The Ising Quantum Chain is Equivalent to a Model of Biological Evolution

E. Baake
Max-Planck-Institut für Entwicklungsbiologie, Spemannstr. 35, D-72076 Tübingen, Germany

M. Baake, H. Wagner
Institut für Theoretische Physik, Universität Tübingen, Auf der Morgenstelle 14, D-72076 Tübingen, Germany
(June 14, 1996)

A sequence space model which describes the interplay of mutation and selection in molecular evolution is shown to be equivalent to an Ising quantum chain. Three explicit examples with representative fitness landscapes are discussed and exactly solved with methods from statistical mechanics.

87.10.+e, 75.10.-b, 64.60.Cn, 05.50.+q

One-dimensional systems, and quantum chains in particular, have long been important tools to understand, at least *approximately*, various physical situations, and there is even a recipe 'How to Reduce Practically Any Problem to One Dimension' [1]. As a complement, we present a problem of biochemical physics that may be mapped *exactly* onto a quantum chain. Selected examples can then be solved without approximation.

In the theory of (molecular) biological evolution, various sequence space models are well established, the best-known being Kauffman's adaptive walk [2] and Eigen's quasispecies model [3]. Whereas the former describes a hill-climbing process of a genetically homogeneous population in tunably rugged fitness landscapes, the latter includes the genetic structure of the population due to the balance between mutation and selection. For equal fitness landscapes, the quasispecies model is thus more difficult to treat than the corresponding adaptive walk.

Some progress was made in [4] through the identification of the quasispecies model with a specific, anisotropic 2D Ising model: The mutation-selection matrix is equivalent to the row transfer matrix, with the mutation probability as a temperature-like parameter, and error thresholds corresponding to phase transitions. This equivalence was exploited to treat simple fitness landscapes as well as spin-glass Hamiltonians with methods from statistical mechanics [5–7]. Of these results, most are approximate or numerical, and the few exact ones in [5] are of limited value as the order parameter was not calculated correctly.

The quasispecies model assumes mutations to originate as replication errors on the occasion of reproduction events. An alternative was introduced in [8] and describes mutation and selection as going on in parallel; we would like to abbreviate it as para-muse (parallel mutation-selection) model. In subsequent investigations [9,10], this model turned out to be both more powerful and structurally simpler than the quasispecies model. Which is the more appropriate one from the biological point of view amounts to the question whether rates of molecular evolution are closer to constant per generation or constant in time – a long-standing, but still unresolved issue [11,12]. Even in the former situation, however, the parallel version is an excellent approximation. Note that both models are sequence space versions of the muse equations of classical population genetics [13].

In this Letter, we will show that, in the same way as the quasispecies model is equivalent to the row transfer matrix of a 2D Ising model, the para-muse model corresponds to the Hamiltonian of an Ising quantum chain. Methods of statistical mechanics will then be employed to treat three sample fitness landscapes exactly, with emphasis on the correct order parameter. More biological implications will be dealt with elsewhere [14].

Evolution and quantum chain. In the framework of sequence space models, genetic information (like nucleic acid strings) is identified with points in (binary) sequence space, e.g. $\{-1,1\}^N$, where $N \in \mathbb{N}$ is the (fixed) length of the string considered; so there are $n = 2^N$ different sequences (or alleles in the language of genetics) termed A_i, $i = 1, \ldots, n$. The composition of an (infinite) population of haploid organisms (they carry one copy of every gene only) under the influence of mutation and selection acting *independently* of each other is described by the following ODE system (for review, see [15]):

$$\dot{x}_i \;=\; x_i(r_i - \sum_{j=1}^{n} r_j x_j) + \sum_{j=1}^{n} m_{ij} x_j \;. \qquad (1)$$

Here, x_i denotes the relative frequency of A_i individuals ($1 \le i \le n = 2^N$), the r_i their (Malthusian) fitness (i.e. the difference between reproduction and death rates), and m_{ij} the rate at which A_j mutates to A_i. When every digit mutates independently at rate $\mu \ge 0$, we have

$$m_{ij} \;=\; \begin{cases} \mu\,, & d_{ij} = 1 \\ -N\mu\,, & i = j \\ 0\,, & \text{otherwise} \end{cases} \qquad (2)$$

where $d_{ij} := d(A_i, A_j)$ is the Hamming distance between A_i and A_j (i.e. the number of positions where the two strings differ). The r_i are, as yet, unspecified; their choice determines the so-called fitness landscape.

The transformation [16]

$$z_i(t) \;:=\; x_i(t) \exp\Big(\sum_{j=1}^{n} r_j \int_0^t x_j(\tau)\, d\tau \Big) \tag{3}$$

reduces the ODE (1) to the linear system

$$\dot{\boldsymbol{z}} \;=\; (M+R)\,\boldsymbol{z}\;, \tag{4}$$

where $R := \mathrm{diag}(r_1, \ldots, r_n)$, and $M = (m_{ij})$ is the (symmetric) matrix of mutation rates. The solution of Eq. (4) then yields a solution of the original Eq. (1) via

$$x_i \;=\; \frac{z_i}{\sum_{j=1}^{n} z_j}\,. \tag{5}$$

The problem is now completely solved as soon as the spectrum of the symmetric matrix $\widetilde{\mathcal{H}} := M+R$ is known. For $\mu > 0$, $\widetilde{\mathcal{H}}$ plus a suitable constant is primitive (i.e. some power of this sum has strictly positive entries only). Consequently, the largest eigenvalue of $\widetilde{\mathcal{H}}$ is unique. It corresponds to the growth rate ('mean fitness') of the equilibrium population, the composition of which is given by the Perron-Frobenius (PF) eigenvector.

Let us first observe that, owing to our two-letter alphabet, the possible occupation numbers are 0 and 1 at every site. With the canonical basis of $\otimes_{i=1}^{N}\mathbb{C}^2$, we can thus rewrite $\mathcal{H} := \widetilde{\mathcal{H}} + N\mu$ in terms of Pauli's matrices:

$$\begin{aligned}\mathcal{H} \;=\; & \mu \sum_{k=1}^{N} \sigma_k^x \;+ \alpha_0 \mathbb{1} \;+ \sum_{k=1}^{N} \alpha_k \sigma_k^z \\ & + \sum_{k<k'}^{N} \alpha_{kk'} \sigma_k^z \sigma_{k'}^z + \sum_{k<k'<k''}^{N} \alpha_{kk'k''} \sigma_k^z \sigma_{k'}^z \sigma_{k''}^z \\ & + \;\ldots \text{ (terms up to } N\text{-fold products).}\end{aligned} \tag{6}$$

Here, $\sigma^x := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\sigma^z := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, and

$$\sigma_k^a \;:=\; \mathbb{1} \otimes \ldots \otimes \mathbb{1} \otimes \sigma^a \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1} \tag{7}$$

with $a \in \{x, z\}$ and σ^a in the k-th place. The σ_k^x flips the k-th digit ('mutation'), and σ_k^z measures the value of the variable (the spin) at site k. As the products of σ_k^z matrices up to N-th order span the space of $2^N \times 2^N$ diagonal matrices, the most general fitness landscape is covered. Note that, due to the projective properties of Eq. (1), all Hamiltonians that differ only in their constant terms lead to the same equilibrium states $\boldsymbol{x}$.

Obviously, $\mathcal{H}$ is, at the same time, the Hamiltonian of an Ising quantum chain in a transverse magnetic field, with general interactions within the chain. (With an automorphism of the Pauli-algebra, namely $\sigma^x \to \sigma^z$, $\sigma^y \to \sigma^y$, $\sigma^z \to -\sigma^x$, one obtains its more familiar form.) Since energies appear with positive sign here, the PF eigenvalue belongs to the ground state. Even with two-spin interactions only (the 'classical' Ising quantum chain), $\mathcal{H}$ provides a very rich (and difficult) range of fitness landscapes, including e.g. spin-glass landscapes as tackled numerically in [7]. Let us, in what follows, focus on three *exactly solvable* examples from this model class.

Mount Fujiyama landscape. Let us first consider the following simple Hamiltonian which is composed from non-interacting one-site Hamiltonians $\mathcal{H}_k$:

$$\mathcal{H} \;=\; \sum_{k=1}^{N} \mathcal{H}_k \;=\; \sum_{k=1}^{N} (\mu\, \sigma_k^x + \alpha_k\, \sigma_k^z)\,. \tag{8}$$

In the evolution model, this is the case of no interaction between sites, or purely additive fitness (cf. Eq. (1)):

$$r_i \;=\; \sum_{j=1}^{N} \alpha_j s_j^{(i)}\;, \tag{9}$$

where $s_j^{(i)}$ is the value of the variable at site j of sequence A_i. This is similar to Kauffman's landscape of an N-dimensional 'Fujiyama' peak [2] (with different scaling).

The spectrum of $\mathcal{H}$ (and $\widetilde{\mathcal{H}}$) may be composed of the eigenvalues of the $\mathcal{H}_k$, $\lambda_k = \pm\sqrt{\alpha_k^2 + \mu^2}$. (A similar structure pertains in the transfer matrix of the corresponding quasispecies model, as exploited in [5,17].)

For $\alpha_j \equiv \alpha$, the ground state energy per spin (or mean fitness of the equilibrium population per site) is:

$$w \;:=\; \widetilde{\lambda}_{\max}/N \;=\; \alpha(-h + \sqrt{1+h^2})\,, \tag{10}$$

where $\widetilde{\lambda}_{\max}$ is the largest eigenvalue of $\widetilde{\mathcal{H}}$, and $h := \mu/\alpha$.

To characterize the genetic structure of the population, the average surplus u of sites with value $+1$,

$$u \;:=\; \frac{\sum_{i=1}^{n} u_i\, v_i}{\sum_{j=1}^{n} v_j}, \quad u_i \;:=\; \frac{1}{N} \sum_{j=1}^{N} s_j^{(i)}\;, \tag{11}$$

is appropriate, where v_i are the components of the PF eigenvector of $\mathcal{H}$. In the present case, we find

$$u \;=\; \frac{1}{h + \sqrt{1+h^2}} \;=\; \frac{w}{\alpha} \tag{12}$$

as it must be, since fitness is proportional to the surplus in this case (cf. Eq. (9)). Both w and u vanish smoothly when $\mu \to \infty$, and no phase transition occurs in the macroscopic (or thermodynamic) limit (see Figs. 1, 2).

Onsager landscape. The previous example relied on a landscape with a single global maximum that can be reached from any sequence in the space by mutational steps each of which increases fitness. As a case study of a more rugged (and more realistic) landscape, let us discuss the well-known Ising quantum chain in a transverse field, with constant coupling γ of nearest-neighbour spins,

$$\mathcal{H} \;=\; \mu \sum_{k=1}^{N} \sigma_k^x + \gamma \sum_{k=1}^{N} \sigma_k^z \sigma_{k+1}^z\;. \tag{13}$$

We assume cyclic boundary conditions, i.e. $\sigma_{N+1}^z = \sigma_1^z$. Let k_i be the number of domain walls in a given sequence

A_i. Then, $r_i = \gamma(N - 2k_i)$. For $\gamma > 0$ and N even (to which we will restrict ourselves for simplicity), the sequences of highest fitness are $\pm(1, \ldots, 1)$. A spin flip at site j implies a change in fitness of $-2s_j(s_{j-1}+s_{j+1})$. Let $L(A_i)$ be the length of the shortest domain in A_i. Then, L is the minimum number of mutations required to reach, from A_i, a sequence with higher fitness. In particular, any A_i with $L(A_i) > 1$ is a ridge (i.e. mutations are either deleterious or selectively neutral).

As is well-known [18–20], the Hamiltonian (13) is solvable and can be transformed into a system of non-interacting fermions with energies $2\gamma\Lambda_k$ where

$$\Lambda_k \;=\; \left(1 + h^2 + 2h\cos(\frac{k\pi}{N})\right)^{1/2}\,,\; 0 \le k < N\,, \qquad (14)$$

and $h = \mu/\gamma$. In the thermodynamic limit, one obtains for the ground state energy per spin of $\widetilde{\mathcal{H}}$:

$$w \;=\; \frac{2\gamma}{\pi}(1+h)E(\frac{\pi}{2},\theta) - \mu, \quad \theta^2 \;=\; \frac{4h}{(1+h)^2} \qquad (15)$$

with $E(\pi/2, \theta)$ the complete elliptic integral of the second kind; see Fig. 1. Let us recall that, in the thermodynamic limit, the system has a second-order phase transition at $h = 1$ with conformal invariance.

Again, the surplus u (or order parameter) as defined in Eq. (11) is the proper quantity to characterize the biological population. It is similar but not directly related to the physical magnetization; although it is not an observable in the usual (quantum mechanical) sense, it is nevertheless an observable quantity. Going through the somewhat painful exercise of calculating the eigenvectors corresponding to the largest and second-largest eigenvalues [18,19], we find that, in the thermodynamic limit,

$$u \;=\; \begin{cases} (1-h)^{1/2}\,, & 0 \le h < 1 \\ 0\qquad\;\;, & h \ge 1 \end{cases} \qquad (16)$$

in contrast to the (physical) magnetization, which is [19] $m = (1-h^2)^{1/8}$ for $0 \le h \le 1$ (see Figs. 2 and 3).

Mean-field landscape. Nearest neighbour interaction has provided us with a fitness landscape with ridges and neutrality, but it should not be taken too literally as a model of biological interaction. After all, DNA strings serve as templates for proteins, which then fold in three dimensions, thus giving rise to interactions that are long-range and very complicated in the sequence picture.

Let us therefore, as a representative model with long-range interaction, consider the mean-field Hamiltonian

$$\mathcal{H} \;=\; \mu\sum_{k=1}^{N}\sigma_k^x + \alpha\sum_{k=1}^{N}\sigma_k^z + \frac{\gamma}{2N}\sum_{k,\ell=1}^{N}\sigma_k^z\sigma_\ell^z\,. \qquad (17)$$

In the context of evolution, this is a realistic example of fitness landscapes that are invariant under permutation of sites, which are also relevant to the multi-locus theory of population genetics (for review, see [21]). For $\gamma < 0$ ($\gamma > 0$), the fitness landscape is concave (convex). The case of $\gamma < 0$ is the quadratic optimum model extensively used in quantitative genetics (cf. [22]).

It is worth emphasizing that, in contrast to the situation in physics, where interaction is inherently local and a mean-field Hamiltonian is an approximation, Eq. (17) is a *model* as such in the evolutionary context. Thus we may directly use mean-field theory (which is rigorous [23,24] in the thermodynamic limit) to reduce the desired solution of a pure phase equilibrium state to that of a one-site Hamiltonian in a self-consistent field:

$$\mathfrak{h}(\rho) \;=\; \mu\sigma^x + \alpha\sigma^z + \gamma\, m\sigma^z, \qquad (18)$$

where $m = \mathrm{tr}(\rho\,\sigma^z)$. (The spectral problem of $\mathcal{H} + \frac{N\gamma}{2}m^2\mathbb{1}$ is solved through tensor products of these.) The statistical operator ρ with minimal free energy density fulfils the self-consistency equation for the inverse temperature β:

$$\rho \;=\; \frac{\exp(\beta\mathfrak{h}(\rho))}{\mathrm{tr}(\exp(\beta\mathfrak{h}(\rho)))}\,. \qquad (19)$$

We need $\beta \to \infty$ here, where ρ boils down to a projector onto the PF eigenvector of $\mathfrak{h}$. (Let us remark that we let $\beta \to \infty$ *prior* to the thermodynamic limit.) For $\gamma > 0$, $\alpha = 0$, one finds the explicit solution ($h = \mu/\gamma$):

$$m \;=\; \begin{cases} \pm\sqrt{1-h^2}\,, & 0 \le h < 1 \\ 0\qquad\;\;, & h \ge 1\,. \end{cases} \qquad (20)$$

From $\mathfrak{h}$, one obtains the *excitation* spectrum of $\mathcal{H}$ (and also of $\widetilde{\mathcal{H}}$). The correct ground state energy per spin of $\widetilde{\mathcal{H}}$ requires the subtraction of the (parameter-dependent!) constant term $\frac{\gamma}{2}m^2 + \mu$, which finally yields

$$w \;=\; \begin{cases} \frac{\gamma}{2}(1-h)^2\,, & 0 \le h < 1 \\ 0\qquad\;\;, & h \ge 1\,. \end{cases} \qquad (21)$$

For the surplus, one obtains for this landscape

$$u \;=\; \begin{cases} \left(\frac{1-h}{1+h}\right)^{1/2}\,, & 0 \le h < 1 \\ 0\qquad\;\;, & h \ge 1\,. \end{cases} \qquad (22)$$

As in Eq. (16), the exponent of the singular behaviour is 1/2. This should not come as a surprise since u is related to the *surface* (as opposed to bulk) magnetization (compare [7]). The curves of w and u are drawn in Figs. 1 and 2. A second-order phase transition occurs at $h = 1$. Here, the population loses its genetic structure, and the mean fitness reaches the minimum value; consequently, selection ceases to operate. For $\alpha \neq 0$, the $\mathbb{Z}_2$-symmetry is broken and no phase transition can occur; more in [24].

For $\gamma < 0$, $\alpha = 0$, both w and u vanish for all μ in the thermodynamic limit. No phase transition occurs, and, what is more, the system is unaffected by mutation (and hardly affected by it in simulations with $N > 50$).

Concluding remarks. We have demonstrated how the Ising quantum chain can be employed to obtain exact solutions of muse models. The idealized fitness landscapes treated explicitly represent frustration-free situations without, with short-range, and with extremely long-range interaction. Genetic structure fades away gradually with increasing mutation in the (unrealistic) case of independent sites, but vanishes abruptly at finite μ when sites interact, thus imposing an upper limit on the mutation rate in an evolving population. This may be conjectured to be typical of situations with two-site interactions without frustration. However, more realistic multipeaked landscapes must be treated in the future.

Acknowledgements. It is our pleasure to thank T. Gerisch for an introduction to the rigorous treatment of mean-field quantum Hamiltonians, and J. Bellissard and A. Gierer for stimulating discussions.

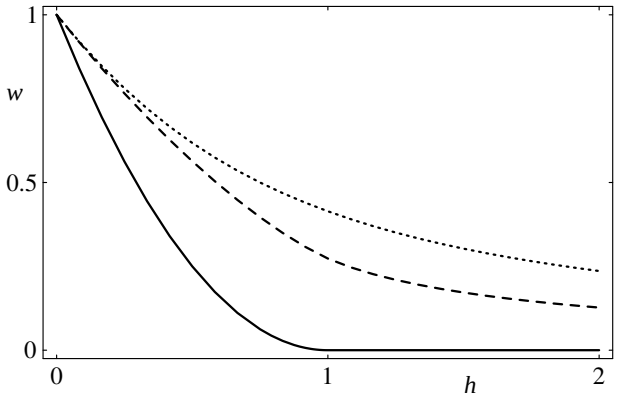

FIG. 1. Mean fitness per spin (w) as defined in Eq. (10). Dotted: Fujiyama landscape (with $\alpha_j \equiv \alpha = 1$); dashed: Onsager landscape(with $\gamma = 1$); solid line: mean-field landscape(with $\alpha = 0$, $\gamma = 2$).

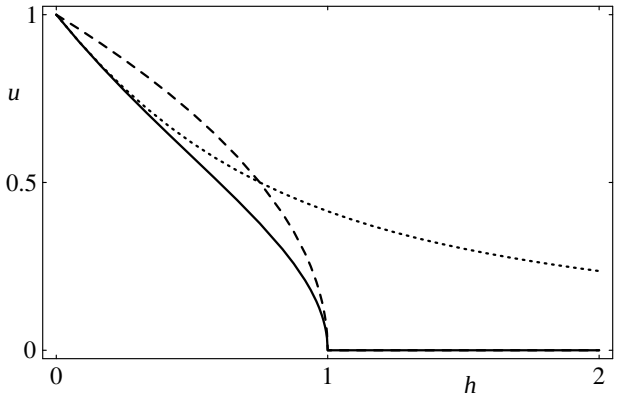

FIG. 2. Average surplus of sites with value +1 (u) as defined in (11), in the macroscopic limit. Dotted: Fujiyama landscape; dashed: Onsager landscape; solid line: mean-field landscape. Parameters as in Fig. 1.

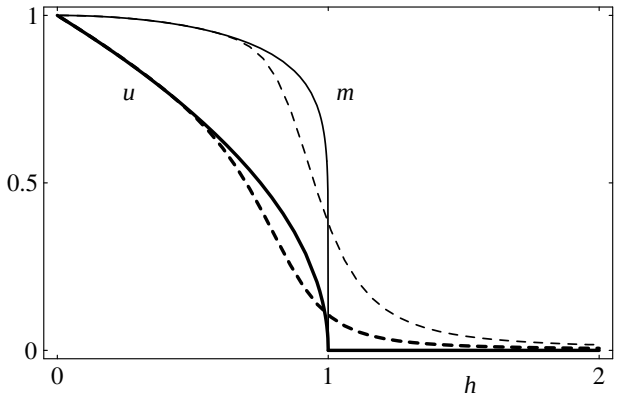

FIG. 3. Physical magnetization (m) and surplus (u) for the Onsager landscape. Solid lines: macroscopic limit; dashed: numerical finite-size calculation for $N = 8$.

[1] D. C. Mattis, in: *Physics in One Dimension*, eds. J. Bernasconi and T. Schneider (Springer, Berlin, 1981).
[2] S. A. Kauffman, *The Origin of Order. Self-Organization and Selection in Evolution* (OUP, New York, 1993).
[3] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[4] I. Leuthäusser, J. Chem. Phys. **84**, 1884 (1986).
[5] I. Leuthäusser, PhD thesis (TU Braunschweig, 1987).
[6] I. Leuthäusser, J. Stat. Phys. **48**, 343 (1987).
[7] P. Tarazona, Phys. Rev. **A45**, 6038 (1992).
[8] E. Baake, J. Biol. Syst. **3**, 343 (1995).
[9] T. Wiehe, E. Baake, and P. Schuster, J. Theor. Biol. **177**, 1 (1995).
[10] E. Baake and T. Wiehe, submitted (unpublished).
[11] M. Kimura, J. Mol. Evol. **26**, 24 (1987).
[12] T. Ohta, Proc. Natl. Acad. Sci. USA **90**, 10676 (1993).
[13] J. Hofbauer, J. Math. Biol. **23**, 41 (1985).
[14] E. Baake and H. Wagner, in preparation (unpublished).
[15] J. Hofbauer and K. Sigmund, *The Theory of Evolution and Dynamical Systems* (CUP, Cambridge, 1988).
[16] C. J. Thompson and J. L. McBride, Math. Biosci. **21**, 127 (1974).
[17] D. S. Rumschitzky, J. Math. Biol. **24**, 667 (1987).
[18] E. Lieb, T. Schultz, and D. Mattis, Ann. Phys. **16**, 407 (1961).
[19] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).
[20] P. Pfeuty, Ann. Phys. **57**, 79 (1970).
[21] N. H. Barton and M. Turelli, Annu. Rev. Genet. **23**, 337 (1989).
[22] R. Bürger and J. Hofbauer, J. Math. Biol. **32**, 193 (1994).
[23] T. Gerisch and A. Rieckers, Z. Naturf. **45a**, 931 (1990); G. A. Raggio and R. F. Werner, Helv. Phys. Acta **62**, 980 (1989).
[24] E. Baake, M. Baake, T. Gerisch, and H. Wagner, in preparation (unpublished).