A Simple Model of Self-organized Biological Evolution

Jan de Boer[1], Bernard Derrida[2,3,4], Henrik Flyvbjerg[2,5], Andrew D. Jackson[1], and Tilo Wettig[1]

[1] *Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3800, USA*

[2] *The Isaac Newton Institute for Mathematical Sciences, 20 Clarkson Road, Cambridge CB4 0EH, U. K.*

[3] *Laboratoire de Physique Statistique, ENS, 24 rue Lhomond, F-75005 Paris, France*

[4] *Service de Physique Théorique, CE-Saclay, F-91191 Gif-sur-Yvette, France*

[5] CONNECT, *The Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark*

(Dated: June 28, 2009)

Abstract

We give an exact solution of a recently proposed self-organized critical model of biological evolution. We show that the model has a power law distribution of durations of co-evolutionary "avalanches" with a mean field exponent 3/2. We also calculate analytically the finite size effects which cut off this power law at times of the order of the system size.

Introduction. Recently, a simple dynamical model for Darwinian evolution on its slowest time scale was introduced by Bak and Sneppen [1, 2]. The model describes an ecosystem of interacting species which evolve by mutation and natural selection. The model is abstract and focusses on only a few important aspects of evolution. It should be thought of as a *coarse grained* description of biological evolution, i.e. a description on the largest time scale. It provides a possible explanation for the characteristic intermittency of actual evolution, called *punctuated equilibrium* by Gould and Eldredge [3, 4], and the apparent scale invariance of extinction events described by Raup *et al.* [5–7]. Its principal idea is that, if life on Earth is a *self-organized critical* dynamical system [8] then intermittency and scale invariance are universal, hence robust consequences that do not depend on details of its dynamics, and so would be present also in much more complicated systems.

The models described in [1, 2] retain the salient features of species evolving by adaptive walks in rugged fitness landscapes and interacting by affecting the shape of each others landscapes, as proposed by Kauffman [9] and analysed in [10, 11]. Even at this level of abstraction details are ignored, however. Thus the state of an ecosystem of N species is characterized simply by N real numbers (x_i), $i = 1, 2, \ldots, N$. The model is completely specified by the following dynamical rule: at each time step, the x_i with minimal value as well as $K - 1$ others chosen at random, are replaced by K new random numbers.

The value of x_i characterizes the effective barrier towards further evolution experienced by the ith species while it exists at a local fitness maximum. The dynamics consists in selecting the species with the lowest barrier value—it is the first to evolve—and replacing that value, and those of $K-1$ other species, with new values. For simplicity, the new values are assumed random, all drawn from the same uniform distribution in the interval $[0, 1]$. The specification of *which* $K - 1$ other species are affected by change in a given species, defines the interactions between species. Here, as in [2] and [11], we assume that the $K - 1$ other species are a random selection among the $N - 1$ other species in the ecology. We assume this randomness is *annealed:* the $K - 1$ species affected by change in a given species are chosen anew every time it changes. This assumption facilitates calculations and does not seem less realistic than other choices.

In the present paper we analyse the model for its mathematical consequences, with little mention of their biological interpretation, using random walk techniques. We first treat the simplest case of $K = 2$ in some detail, then briefly discuss the extension to the general case towards the end. We calculate the distribution of the duration of avalanches for an infinite system and find that the system has the mean field exponent found in [2]. We also calculate the finite size effects which appear as a cutoff in the distribution of avalanche life times.

Master equations. A simple quantity one can consider for this system is the number n of variables x_i which have values less than a fixed value λ. Let $P_n(t)$ denote the probability that this is the case at time t. ¿From the definition of the model one can then write the following master equation for P_n:

$$P_n(t+1) = \sum_{m=0}^{N} M_{n,m} P_m(t) \tag{1}$$

where the matrix $M_{n,m}$ for $n \geq 1$ is given by

$$\begin{aligned} M_{n+1,n} &= \lambda^2 - \lambda^2(n-1)/(N-1) \\ M_{n,n} &= 2\lambda(1-\lambda) + (3\lambda^2 - 2\lambda)(n-1)/(N-1) \\ M_{n-1,n} &= (1-\lambda)^2 + (-3\lambda^2 + 4\lambda - 1)(n-1)/(N-1) \\ M_{n-2,n} &= (1-\lambda)^2(n-1)/(N-1), \end{aligned} \tag{2}$$

and for $n = 0$:

$$M_{0,0} = (1-\lambda)^2 \ ; \quad M_{1,0} = 2\lambda(1-\lambda) \ ; \quad M_{2,0} = \lambda^2. \tag{3}$$

We note from Eq. (2) that if $P_n(0) = 0$ for $n > N$, this property remains true at any later time.

One defines a λ-*avalanche* as the evolution taking place between two successive times where the number n vanishes. Thus if one lets $Q_n(t)$ denote the probability of having n numbers x_i less than λ, given that the avalanche started t time steps ago, $Q_n(t)$ satisfies the same master equation as $P_n(t)$ does, but with $M_{0,n}$ replaced by 0. The probability $q(t)$ of avalanches having duration t is then the probability that an avalanche terminates at time t,

$$q(t) = (1-\lambda)^2 \left(Q_1(t-1) + \frac{1}{N-1} Q_2(t-1) \right), \tag{4}$$

assuming it began at time $t = 0$. Figure 1 shows exact values of $q(t)$ obtained by iterating the master equation for several choices of N. One sees that $q(t)$ is a power law $t^{-3/2}$ at early times with an N-dependent cutoff at late times. As $N \to \infty$ only the power law is seen.

Several other quantities could be obtained from the knowledge of P_n or Q_n. For example, one can calculate the probability distribution for the n'th smallest value $x^{(n)}$ of the N

variables (x_i) in the steady state,

$$\text{Probability}(x^{(n)} = \lambda) = -\frac{\partial}{\partial\lambda}\sum_{m=0}^{n-1} P_m(t). \tag{5}$$

One could also calculate from the $Q_n(t)$ the maximum value of n reached during an avalanche and the total number of different variables involved in an avalanche.

Case of $N = \infty$. The limit $N \to \infty$ makes the analytical approach much easier. We first treat the case of an infinite system and then discuss how the limit $N \to \infty$ is approached. If $N \to \infty$, n being kept fixed, Eqs. (1–3) read

$$\begin{aligned} P_0(t+1) &= (1-\lambda)^2 \left[P_0(t) + P_1(t)\right], && (6)\\ P_1(t+1) &= 2\lambda(1-\lambda)\left[P_0(t) + P_1(t)\right] \\ &\quad +(1-\lambda)^2 P_2(t), && (7)\\ P_2(t+1) &= \lambda^2 \left[P_0(t) + P_1(t)\right] \\ &\quad +2\lambda(1-\lambda)P_2(t) + (1-\lambda)^2 P_3(t), && (8) \end{aligned}$$

and for $n \geq 3$

$$\begin{aligned} &P_n(t+1) = \\ &\qquad \lambda^2 P_{n-1}(t) + 2\lambda(1-\lambda)P_n(t) + (1-\lambda)^2 P_{n+1}(t). \end{aligned} \tag{9}$$

These equations describe a biased random walk with a reflecting boundary a $n = 0$. As $t \to \infty$, $P_n(t)$ evolves to the time-independent solution to this equation. When $\lambda < 1/2$ this solution is a geometric series for $n \geq 2$,

$$\begin{aligned} P_0 &= 1 - 2\lambda && (10)\\ P_1 &= (1-2\lambda)((1-\lambda)^{-2} - 1) && (11)\\ P_n &= (1-2\lambda)\lambda^{2n-2}(1-\lambda)^{-2n}. && (12) \end{aligned}$$

We see that the assumption $n = \mathcal{O}(1)$ is satisfied where P_n is not exponentially small, provided λ remains fixed at a value less than 1/2 in the limit $N \to \infty$. As $\lambda \to 1/2$, all the $P_n \to 0$, meaning that the probability that n remains of order 1 vanishes. The scaling limit $\lambda \to 1/2$ and $N \to \infty$ is discussed below. For $\lambda > 1/2$, Eqs. (6–9) predict that $P_n \to 0$ as $t \to \infty$, and this is because the distribution in the steady state is peaked around $n = (2\lambda - 1)N$.

For $\lambda < 1/2$, one can also calculate $q(t)$ by using the biased random walk picture. An avalanche started at time $t = 0$ has $P_0(0) = 1$, hence has initial condition $Q_1(1) = 2\lambda(1-\lambda)$, $Q_2(1) = \lambda^2$, $Q_n(1) = 0$ for $n \geq 3$. Using the method of images [12, p. 236], one finds

$$Q_n(t) = \frac{2n(2t+1)!\ \lambda^{t+n-1}(1-\lambda)^{t-n+1}}{(t+n+1)!\ (t-n+1)!} \tag{13}$$

Using Eq. (4) with $N = \infty$, one gets for the probability that an avalanche terminates at time t:

$$q(t) = \frac{(2t)!}{(t+1)!\ t!} \lambda^{t-1}(1-\lambda)^{t+1} . \tag{14}$$

It is easy to calculate the generating function of $q(t)$,

$$\sum_{t=1}^{\infty} x^t q(t) = \frac{1 - 2x\lambda(1-\lambda) - [1 - 4x\lambda(1-\lambda)]^{1/2}}{2\lambda^2 x}. \tag{15}$$

One can then check that $q(t)$ is normalised and that the average value of t diverges as $\lambda \to 1/2$,

$$\langle t \rangle = \sum_{t=1}^{\infty} t q(t) = (1 - 2\lambda)^{-1}. \tag{16}$$

For large values of t, the asymptotic form of Eq. (14) is

$$q(t) \simeq \frac{(1-\lambda)\,[4\lambda(1-\lambda)]^t}{t^{3/2}\ \lambda\sqrt{\pi}}, \tag{17}$$

and when $\lambda \to 1/2$ this gives a power law with the usual mean field exponent 3/2 of self-organized criticality [2].

Scaling limit $N \to \infty$ and $\lambda \to 1/2$. As $\lambda \to 1/2$, finite size effects start to become important. We consider now the scaling limit

$$N \to \infty \quad \text{with} \quad \lambda = \frac{1}{2} + \frac{\alpha}{\sqrt{N}}. \tag{18}$$

In this limit, P_n becomes a function of $n/\sqrt{N}$

$$P_n = \frac{1}{\sqrt{N}} f(\frac{n}{\sqrt{N}}) \tag{19}$$

and by requiring that Eq. (19) is a steady state solution of Eq. (1), one finds for large N that f should satisfy

$$\frac{1}{4}\frac{d^2}{dx^2} f + (x - 2\alpha)\frac{d}{dx} f + f = 0 \tag{20}$$

which leads to the following expression for the steady state probabilities P_n in the scaling regime (18):

$$f(x) = e^{-2(x-2\alpha)^2} \left(\int_0^\infty dy e^{-2(y-2\alpha)^2} \right)^{-1} . \tag{21}$$

In the scaling regime (18), the distribution of durations of avalanches can also be calculated. The distribution $Q_n(t)$ defined as above (see paragraph between Eq. (4) and Eq. (5)) has the scaling form

$$Q_n(t) = \frac{1}{\sqrt{N}} g(\frac{n}{\sqrt{N}}, \frac{t}{N}). \tag{22}$$

$Q_n(t)$ satisfies the same equation as $P_n(t)$ except at the boundary $n = 0$. Consequently, one can substitute the scaling form of $Q_n(t)$ in the master equation, and find that the function $g(x, \tau)$ satisfies

$$\frac{\partial g}{\partial \tau} = g + (x - 2\alpha) \frac{\partial g}{\partial x} + \frac{1}{4} \frac{\partial^2 g}{\partial x^2}. \tag{23}$$

We shall only discuss the critical case $\alpha = 0$ here, because this makes the solution of Eq. (23) easier. In the limit $\tau \to 0$ the solution must coincide with the expression (13) in the limit $1 \ll n \ll t$, i.e. with

$$g(x, \tau) = \frac{4x}{\sqrt{\pi} \tau^{3/2} \sqrt{N}} \exp(-x^2/\tau) \tag{24}$$

With this initial condition, the solution of Eq. (23) is

$$g(x, \tau) = \frac{4x}{\sqrt{\pi}\sqrt{N}} \left(\frac{2}{1 - e^{-2\tau}} \right)^{3/2} e^{-\tau} e^{-2x^2/(1-e^{-2\tau})}. \tag{25}$$

Using Eq. (4), one then finds

$$q(t) = \frac{\sqrt{8}}{\sqrt{\pi} N^{3/2}} \; \frac{e^{2t/N}}{(e^{2t/N} - 1)^{3/2}}. \tag{26}$$

For $t/N \ll 1$ this result is identical to Eq. (17) at $\lambda = 1/2$, as it should be. Figure 2 shows the same results as Figure 1, but in a scaling form, i.e. $t^{3/2} q(t)$ versus t/N. As N increases, the results agree better and better with the scaling form (26). The average duration of an avalanche at the critical point follows also from Eq. (26),

$$\langle t \rangle = \sum_t q(t) \; t \simeq \sqrt{2\pi N}. \tag{27}$$

Case of generic K-value. Let us now discuss briefly the model for general K. The matrix elements $M_{n,m}$ in Eq. (1) become for $m \geq 1$

$$M_{n,m} = \sum_{k=1}^{N} B_{n-m+k;K} H_{k-1;K-1;m-1;N-1}, \tag{28}$$

and for $m = 0$

$$M_{n,0} = B_{n;K}. \tag{29}$$

Here

$$B_{k;K} = \begin{pmatrix} K \\ k \end{pmatrix} \lambda^k (1-\lambda)^{K-k} \tag{30}$$

is the binomial distribution and

$$H_{k-1;K-1;m-1;N-1} = \begin{pmatrix} m-1 \\ k-1 \end{pmatrix} \begin{pmatrix} N-m \\ K-k \end{pmatrix} \Big/ \begin{pmatrix} N-1 \\ K-1 \end{pmatrix}. \tag{31}$$

is the hypergeometric distribution [13, 26.1.21]. In Eq. (28) the hypergeometric distribution gives the probability that $K-1$ numbers randomly chosen among $N-1$ yield $k-1$ with values less than λ when there are $m-1$ such numbers among the $N-1$. The binomial distribution gives the probability that out of K numbers assigned equiprobable random values between 0 and 1, $n-m+k$ numbers are given values less than λ. Each value of k denotes a different way that m values less than λ may change into n such values in a single time step. We have everywhere used the convention, or analytical extension,

$$\begin{pmatrix} K \\ k \end{pmatrix} = 0 \quad \text{for} \quad k \leq -1 \quad \text{or} \quad k \geq K+1, \quad k \text{ integer.} \tag{32}$$

For $n = \mathcal{O}(1)$ and $N \to \infty$, the limit of the master equation is obtained by using $H_{k-1;K-1;m-1;N-1} = \delta_{k,1}$ in that limit, hence the generalization of Eqs. (6–9) is

$$P_n(t+1) = B_{n;K} P_0(t) + \sum_{k=0}^{\min(K,n)} B_{k;K} P_{n-k+1}(t). \tag{33}$$

As $t \to \infty$, $P_n(t)$ evolves to the time-independent solution to this equation, which can be calculated using generating functions. For $\lambda < 1/K$ this solution is a sum of $K-1$ geometric series for $n \geq 2$,

$$P_0 = 1 - K\lambda \tag{34}$$

$$P_1 = (1-K\lambda)((1-\lambda)^{-K} - 1) \tag{35}$$

$$P_n = (1-K\lambda) \sum_{k=1}^{K-1} \frac{(1-z_k)(1-\lambda+\lambda z_k)}{1-\lambda-(K-1)\lambda z_k} z_k^{-n} \tag{36}$$

where (z_k), $k = 1, \ldots, K-1$ are the $K-1$ roots of the polynomial $(1-\lambda+\lambda z)^K - z$ which differ from 1. These roots all have modulus larger than 1 for $\lambda < 1/K$, so P_n decreases exponentially fast with increasing n. Thus the assumption $n = \mathcal{O}(1)$ is satisfied, provided λ remains less than $1/K$ in the limit $N \to \infty$. The critical value is $\lambda = 1/K$.

The calculation in the scaling limit done above for $K = 2$ is easily extended to generic values for K. If in this case one defines α by

$$\lambda = \frac{1}{K} + \frac{\alpha}{\sqrt{N}},$$

one finds that Eq. (23) becomes

$$\frac{\partial g}{\partial \tau} = (K-1)g + [(K-1)x - K\alpha]\,\frac{\partial g}{\partial x} + \frac{K-1}{2K}\frac{\partial^2 g}{\partial x^2}. \tag{37}$$

By a rescaling of τ, x, and α, one recovers Eq. (23). As a consequence, we expect that for a finite system the expression which generalizes Eq. (27) becomes, for $\lambda = 1/K$ (hence for $\alpha = 0$)

$$\langle t \rangle = a(K)\sqrt{N}, \tag{38}$$

where $a(K)$ is a constant which depends on K.

Conclusion. We have obtained exact expressions for several steady state properties of the random neighbor model, in particular an exact expression for the finite size effects which cut off the $t^{-3/2}$-law.

For the same model a number of other quantities could be calculated, for example correlation functions like the probability of having n' x_i-values less than λ' at time t, assuming there were n such values less than λ at time $t = 0$. One could also extend the model by, instead of removing at each time step the minimal x_i-value, allow to remove any x_i with a probability which depends on x_i.

An interesting question would be to extend our analytical results to other choices for the species affected by change in a given species. In the limit $N \to \infty$ with other parameters held fixed, results obtained in the present paper with *annealed* randomness of species interactions should remain valid for a model with quenched random interactions [14], hence also for any mixture of quenched and annealed interactions.

Acknowledgments. JdB was partialy supported by the National Science Foundation under grant no. PHY 93-09888. HF was partially supported by the Danish Natural Science Research Council under grant no. 11-0244-1. TW and ADJ were partially supported by the US Department of Energy under grant no. DE-FG02-88ER 40388. BD and HF thank the Isaac Newton Institute for its hospitality.

FIG. 1: Dotted and dashed lines: distribution of avalanche life times $q(t)$ at critical point for $N = 5$, 25, 125, 625, and 3125. Full line: analytical expression (26). All cases have $K = 2$ and $\lambda = 1/2$. The results for finite N are exact, obtained by iterating the master equation (1–3) numerically.

FIG. 2: Same results as Fig. 1, except that we show $t^{3/2}q(t)$ versus t/N so that the pure $t^{-3/2}$ power law behaviour corresponds to a straight horizontal line.

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

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

[3] S. Gould and N. Eldredge, Paleobiology, **3,** 115 (1977).

[4] N. Eldredge and S. Gould, Nature **332,** 211 (1988).

[5] D. M. Raup and J. J. Jr. Sepkoski, Science **215**, 1501 (1982).

[6] D. M. Raup, Science **231**, 1528 (1986).

[7] D. M. Raup and G. E. Boyajian, Paleobiology **14**, 109 (1988).

[8] P. Bak, C. Tang and K. Wiesenfeld, Phys. Rev. A **38**, 364 (1987).

[9] S. A. Kauffman, *The Origins of Order,* (Oxford University Press, New York, Oxford, 1993.)

[10] H. Flyvbjerg and B. Lautrup, Phys. Rev. A **46**, 6714 (1992).

[11] P. Bak, H. Flyvbjerg and B. Lautrup, Phys. Rev. A **46**, 6724 (1992).

[12] K. F. Riley, Mathematical Methods for the Physical Sciences, Cambridge University Press, 1974.

[13] M. Abramowitz and I. A. Stegun (eds), Handbook of Mathematical Functions, ninth printing, Dover Publications, Inc., New York.

[14] B. Derrida and G. Weisbuch, J. Physique **47**, (1986) 1297