

## An evolutionary version of the random energy model

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1993 J. Phys. A: Math. Gen. 26 L1195

(<http://iopscience.iop.org/0305-4470/26/23/001>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.68

The article was downloaded on 01/06/2010 at 20:09

Please note that [terms and conditions apply](#).

## LETTER TO THE EDITOR

# An evolutionary version of the random energy model

Silvio Franz<sup>†</sup>, Luca Peliti<sup>‡§||</sup>, and Mauro Sellitto<sup>‡</sup>.

<sup>†</sup> Laboratoire de Physique Théorique, Ecole Normale Supérieure, 24, rue Lhomond, F-75231 Paris Cedex 05, France

<sup>‡</sup> Dipartimento di Scienze Fisiche and Unità INFM, Università 'Federico II', Mostra d'Oltremare, Pad. 19, I-80125 Napoli, Italy

<sup>§</sup> Boursier Henri de Rothschild, Institute Curie, Section de Physique et Chimie, Laboratoire Curie, 11, rue Pierre et Marie Curie, F-75231 Paris Cedex 05, France

Received 26 July 1993

**Abstract.** We consider Eigen's quasispecies model of molecular evolution in the case in which the reproduction rates of different molecular species are quenched independent random variables. We show, by analogy with the random energy model, the existence of two phases: a 'neutral' phase, where no adaptation effects are exhibited, and a (pathological) 'adapted' phase, where the population is made of identical, optimally adapted molecules, and does not evolve.

One of the problems in understanding the origin of life lies in the identification of the mechanism by which starting from the autocatalysis of simple replicating molecules, a Darwinian evolutionary mechanism based on reproduction, mutation and selection may set in. The first steps in this direction were made by Manfred Eigen more than twenty years ago (Eigen 1971). In this letter he introduced a model of an evolving population of self-replicating entities (which may be identified with the ancestors of RNA molecules), which we shall call the *quasispecies model*. He was able to show that the eventual composition reached by this population results from a compromise between replication efficiency and frequency of mutations, which recalls the compromise between energy minimization and entropy maximization which determines thermal equilibrium. Moreover, he introduced the notion of the *error threshold* in mutation rate: when the rate is smaller than this threshold, the population is made up of molecules of rather similar primary structure (nucleotide sequence). Above this threshold, the population becomes closer and closer to a random sample. The error threshold is analogous to an order-disorder transition in statistical mechanics.

This analogy was put on a firm basis by Leuthäusser (1987). Indeed, a version of the quasispecies model may be cast in a form amenable to a treatment by the classical methods of *equilibrium statistical mechanics*. More recently, Tarazona (1992) exploited Leuthäusser's formulation to investigate numerically several simple forms of the quasispecies model and to highlight the analogy between the error threshold and phase transitions.

In this letter we show that the methods of the statistical mechanics of spin glasses allow us to investigate analytically the behaviour of the quasispecies model in some simple cases. We consider in particular the evolutionary version of Derrida's random

|| To whom correspondence should be addressed.

energy model (REM) (Derrida, 1981). This corresponds, in the quasispecies context, to the 'rugged fitness landscapes' introduced in evolutionary models by Kauffman and Levin (1987). We shall see that, in this case, the error threshold corresponds to a freezing transition in sequence space: for mutation rates smaller than the threshold, the population is made of molecules with identical primary structure and does not evolve.

Consider a population of self-reproducing molecules, whose structure may be described by a collection of  $N$  binary variables,  $S_i = \pm 1$ ,  $i = 1, 2, \dots, N$ . Then the fraction  $x_S$  of molecules of structure  $S = (S_1, S_2, \dots, S_N)$  obeys the following evolution equation:

$$x_S(t+1) = \frac{1}{Z(t)} \sum_{S'} W_{SS'} A(S') x_{S'}(t) \quad (1)$$

where  $A(S)$  is the average number of offspring that a molecule of structure  $S$  produces at the next generation (if one assumes infinite environmental carrying capacity) and  $W_{SS'}$  is the conditional probability that the reproduction of a molecule of structure  $S'$  effectively produces a molecule of structure  $S$ , and therefore represents the effects of mutations. A convenient expression for the matrix  $W$  is given by

$$W_{SS'} = w^{d(S,S')} (1-w)^{N-d(S,S')} \quad (2)$$

where  $0 < w < 1$  is the probability of having one mutation per unit and per generation, and

$$d(S, S') = \frac{1}{2} \sum_{i=1}^N (1 - S_i S'_i) \quad (3)$$

is the number of different units in the structures  $S$  and  $S'$  respectively. The factor  $Z(t) = \sum_S A(S) x_S$  ensures the normalization of  $x_S(t)$  at any generation. In order to derive equation (1) one assumes that 'generations' of the self-reproducing molecules are non-overlapping and that the number of molecules in the population is sufficiently large to neglect fluctuations in the  $x_S(t)$ .

It was shown by Leuthäusser (1987) that the quasispecies model corresponds to a problem in *equilibrium* statistical mechanics. Indeed, the matrix

$$T_{SS'} = W_{SS'} A(S') \quad (4)$$

can be considered as the transfer matrix for an Ising-like system of Hamiltonian

$$-\beta H[S(t)] = \sum_{t=1}^T \sum_{i=1}^N \beta S_i(t) S_i(t+1) + \sum_{t=1}^T \ln[A(S(t))] \quad (5)$$

where the 'inverse temperature'  $\beta$  is given by

$$\beta = -\frac{1}{2} \ln[w/(1-w)] \quad (6)$$

and an irrelevant additive constant has been understood. It is important to keep in mind that the quantities one naturally computes within this approach are the *unrenormalized* fractions  $y_S(t)$ , which satisfy the evolution equation

$$y_S(t+1) = \sum_{S'} T_{SS'} y_{S'}(t). \quad (7)$$

The relation between the  $y$ 's and the  $x$ 's is obviously given by

$$x_S(t) = y_S(t) \left/ \sum_S y_S(t) \right. \quad (8)$$

The 'observables' (population averages) are obtained by averaging over the  $x$ 's, i.e. by normalizing the sum over the  $y$ 's. If the reproduction rates are random variables, we are interested in the corresponding expectation values of the observables, which can be obtained by replica methods.

Following Derrida (1981), we assume that the reproduction rates  $A(S)$  are given by independent 'quenched' random variables for each different structure  $S$ . We assume in fact  $A(S) = \exp(kE(S) + \beta h \sum_i S_i)$ , where the 'energy'  $E(S)$  is a Gaussian random variable of zero mean and variance equal to  $N/2$ . We have also introduced an 'external field'  $h$ , controlling the relative ratio of positive to negative  $S$ 's. Introducing the replica indices  $a = 1, 2, \dots, n$ , the overlap matrix

$$Q^{ab}(tt') = \frac{1}{N} \sum_i S_i^a(t) S_i^b(t') \quad (9)$$

and performing the average over the 'energy' distribution, we obtain

$$[Z^n]_{\text{av}} = \text{Tr}_{S_i^a(t)} \exp \left\{ \frac{k^2}{4} \sum_{ab} \sum_{tt'} \delta^{(K)}(Q^{ab}(tt'), 1) + \frac{\beta}{N} \sum_i \sum_q \sum_t (S_i^q(t) S_i^q(t+1) + h S_i(t)) \right\} \quad (10)$$

where  $\delta^{(K)}(x, y)$  denotes a Kronecker delta. In this formulation, we recognize a generalization of the problem considered by Franz *et al* (1992, 1993), where several real replicas of the same spin glass are constrained to have fixed values of the overlaps between themselves. In our case, the number of replicas,  $T$ , is assumed to go eventually to infinity, and the constraint is applied only between successive 'generations'. We may distinguish two cases: (1) 'frozen' behaviour: for some replicas  $(a, b)$  one has  $\lim_{|t-t'| \rightarrow \infty} Q^{ab}(tt') = 1$ ; (2) 'evolving' behaviour: for any  $t \neq t'$ , one has  $|Q^{ab}(tt')| < 1, \forall (a, b)$ . Moreover we may assume a one-step replica symmetry breaking scheme (Gross and Mézard, 1984), introducing a variational parameter  $m (0 < m \leq 1)$ . We thus obtain  $[Z^n] = \exp(-NG)$ , where we express the 'free energy'  $G$  in terms of the 'magnetization'  $\mu = \langle S_i^a(t) \rangle$ , and we have

$$-\frac{G}{n} = \begin{cases} \frac{k^2}{4} mT^2 + \beta T + \frac{1}{m} s(\mu) + \beta h \mu T & \text{case (1)} \\ \frac{k^2}{4} mT + \frac{1}{m} T \ln \zeta_1(\beta m, h) & \text{case (2).} \end{cases} \quad (11)$$

We remind that  $T$  is the number of generations we are considering. We wish eventually to let  $T \rightarrow \infty$ . In the previous formula, we have introduced the following expressions:

$$s(\mu) = \ln 2 - \frac{1}{2} [(1 - \mu) \ln(1 - \mu) + (1 + \mu) \ln(1 + \mu)] \quad (12)$$

$$\zeta_1(\beta, h) = e^{\beta} \cosh \beta h + \sqrt{e^{-2\beta} + e^{2\beta} \sinh^2 \beta h}. \quad (13)$$

We must now identify the saddle point with respect to  $m$  and  $\mu$ . A straightforward analysis yields the following solutions: for case (1)

$$m = \frac{2\sqrt{s(\mu)}}{kT}; \quad \tanh^{-1}\mu = \beta mhT \tag{14}$$

in the absence of replica symmetry breaking we set  $m=1$  and obtain

$$\tanh^{-1}\mu = \beta hT. \tag{15}$$

The corresponding values of  $G$  are given by

$$-\frac{G}{n} = \begin{cases} kT\sqrt{s(\mu)} + \beta T + \beta h\mu T & m < 1 \\ (k^2/4)T^2 + s(\mu) + \beta T + \beta h\mu T & m = 1. \end{cases} \tag{16}$$

Since we have to maximize  $G$  with respect to  $m$ , we see that the replica symmetric solution must always be discarded, in the  $T \rightarrow \infty$  limit, for case (1). This is confirmed by an analysis via the original Derrida (1981) probabilistic method.

For case (2) we obtain, when  $m < 1$ :

$$\frac{k^2}{4} = \frac{1}{m^2} \left[ \ln \zeta_1(\beta m, h) - \beta m \left. \frac{\partial \ln \zeta_1}{\partial \beta} \right|_{\beta = \beta m} \right] = \frac{1}{m^2} S_1(\beta m, h). \tag{17}$$

This equation must be supplemented by the equation connecting  $h$  and  $\mu$ :

$$\mu = \frac{e^{\beta m} \sinh \beta mh}{\sqrt{e^{-2\beta m} + e^{2\beta m} \sinh^2 \beta mh}} \tag{18}$$

The corresponding values of the 'free energy'  $G$  are given by

$$-\frac{G}{n} = \begin{cases} \frac{k}{2} T \frac{[S_1(\beta m, h) + \ln \zeta_1(\beta, h)]}{(S_1(\beta m, h))^{1/2}} & m < 1 \\ \frac{k^2}{4} T + T \ln \zeta_1(\beta, h); & m = 1. \end{cases} \tag{19}$$

For small values of  $\beta$  and  $k$ , it is easy to see that the stable phase is the replica symmetric ( $m=1$ ) one of case (2). When  $h=0$ , it is easy to identify the line  $k=k_c(\beta)$  in which the free energy of this phase becomes equal to that of the replica symmetry breaking (RSB) one for case (1) and (2) respectively; the second transition line,  $k=k'_c(\beta)$ , is most easily identified by setting  $m=1$  in (17):

$$\begin{aligned} \text{Case (1)} \quad k_c(\beta) &= 2(\sqrt{\ln 2} - \sqrt{\beta - \ln \cosh \beta}) \\ \text{Case (2)} \quad k'_c(\beta) &= 2\sqrt{\ln 2 \cosh \beta - \beta \tanh \beta}. \end{aligned} \tag{20}$$

The two transition lines are shown in figure 1. It is clear that the first line always prevails. When the 'magnetic field'  $h$  is non-zero, one has to solve the corresponding equations numerically. Nevertheless the general aspect of the transition curves remains

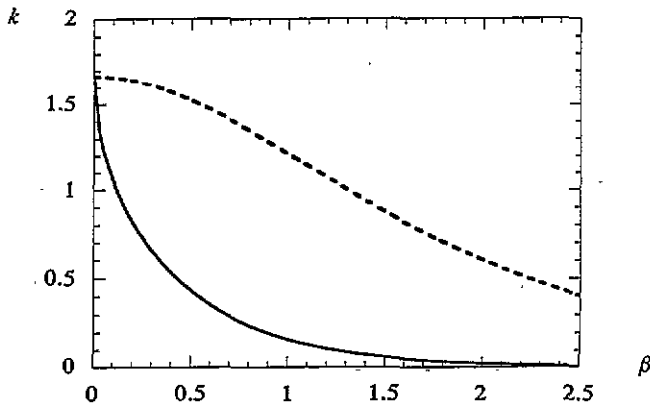


Figure 1. Transition lines  $k_c(\beta)$  (full line) and  $k'_c(\beta)$  (dashed line) in the  $(k, \beta)$  plane for  $h=0$ .

the same. We have thus identified the transition line  $k_c(\beta)$  above which the population is frozen in an adaptation optimum: all individual molecules are identical at a given generation and from generation to generation.

Below the transition line, the population does not exhibit adaptation, and its behaviour is similar to a population evolving in a 'flat fitness landscape' (Derrida and Peliti, 1991), i.e. in the absence of selection. A reminder of the replication mechanism is exhibited by the self-overlap  $Q^{aa}(t't') = \tanh^{t-t'} \beta$ , which shows that the correlations in the 'average genotype' decay like those of a one-dimensional Ising model.

The model we have just discussed is pathological in the sense that when adaptation takes place all the molecules of the population become identical. Selection leads therefore to the persistence of a single molecular species. This is analogous to the lack of thermal fluctuations which characterises the REM spin glass. When the 'fitness landscape'  $A(\mathcal{S})$  is smoother—i.e. when some correlations between the values of  $A(\mathcal{S})$  for similar structures  $\mathcal{S}$  exist—mutations lead to some differentiation within the population. Smoother landscapes are obtained, for example, by considering the  $NK$  landscapes discussed by Kauffman (1989) or the 'Hopfield landscape' investigated by Leuthäusser (1987) and Tarazona (1992). The analytical methods of the statistical physics of spin glasses may also be applied to the last case, at least close enough to the error threshold (Sellitto and Peliti, in preparation).

## References

- Derrida B 1981 *Phys. Rev. B* **24** 2613  
 Derrida B and Peliti L 1991 *Bull. Math. Biol.* **53** 355  
 Eigen M 1971 *Naturwissenschaften* **58** 465  
 Franz S, Parisi G and Virasoro M A 1992 *J. Phys. I France* **2** 1869  
 Franz S, Parisi G and Virasoro M A 1993 *Europhys. Lett.* **22** 405  
 Gross D J and Mézard M 1984 *Nucl. Phys. B* **240** [FS12] 431  
 Kauffman S A 1989 *Complex Systems: SFI Studies in the Sciences of Complexity* ed D E Stein (Reading, MA: Addison-Wesley) p 618  
 Kauffman S A and Levin S 1987 *J. Theor. Biol.* **128** 11  
 Leuthäusser I 1987 *J. Stat. Phys.* **48** 343  
 Tarazona P 1992 *Phys. Rev. A* **45** 6038