Genetic information and self-organized criticality

P. R. WILLS[1](*), J. M. MARSHALL[1,2] and P. J. SMITH[1]

[1] *Department of Physics, University of Auckland*
Private Bag 92019, Auckland, New Zealand
[2] *Department of Biomathematics, UCLA - Los Angeles, CA 90095-1766 USA*

Abstract. – The numerical fitnesses of species defined in the Bak-Sneppen model of self-organized criticality are interpreted as binary strings. This allows new species to be generated by mutation of survivors. It is shown that selection in Bak-Sneppen systems defined on both uniform and random lattices produces genotypes in conformity with the Eigen criterion for the accumulation of genetic information in macromolecular sequences.

The Bak-Sneppen (BS) model [1] of self-organized criticality (SOC) mimics certain gross features of global evolutionary dynamics. It has been used to explain the frequency distribution of extinction events of widely varying magnitude deduced from the fossil record [2] and other biological phenomena of apparently fractal origin [3]. The model has been applied to mutation in bacterial populations [4,5], but little effort has been made explicitly to take account of phenomena that arise directly from biological mechanisms.

Here we consider a variant of SOC in which species are generated through genetic replication (inheritance and mutation) and the selective fitness of a species is a phenotypic property. We find that the BS model conforms to the Eigen criterion [6,7] for the accumulation of genetic sequence information. This establishes for the first time a direct connection between disparate approaches to studying evolution: extremal dynamics on lattices of variable dimension; and population dynamics on sequence-encoded fitness landscapes.

The basic one-dimensional BS model consists of a linear array of cells, each of which contains a species that has an assigned fitness (or "barrier") B, selected from a uniform distribution, typically in the range [0 1]. At each time-step the species with the lowest fitness is eliminated, along with its lattice neighbours, and each is replaced by a new species with a randomly assigned fitness value. The system evolves eventually to a critical state characterized by a threshold value of the fitness B_c which, in the thermodynamic limit, determines the ultimate fate of any species: death or survival. The model has been studied in high-dimensional systems [8,9] and applied to the study of random [10], branched [11] and scale-free [12] networks.

(*) E-mail: `p.wills@auckland.ac.nz`

Asymmetric (directed) variants of the model have also been studied, allowing interactions to be interpreted as predator-prey relationships [13, 14].

On the other hand, the quasi-species model of natural selection [6, 7] considers the error-prone replication of genetic sequences in spatially homogeneous systems. Formulation of the sequence population dynamics in terms of ordinary differential equations allows the derivation of an error-threshold criterion for the occurrence of selection in the system. The criterion can be recast as a measure of the number of symbols that can be stably reproduced for any given value of the error rate in symbol-copying. Eigen's model is normally described in terms of standard features of molecular biological systems (metabolism, replication and mutation), but it is quite general and has wider applicability. Our current interest is to show that the BS model can be fitted into the biological paradigm of quasi-species selection.

Our goal of giving a more biological representation of individual species in the BS model is easily achieved. We associate a binary string (the genes) with each species in the array. This has two effects. First, it allows us to calculate fitness from a genotype-phenotype mapping, an archetypal feature of real organisms. Second, it provides for an important modification to the rules of the Bak-Sneppen model: the new species that replace those eliminated at each time-step are created by mutation of surviving species, rather than being chosen from a pre-assigned uniform distribution of fitness. The final result is that the stable storage of genetic information can be measured in systems governed by extremal dynamics.

We formulate the basic BS system in terms of an elementary genotype-phenotype mapping: a binary string $\boldsymbol{b} = b_M b_{M-1} \ldots b_1$ comprised of M bits $b_i \in 0, 1$ representing a species' genes is interpreted as a real number in the range [0 1],

$$B = \sum_{i=1}^{M} 2^{i-1-M} b_i \,. \tag{1}$$

This number is taken to be the fitness of the species' phenotype. At each time-step, the species with the smallest fitness $B_{\min}$ and its lattice neighbours are replaced with new species that are constructed by (asexually) generating variant offspring from surviving "parent" species. When each offspring's genes are produced by randomizing all bits of its parent's genes, the fitness of offspring is distributed uniformly, conforming to the original BS-*ansatz*. It is then a small step to enquire into the extremal dynamics of the system when genealogies arise by partial rather than full randomization of parent genes. The appropriate parameter for measuring the degree of randomization of genes during reproduction is the bit-copy error rate ε. A value of $1/2$ corresponds to an average of $M/2$ in the number of bits mutated during M-bit reproduction, equivalent to choosing the offspring's fitness from a uniform distribution. We have investigated the effect of progressively decreasing ε from its upper limit of $1/2$ (the original BS-*ansatz*) to a value of about $1/M$.

Starting from a random population of binary sequences, the critical threshold value of the fitness $B_{\rm c}$ can be determined from the final stationary value of the maximum B^* struck by the minimum fitness $B_{\min}$ occurring at each time-step (fig. 1). This procedure for determining $B_{\rm c}$ was suggested by Bak and Paczuski [15]. When the offspring replacing eliminated species are generated by mutating the bits of the genetic sequences of randomly chosen surviving parents at a rate ε less than $1/2$, the system still evolves to a critical state. The critical value of the fitness $B_{\rm c}$ becomes arbitrarily close to 1 as ε decreases (table I), limited only by the precision 2^{-M} to which fitnesses are determined. The increase in $B_{\rm c}$ with decreasing ε demonstrates the expected strengthening pressure for selection of 1's rather than 0's in the position(s) of leading magnitude ($i = M, M-1 \ldots$ in eq. (1)).

We have confirmed that our systems belong to the same universality class as the original
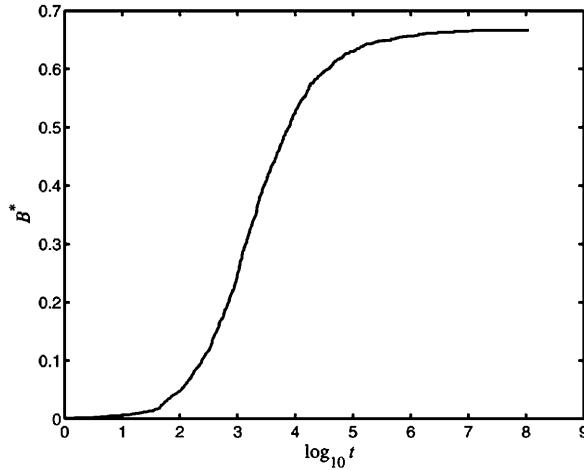
Fig. 1 – Determination of the critical threshold fitness for BS dynamics. The parameter B^* represents the maximum value of minimum fitness $B_{\min}$ struck during the period up until time t (measured in arbitrary steps). The starting distribution was $N = 2048$ strings of length $M = 52$ bits. New species were generated by mutating the sequence of bits of surviving species with an error rate of $\varepsilon = 1/2$.

BS model by measuring several critical exponents (fig. 2). For the probability that the global minimum $B_{\min}$ occurs on a site of age t, $\rho(t) \sim t^{-\alpha}$, we obtain $\alpha = 1.93 \pm 0.03$; for the distribution of first return times to any lattice site, $P_{\mathrm{f}}(t) \sim t^{-\tau_{\mathrm{f}}}$, $\tau_{\mathrm{f}} = 1.56 \pm 0.02$; and for the distribution of all return times, $\tau_{\mathrm{a}} = 0.45 \pm 0.02$. These values are statistically indistinguishable from the accurate values reported in [9] for the original BS model even though they refer to a system with $B_{\mathrm{c}} = 1 - 10^{-10}$.

For any given value of ε the strings in a large population in the critical state are characterized by a relatively sharp cut-off between series of leading 1's and trailing series of bits in which 1's and 0's appear with close to the same probability (fig. 3). The cut-off position indicates the number of bits of genetic information ν that, in the limit of large N, the dynamic system is able to store in spite of errors in the generation of offspring from parents. As N becomes small, the patterns in fig. 3 become less sharp and they are displaced to the left. In the lower limit of $N = 4$ appropriate to a BS system arranged on a 1D circular lattice, the generation of 3 new mutant sequences at each time step will clearly affect the average value of the leading bits in a way which is diluted out in the upper limit $N \to \infty$.

TABLE I – *Variation of the critical threshold fitness B_{c} with the bit-copy error rate ε for BS dynamics on a circular 1D lattice comprised of $N = 512$ points. Sequences were of length $M = 52$ bits and fitnesses were determined to a precision of $2^{-52} \approx 2.2 \times 10^{-16}$.*

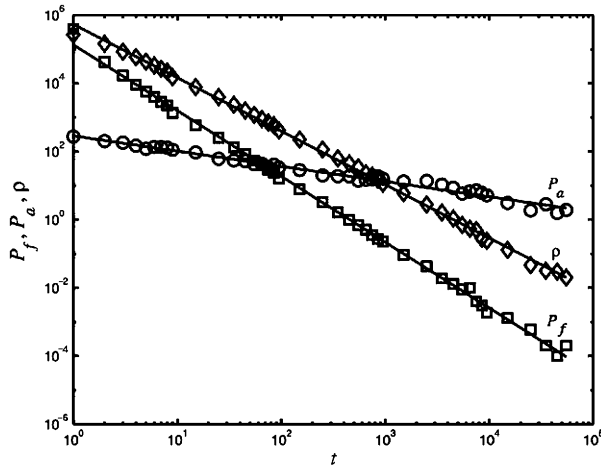| ε	$1 - B_{\mathrm{c}}$
0.5	3.34×10^{-1}
0.10	8.83×10^{-4}
0.05	5.87×10^{-7}
0.04	1.32×10^{-8}
0.03	2.89×10^{-11}
0.025	2.11×10^{-13}

Fig. 2 – Power law distributions characterising critical dynamics in modified BS systems. We show the probability $\rho(t)$ that the global minimum $B_{\min}$ occurs on a site of age t (diamonds), the distribution $P_{\mathrm{f}}(t)$ of first return times to any lattice site (squares), and for the distribution $P_{\mathrm{a}}(t)$ of all return times (circles). Results were obtained from a system with $N = 1024$, $M = 64$, and $\varepsilon = 1/32$.

The critical fitness B_{c} has an approximate digital representation $\boldsymbol{b}_{\mathrm{c}}$ consisting of ν 1's followed by $M - \nu$ 0's and corresponding to the algebraic approximation

$$B_{\mathrm{c}} \approx \sum_{i=M-\nu+1}^{M} 2^{i-1-M}. \tag{2}$$

The maximal fractional error in this estimate of B_{c} is $2^{-\nu}$. Any number less than this estimate
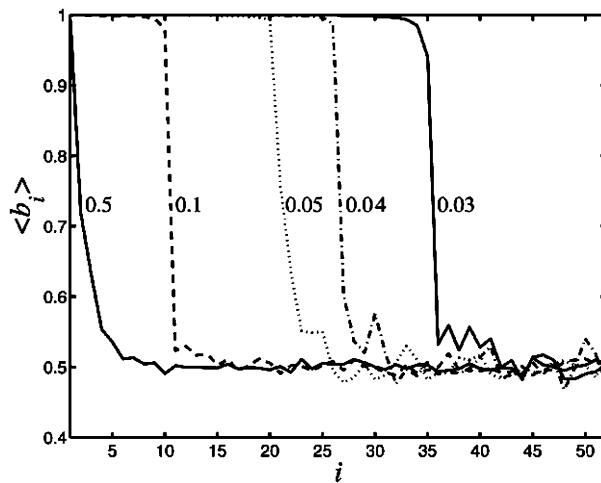

Fig. 3 – Bit structure of self-organized population. For decreasing values of the bit-copy error rate ε (labeled), the time- and population-averaged value $\langle b_i \rangle$ of each sequence-bit (index $1 \leq i \leq M$) was determined over about 10^7 time-steps for a critical population of $N = 729$ sequences of length $M = 52$ bits.

of B_c must have fewer than ν leading 1's. The information storage capacity of the system can then be evaluated from the precise determination of B_c:

$$\nu = -\log_2(1 - B_c). \tag{3}$$

The number of bits stored is not necessarily an integer and the result obtained for ν is insensitive to the chosen string-length M, provided it is significantly greater than ν. When $\varepsilon = 1/2$ we find $B_c \approx 2/3$ suggesting $\nu = 1.59$. With $M = 1 < \nu$, the system does not attain criticality.

The Eigen stability criterion for quasi-species selection [6] provides a measure of the maximum number of symbols ν that can be maintained through error-prone sequence copying. The criterion is expressed in terms of species-specific rate constants for replication, A_i, and degradation, D_i, appearing in the differential equations for the time variation of population numbers n_i of different sequences. For a fixed total population of binary sequences of length M the equations are [7]

$$\dot{n}_i = \left[A_i(1 - \varepsilon)^M - D_i - \bar{E}(t)\right]n_i + \sum_{j \neq i} A_j \varepsilon^{h(i,j)}(1 - \varepsilon)^{M-h(i,j)} n_j \,, \tag{4}$$

where $h(i,j)$ is the Hamming distance between two sequences, $\bar{E}(t) = \sum_k E_k n_k / \sum_k n_k$ is the "mean productivity" and $E_k = A_k - D_k$. The upper limit for the information content ν (in bits) of binary sequences that survive natural selection is related to the symbol-copy error rate ε through [16]

$$\nu = -\frac{\ln \Theta_m}{\ln(1 - \varepsilon)} \,, \tag{5}$$

where the "superiority parameter" $\Theta_m = \frac{A_m}{E_{j+m} + D_m}$ measures the advantage in growth of the master sequence $\boldsymbol{b}_m$ (bit-complement of the sequence space origin) relative to its mutants. It is evident that in our BS systems the replication rate A_i of all sequences is identical because parent-species are chosen at random from the surviving population at each time step. Also, it is found that in the critical state n_i is uniform for all sequences with the correct ν leading bits and zero otherwise. However, the degradation rates D_i are not intrinsic constants belonging to species as conceived in the Eigen model. Rather, they have a complicated dependence on both the population vector $\boldsymbol{n}$ and the spatially correlated structure of avalanches that occur when the neighbours of the species with minimum fitness are destroyed at each time step. This precludes any simple analytical definition of a putative Θ_m, even by using the methodology developed by McCaskill [17].

Figure 4 shows that genetic information storage in BS systems nevertheless conforms to the Eigen criterion, not only on simple one-dimensional lattices but also in higher-dimensional systems. The values of Θ_m determined by least-squares fitting to the data in fig. 4 are 2.807, 1.466 and 1.245 (all ± 0.001 std. err.) for 1, 2 and 3D systems, respectively. The fact that Θ_m turns out to be so well determined for BS systems with critical dynamics indicates that the time-averaged degradation rate D_i for each species is a well-defined parameter in spite of the complexity of the selection criterion that generates SOC.

We have also investigated BS dynamics with varying bit-copy error rates on random networks with connections Poisson-distributed between nodes. For any given value of the mean degree of connectivity $\langle z \rangle$ greater than the critical value of unity [10], the critical dynamics still conform to the Eigen criterion, but Θ_m decreases with $\langle z \rangle$ (fig. 5), giving the empirical relation $\ln \Theta_m \approx \frac{1}{\langle z \rangle}$. The information storage capacity of replicating genes in an SOC system subject to the elementary BS selection criterion approximates to the relationship $\nu \approx -\frac{1}{\langle z \rangle \ln q}$, where $q = 1 - \varepsilon$. This result applies to random networks of low relative connectivity $(1 < \langle z \rangle \ll N)$.
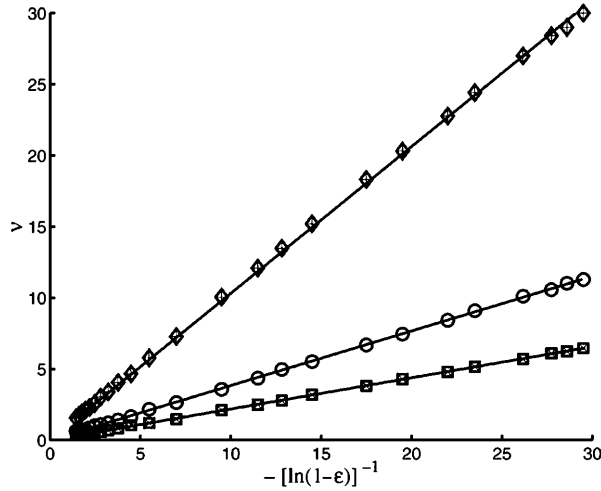
Fig. 4 – Conformity of SOC systems to Eigen's criterion. The information carrying capacity ν of a critical population of species was determined by using eq. (3) for different values of the bit-copy error rate ε. Results are shown for BS dynamics on symmetrical, closed lattices of 1 (diamonds), 2 (circles) and 3 (squares) dimensions. Lines of best fit have been calculated according to eq. (4).

Clearly, the surviving population in SOC systems of the BS sort is a quasi-species as defined by Eigen [6]. However, there is a fundamental difference between more classical evolutionary systems and SOC dynamics, arising from the selection criterion that is applied to species in SOC systems. BS selection is a two-step process. First, the species with the global minimum of fitness is eliminated. When this first step constitutes the entire selection process, the system does not evolve to a critical state [18]. In the second step of BS selection, the lattice neighbours of the species of minimum fitness are eliminated. Without the influence of lattice
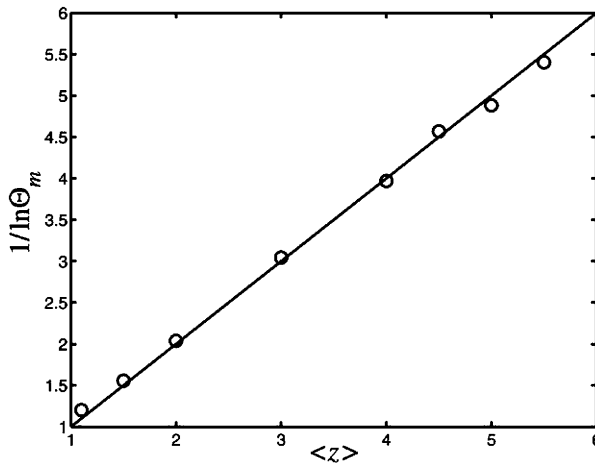

Fig. 5 – Eigen selection behaviour on random SOC networks. The logarithm of the superiority parameter Θ_m was determined from the slope of graphs of the form of fig. 4 for random networks of Poisson-distributed connectivity of mean degree $\langle z \rangle$. The line corresponds to eq. (5).

network relationships on the selection procedure, the systems behave like the much-studied well-stirred reactor [6, 7, 16], but the extremal selection criterion maintains the system in a critical state [19], rather than fluctuating about a stable fixed attractor defining a classical quasi-species [6]. The inclusion of lattice interactions in the complete BS selection criterion has a further complicated influence on the value of the superiority parameter Θ_m that characterizes the system in the critical state.

It will be of interest to study the behaviour of our extended BS model on Barabási-Albert–type scale-free networks [20]. It has been shown [12] that the standard BS model on such networks ($\varepsilon = 1/2$ in our model) leads to dynamics in which the critical barrier B_c depends on the population size N and that $B_c \to 0$ as $N \to \infty$, implying that the information-storage capacity of sequences is reduced as the size of the system increases. It is reasonable to expect a reduction in the bit-copy error rate will be able to compensate for an increase in the size of the network to maintain a given information-storage capacity (value of B_c). Such an effect may be relevant to biological phenomena that depend on global properties of ecosystems.

Punctuated equilibria arise in systems governed by extremal dynamics on smooth single-peaked landscapes corresponding to genotype-phenotype mappings like eq. (1). The same phenomenon is observed in population-dynamic systems described by ODEs when the fitness landscape has certain neighbourhood relationships [21,22]. Further study of genetics in critical systems and comparison with the phylogenies of conserved and non-conserved symbols in actual genetic sequences may ultimately provide an empirical test of the bold postulate of Bak [23] that biological evolution is a self-organized critical process.

REFERENCES

[1]   Bak P. and Sneppen K., *Phys. Rev. Lett.*, **71** (1993) 4083.
[2]   Sneppen K., Bak P., Flyvbjerg H. and Jensen M. H., *Proc. Natl. Acad. Sci. USA*, **92** (1995) 5209.
[3]   Sole R. V., Manrubia S. C., Benton M., Kauffman S. and Bak P., *TREE*, **14** (1999) 156.
[4]   Bose I. and Chaudhuri I., *Int. J. Mod. Phys. C*, **12** (2001) 675.
[5]   Donangelo R. and Fort H., *Phys. Rev. Lett.*, **89** (2002) 038101.
[6]   Eigen M., *Naturwiss.*, **10** (1971) 465.
[7]   Eigen M., McCaskill J. and Schuster P., *Adv. Chem. Phys.*, **75** (1989) 149.
[8]   Ray T. S. and Jan N., *Int. J. Mod. Phys. C*, **12** (1994) 1375.
[9]   De Los Rios P., Marsili M. and Vendruscolo M., *Phys. Rev. Lett.*, **80** (1998) 5746.
[10] Christensen K., Donangelo R., Koiller B. and Sneppen K., *Phys. Rev. Lett.*, **81** (1998) 2380.
[11] Datta A. S., Christensen K. and Jensen H. J., *Europhys. Lett.*, **50** (2000) 162.
[12] Moreno Y. and Vazquez A., *Europhys. Lett.*, **57** (2002) 765.
[13] Vendruscolo M., De Los Rios P. and Bonesi L., *Phys. Rev. E*, **54** (1996) 6053.
[14] Stauffer D. and Jan N., *Int. J. Mod. Phys. C*, **11** (2000) 147.
[15] Bak P. and Paczuski M., *Proc. Natl. Acad. Sci. USA*, **92** (1995) 6689.
[16] Kuppers B. O., *Molecular Theory of Evolution* (Springer, Berlin) 1983.
[17] McCaskill J., *J. Chem. Phys.*, **80** (1984) 10.
[18] Stauffer D. and Newman M. E. J., *Phys. Rev. Lett.*, **72** (1994) 4045.
[19] Flyvbjerg H., Sneppen K. and Bak P., *Phys. Rev. Lett.*, **71** (1993) 4087.
[20] Barabási A.-L. and Albert R., *Science*, **286** (1999) 509.
[21] Fontana W. and Schuster P., *Science*, **280** (1998) 1451.
[22] Stadler B. M. R., Stadler P. F., Wagner G. and Fontana W., *J. Theor. Biol.*, **213** (2001) 241.
[23] Bak P., *How Nature Works* (Oxford University Press, Oxford) 1997.