a mean field equation and a mean field simulation. The model is in a different universality class of self-organized critical behavior than the Bak-Sneppen model, and in addition shows a nontrivial fitness probability distribution. [S1063-651X(96)02012-0]

PACS number(s): 64.60.−i

I. INTRODUCTION

Much attention has recently focused upon nonequilibrium systems displaying self-organized criticality (SOC), a concept introduced by Bak, Tang, and Wiesenfeld [1]. SOC systems appear to be widespread in nature, including sandpiles [1], earthquakes [2], creep phenomena [3], material fracturing [4], fluid displacement in porous media [5], interface growth [6], and river networks [7].

Recently Bak and Sneppen (BS) [8,9] introduced a SOC model describing an ecosystem of interacting species evolving by mutation and selection, and capable of reproducing the punctuated equilibrium features [10] of evolution as inferred from observations of fossil records [11]. The signature of this phenomenon is the power law distribution of evolutionary avalanches, clearly showing that mutations (and extinction) may be episodic at all scales [10,11].

It has been debated if changing the microscopic dynamical rules in the sandpile model does or does not change the SOC universality class of the system [12,13]. It is therefore interesting to study the robustness of BS-type models when the interaction rules between species are changed.

In the BS model an ecosystem is represented by N interacting species. Each species is characterized only by its fitness. Such fitness is proportional to the average number of offsprings that an individual of the species may have in the given environment [14]. This definition also accounts for the greater resistance to mutations of the fitter species: mutations must propagate over a greater number of individuals to become a genetic trait of the species.

The species which is the most likely to mutate is the one with the lowest fitness value, because it is the one feeling the strongest evolutionary pressure (it evolves or becomes extinct). Genetic mutations, which occur at random, will change the fitness of the species at random as well. A change in the fitness of a species will in turn change the *local* environment for the species that are more dependent on it. As a consequence those species will also change their fitness (which is environment dependent).

It is therefore natural to distinguish between two different mechanisms through which the fitness of a species can change: a primary one due to spontaneous mutations, that involves the species with the lowest fitness, and a secondary (or induced) one, due to changes in the environment caused by the spontaneous mutation of another species. Without the offspring-related definition of fitness it would be difficult to understand why and how primary mutations induce secondary ones [15]. In our model such observations translate into microscopic dynamical rules, according to which secondary mutations are nontrivially correlated to primary ones.

The paper is organized as follows: in Sec. II the Bak-Sneppen model is reviewed, and a model with correlation among evolving species is introduced and motivated. In Sec. III results from numerical simulations of this model are presented and discussed, and a further model is introduced in order to check the robustness of its universality class; the random neighbor version of these models are also numerically analyzed. In Sec. IV a mean field solution of the three models (BS and the two correlated models) is proposed and compared with numerical simulations. In Sec. V conclusions are drawn.

II. MODEL

In the BS model species are arranged on a lattice, and interactions are among nearest neighbors. To each species is assigned a fitness represented by a random number x with uniform probability distribution in $[0,1]$. Evolution takes place in the following way: at each time step the species with the minimum fitness x_i is the most likely to mutate. Thus it is selected, and the value of its fitness is redrawn. Together with this, the fitnesses of its nearest neighbors also change at random with the same uniform distribution in $[0,1]$. The BS model does not introduce any correlation between primary and secondary mutations except causality: primary mutations induce secondary ones.

The SOC behavior of the system can be seen from the distribution $P(s)$ of the number s of causally connected mutation events, or avalanches. In one dimension (1D) it is found that $P(s)\sim s^{-\tau}$ [8], and the most accurate numerical estimate of the exponent is $\tau\simeq 1.073(3)$ [16]. The distribution $P(t)$ of the time t of first return (the interval between two successive mutations of the same species i) displays a power law behavior as well, $P(t)\sim t^{-\tau_{\mathrm{first}}}$, with $\tau_{\mathrm{first}}\simeq 1.58$. The same observation applies to the distribution of any return, with an exponent $\tau_{\mathrm{all}}\simeq 0.42$. An interesting feature of

the distribution $P(x)$ of the fitness in the BS model is that species with a fitness below a threshold value $x_c \simeq 0.667$ are likely to undergo a rapid extinction, that is, $P(x)=0$ for $x<x_c$ [8]. Above x_c, $P(x)$ is independent of x.

The BS model can be considered as a food chain where the mutation of a species randomly affects the fitness of its neighbors. It is legitimate to inquire about the motivations and effects of a different microscopic dynamical rule. We introduce a more biologically motivated correlation between species, assuming a predator-prey relation as given, e.g., by the well known Lotka-Volterra equations [17].

If a species increases its fitness, this will affect in different ways the *local* environment of those species that most interact with it. Species that feed on it will find a greater number of prey, and this means that they will have better chances to propagate their genes, and their fitness will be higher than before. Conversely, prey (or competitors in general), will find a fitter predator or competitor, thus finding a more hostile environment: their fitness will be lower than before. The opposite rule applies in the case the species with the lowest fitness value further weakens. Rules of this kind are often found in the biological literature [18].

Operationally this implies that if $x_i(t+1)<x_i(t)$, then $x_{i+1}(t+1)$ is extracted in $[0,x_{i+1}(t)]$ and $x_{i-1}(t+1)$ is extracted in $[x_{i-1}(t),1]$. The fitness of species $i+1$ which is *predator* over species i is decreased randomly and the fitness of species $i-1$ which is the *prey* of species i is increased randomly. In the opposite case, when $x_i(t+1)>x_i(t)$, then $x_{i+1}(t+1)$ is extracted in $[x_{i+1}(t),1]$ and $x_{i-1}(t+1)$ is extracted in $[0,x_{i-1}(t)]$.

III. NUMERICAL SIMULATIONS

We perform simulations for a 1D system with the rules described in Sec. II. The correlated model introduced in this paper still shows the presence of a threshold in the distribution $P(x)$. A major difference with respect to the original BS model is that above such a threshold $P(x)$ is not a constant. From mean field calculations, which we will discuss in the next section, Sec. IV, we expect an algebraic behavior for $P(x)$ for $x \to 1$,

$$P(x)=A(1-x)^{-\alpha}, \tag{1}$$

where, from the normalization condition, $A=(1-\alpha)/2(1-x_c)^{1-\alpha}$. The fitness probability distribution obtained numerically, shown in Fig. 1, is well represented by a power law in the entire interval $[x_c,1]$. We estimate $\alpha=0.51(1)$ and $x_c=0.75(1)$, [note that $P(x_c)\sim 2$]. This distribution has a very simple meaning, and implies that the greater the probability to find a species, the higher is its fitness. From a scaling analysis of our numerical data, we obtain the avalanche exponent $\tau=1.04(2)$ (see the inset in Fig. 2).

For the distribution $P_{\text{first}}(t)$ of the first return time and for the distribution $P_{\text{all}}(t)$ of any return time we obtain, respectively, $\tau_{\text{first}}=1.40(2)$ as shown in Fig. 2, and $\tau_{\text{all}}=0.60(1)$. We note that the relation $\tau_{\text{first}}+\tau_{\text{all}}=2$ holds within the error as for the BS model [19].

The correlated model introduced in this paper belongs to a different universality class from the BS model. We enquire about its robustness by introducing a second model (model B, whereas the previous model will be referred to as model A in what follows), with a modified microscopic rule. The primary mutation induces secondary mutations of the nearest and next to the nearest species in such a way that if it increases its fitness then the two *predators*, that is, the two species on the right side, will increase their fitness too; the two species on the left will instead have an alternate behavior, the nearest one decreasing its fitness, the next nearest one increasing it. In the opposite case, when species i decreases its fitness, species $i+1$ and $i+2$ decrease their fitness as well, whereas species $i-1$ increases its fitness and species $i-2$ decreases it. As odd as it may seem, this rule is just the simplest extension of the above-mentioned predator-prey relation.

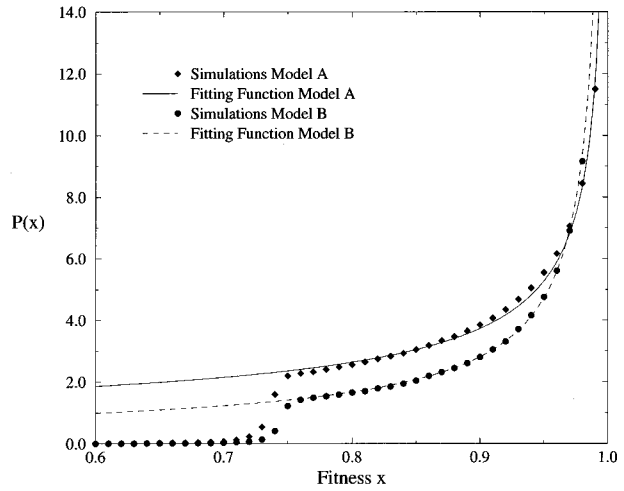

FIG. 1. Fitness probability distributions for model A (diamonds) and model B (circles); symbols correspond to results from simulations (1D nearest neighbor model), the fitting functions are proportional to $(1-x)^{-0.5}$ (model A, solid line) and to $(1-x)^{-0.75}$ (model B, dashed line).

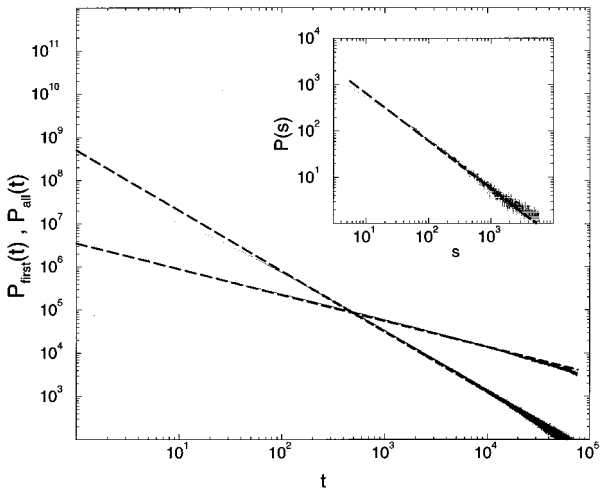

FIG. 2. Return time probability distributions for model A. We numerically obtain $\tau_{\text{first}}=1.40(2)$ (lower curve) and $\tau_{\text{all}}=0.60(1)$ (upper curve). The avalanche probability distribution $P(s)$ is shown in the inset ($\tau=1.04\pm 0.02$).

Simulations of the evolution of this second model show again that the fitness distribution exhibits a threshold, and that above it $P(x)$ increases as $(1-x)^{-\alpha}$ with $\alpha=0.748(3)$ (Fig. 1). Again the behavior of $P(x)$ above the threshold seems to be a power law, obeying law (1). The critical exponents τ, τ_{first}, and τ_{all} are unchanged within the error. We can thus conclude that the two models belong to the same universality class [which is determined from the avalanche and return time exponents, and not from $P(x)$, which is much more like an order parameter].

In order to check the previous results, we perform numerical simulations of the random neighbor version of both models; that is, the species that undergo secondary mutations are chosen at random on the left and right sides of the one affected by a primary mutation. The results are qualitatively in agreement with those from the nearest neighbor models: the fitness distribution shows both a threshold and a divergence for $x\to 1$, and the divergence exponents agree within the error with those from simulations for the nearest neighbor model. The avalanche and return time exponents are all 1.50(1) for both models, in agreement with the random neighbor version of the BS model and with the theoretical predictions for it, $\tau=\tau_{\text{first}}=\tau_{\text{all}}=\frac{3}{2}$ [9,20]. Actually this does not come as a surprise, since the random neighbor version of the models can be formulated in terms of branching processes, where the exponents $\frac{3}{2}$ come independently out of the details of the microscopic rules [21].

We thus identify a *mean field* universality class to which the random neighbor versions of the BS and correlated models all belong. The *mean field* predictions for $P(x)$ are quite in agreement with the real behavior of the system, as they should be if we interpret the fitness distribution as some sort of order parameter.

Actually, if this is the case, we expect the mean field approximation to describe the system far from the critical point better, which in this case is the fitness threshold.

IV. MEAN FIELD SOLUTION

We give a mean field analysis of the models in order to check the divergence of $P(x)$ for x close to 1. Following [22], we write a master equation for the problem.

$$P(\mathbf{x};t+1)=\sum_{i=1}^{N}\int_0^1 d\mathbf{x}' P_a(i;\mathbf{x}')T(i;\mathbf{x}',\mathbf{x})P(\mathbf{x}';t), \quad (2)$$

where $P(\mathbf{x};t)$ is the probability of the fitness configuration $\mathbf{x}=\{x_1,\ldots,x_N\}$ at time t; $P_a(i;\mathbf{x})$ is the probability that site i is active given the configuration $\mathbf{x}$,

$$P_a(i;\mathbf{x})=\prod_{j\neq i}\theta(x_j-x_i); \quad (3)$$

and $T(i;\mathbf{x}',\mathbf{x})$ is the transition probability from configuration $\mathbf{x}'$ to configuration $\mathbf{x}$ if i is the active site.

Bak and Sneppen model: only three sites change their fitness, taking it at random from a uniform probability distribution between 0 and 1. Then the expression for $T(i;\mathbf{x}',\mathbf{x})$ is

$$T(i;\mathbf{x}',\mathbf{x})=\prod_{j\neq i,i\pm1}\delta(x'_j-x_j). \quad (4)$$

The mean field approximation turns out to be just an ansatz for the form of $P(\mathbf{x})$ at the stationariety (no more time dependence):

$$P(\mathbf{x})=\prod_{i=1}^{N}p(x_i). \quad (5)$$

Substituting Eq. (5) into Eq. (2), integrating over $N-1$ of the N possible x_i variables between 0 and 1, and integrating over the last variable between x and 1 we obtain an equation which is the same as the one presented in [9]:

$$\left(1-\frac{2}{N-1}\right)Q^N(x)+\frac{2N}{N-1}Q(x)+3(x-1)=0, \quad (6)$$

where $Q(x)=\int_x^1 p(x')dx'$. The solution of this equation is [9]

$$P(x)=\begin{cases}0, & x<\frac{1}{3}\\ \frac{3}{2}, & x>\frac{1}{3}\end{cases} \quad (7)$$

In the limit $x\to 1$, we try a solution of the form

$$p(x)\sim A(1-x)^{-\alpha}. \quad (8)$$

Cancellation of the leading terms on the left-hand side of Eq. (6) yields $\alpha=0$ and $A=\frac{3}{2}$, consistent with Eq. (7).

Starting from Eq. (2), we can derive a mean field solution for our models changing the transition probability $T(i;\mathbf{x}',\mathbf{x})$.

Model A: three sites are involved in the change, the active one and its two nearest neighbors, as in the BS model. If i is the active site, then site $i+1$ changes in the same direction, and site $i-1$ in the opposite one. The corresponding transition probability can be written as

$$\begin{aligned}T(i;\mathbf{x}',\mathbf{x})=&\left[\prod_{j\neq i,i\pm1}\delta(x'_j-x_j)\right]\Bigg[\theta(x_i-x'_i)\frac{1}{1-x'_{i+1}}\\ &\times\theta(x_{i+1}-x'_{i+1})\frac{1}{x'_{i-1}}\theta(x'_{i-1}-x_{i-1})\\ &+\theta(x'_i-x_i)\frac{1}{x'_{i+1}}\theta(x'_{i+1}-x_{i+1})\\ &\times\frac{1}{1-x'_{i-1}}\theta(x_{i-1}-x'_{i-1})\Bigg]. \quad (9)\end{aligned}$$

The mean field ansatz is again Eq. (5); proceeding through the same steps as before we obtain the following mean field equation:

$$\begin{aligned}&\left(1-\frac{2}{N-1}\right)Q^N(x)+\frac{2N}{N-1}Q(x)+(x-1)\\ &-\frac{N}{N-1}\int_0^1 p(x')\left[\frac{2x'-x}{x'}\theta(x'-x)+\frac{1-x}{1-x'}\theta(x-x')\right]\\ &\times[1-Q^{N-1}(x')]dx'=0. \quad (10)\end{aligned}$$

We try a solution of the form (8). After some simple algebra, cancellation of the leading terms on the left-hand side of Eq. (10) gives $\alpha=\frac{1}{2}$, in agreement with both true and mean field simulations.

Model B: this model involves the simultaneous change of five sites, the i site and its nearest and next to the nearest neighbors according to the above-mentioned rules.

The transition probability reads

$$T(i;\mathbf{x}',\mathbf{x})=\left[\prod_{j\neq i,i\pm1,i\pm2}\delta(x'_j-x_j)\right]\left[\theta(x_i-x'_i)\frac{1}{1-x'_{i+2}}\theta(x_{i+2}-x'_{i+2})\frac{1}{1-x'_{i+1}}\theta(x_{i+1}-x'_{i+1})\right.$$
$$\times\frac{1}{x'_{i-1}}\theta(x'_{i-1}-x_{i-1})\frac{1}{1-x'_{i-2}}\theta(x_{i-2}-x'_{i-2})+\theta(x'_i-x_i)\frac{1}{x'_{i+2}}\theta(x'_{i+2}-x_{i+2})\frac{1}{x'_{i+1}}\theta(x'_{i+1}-x_{i+1})$$
$$\left.\times\frac{1}{1-x'_{i-1}}\theta(x_{i-1}-x'_{i-1})\frac{1}{x'_{i-2}}\theta(x'_{i-2}-x_{i-2})\right]. \tag{11}$$

The mean field approximation leads now to the equation

$$\left(1-\frac{4}{N-1}\right)Q^N(x)+\frac{4N}{N-1}Q(x)+(x-1)-\frac{N}{N-1}\int_0^1 p(x')\left[\frac{4x'-x}{x'}\theta(x'-x)+3\frac{1-x}{1-x'}\theta(x-x')\right][1-Q^{N-1}(x')]dx'$$
$$+\frac{2N}{N-1}\int_0^1 p(x')\left[\int_0^{x'}Q^{N-1}(x'')dx''-x'Q^{N-1}(x')\right]\left[\frac{x}{x'}\theta(x'-x)+\frac{1-x}{1-x'}\theta(x-x')\right]=0. \tag{12}$$

In the limit $x\to1$, using Eq. (8), we obtain $\alpha=\frac{3}{4}$, again in agreement with both kinds of simulations.

We could further generalize the model, allowing N sites to be involved in the mutation event, with a simple extension of the rule that led from model A to model B. In that case we obtain $\alpha=(N-2)/(N-1)$. The theoretical limit, as is clearly visible, is $\alpha\to1^-$ when $N\to\infty$, which ensures the integrability of the fitness probability distribution. Since the value of α depends on the number of species involved in the change, we argue that it is also strongly dependent on the dimensions of the system.

The mean field approach we used is very powerful in the determination of $P(x)$, at least far from the threshold x_c, as we could expect from a mean field solution far from the critical point.

V. CONCLUSIONS

The recent introduction [8] of a very simple model able to reproduce some of the features that seem to be important to understand the evolution of life, such as *punctuated equilibrium* [10,11], stimulated research aimed at both better understanding the properties of the model itself [20] and introducing more realistic features [23]. We adopted a definition of fitness that allows us to introduce a *predator-prey* correlation among the species, which we believe is an essential ingredient in the description of an ecosystem. We presented numerical simulations of two related models based on this dynamics and were able to describe their behavior using a mean field approach.

The analysis of these models clearly shows that they are still SOC systems, and that they belong to a different universality class than the Bak-Sneppen model. This is an important result, in fact, the BS model turns out to be sensitive to changes of the microscopic dynamical rules. As has been recently observed, this is not a feature characteristic only of SOC models based on extremal dynamics, but also of sandpile models [13].

Also the fitness probability distribution $P(x)$ is nontrivially affected by the introduction of correlations among the species. There are no species with a fitness below a certain threshold (just as in the BS model), and above it there is a higher probability of finding species with high fitness values.

Our model can still be considered a branching process, as it is signaled by its mean field exponents. Moreover the law $\tau_{\text{first}}+\tau_{\text{all}}=2$ is still respected.

ACKNOWLEDGMENTS

It a pleasure to thank Amos Maritan for valuable suggestions. This work was supported by Consiglio Nazionale delle Ricerche under Contract No. 95.01056.CT12.

[1] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987); Phys. Rev. A **38**, 364 (1988).
[2] J. M. Carlson and J. S. Langer, Phys. Rev. Lett. **62**, 2632 (1989).
[3] S. I. Zaitsev, Physica A **189**, 411 (1992)
[4] A. Petri, G. Paparo, A Vespignani, A. Alippi, and M. Costantini, Phys. Rev. Lett. **73**, 3423 (1994); P. Diodati, F. Marchesoni, and S. Piazza, *ibid.* **67**, 2239 (1991).
[5] D. Wilkinson and J. F. Willemsen, J. Phys. A **16**, 3365 (1983); M. Cieplak and M. O. Robbins Phys. Rev. Lett. **60**, 2042 (1988)
[6] K. Sneppen, Phys. Rev. Lett. **69**, 3539 (1992); K. Sneppen, *ibid.* **71**, 101 (1993).
[7] A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, E. Ijjasz-Vasquez, and R. L. Bras, Phys. Rev. Lett. **70**, 822 (1993); A. Maritan, F. Colaiori, A. Flammini, M. Cieplak, and J. R. Banavar, Science **272**, 984 (1996); G. Caldarelli, A. Giacometti, A. Maritan, I. Rodriguez-Iturbe, and A. Rinaldo (unpublished).
[8] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).
[9] H. Flyvbjerg, K. Sneppen, and P. Bak, Phys. Rev. Lett. **71**, 4087 (1993).
[10] S. J. Gould, Paleobiology **3**, 135 (1977); N. Eldredge and S. J. Gould, Nature (London) **332**, 211 (1988).
[11] M. D. Raup, Science **251**, 1530 (1986).
[12] L. Pietronero, A. Vespignani, and S. Zapperi, Phys. Rev. Lett. **72**, 1690 (1994).
[13] A. Ben-Hur and O. Biham, Phys. Rev. E **53**, R1317 (1996)
[14] L. Peliti (unpublished).
[15] H. F. Chau, L. Mak, and P. K. Kwok, Physica A **215**, 431 (1995).
[16] P. Grassberger, Phys. Lett. A **200**, 277 (1995)
[17] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Springer Series in Synergetics Vol. 13 (Springer, New York, 1985).
[18] See, for example, M. Higashi and H. Nakajima, Math. Biosci. **130**, 99 (1995), and references therein.
[19] S. Maslov, M. Paczuski, and P. Bak, Phys. Rev. Lett. **73**, 2162 (1994).
[20] J. de Boer, B. Derrida, H. Flyvberg, A. D. Jackson, and T. Wettig, Phys. Rev. Lett. **73**, 906 (1994); J. de Boer, A. D. Jackson, and T. Wettig, Phys. Rev. E **51**, 1059 (1995).
[21] G. Caldarelli, C. Tebaldi, and A. L. Stella, Phys. Rev. Lett. **76**, 4983 (1996).
[22] L. Anton (unpublished); A. Maritan (private communication).
[23] M. E. J. Newman and B. W. Roberts, Proc. R. Soc. London Ser. B **260**, 31 (1995).