# Evolutionary strategies of optimization

Torsten Asselmeyer, Werner Ebeling, and Helge Rosé
*Institute of Physics, Humboldt University Berlin, Invalidenstrasse 110, D-10115 Berlin, Germany*
(Received 16 December 1996)

Evolutionary algorithms have proved to be a powerful tool for solving complex optimization problems. The underlying physical and biological strategies can equally be described by a Schrödinger equation. The properties of the dynamics of optimization are encoded in the spectrum of the Hamiltonian. Analytic solutions and convergence velocity of the dynamics are calculated and compared with simulations of the corresponding algorithms. The connection between physical and biological strategies is analyzed. Mixing both strategies creates a basic class of evolutionary algorithms improving robustness and velocity of optimization. [S1063-651X(97)00407-8]

PACS number(s): 05.40.+j, 05.90.+m, 03.65.Db

## I. INTRODUCTION

The idea of solving optimization problems by means of evolutionary principles is a noteworthy example demonstrating the fruitful confluence and interaction of various fields of science. Optimization is not only an important technological question, it is a fundamental principle governing the dynamical laws of physics and the processes of biological evolution. Knowledge of natural systems becomes a valuable source of inspiration for constructing and investigating algorithms solving very complex optimization problems at a new technological level.

The past decade, in particular, was marked by a great development of optimization algorithms that may be collectively referred to as *evolutionary algorithms*, which consist of simulated annealing [1], evolution strategies [2], genetic algorithms [3], evolutionary programming [4], and genetic programming [5].

The rich variety of representatives seems to suggest that there is a long way to go to unification, but in fact, all of these algorithms are characterized by two general dynamic properties. The dynamics of an evolutionary algorithm may be regarded by a motion $x(t)$ in a search space $X$ visiting all global optima of a function $F(x)$ on $X$, and we can realize that the motion $x(t)$ is not deterministic: there is a stochastic generation of *alternatives* of further motion. The realization of one of them takes place by an *assessment* related to $F(x)$. All evolutionary algorithms realize the processes of generation and assessment of alternatives, i.e., *mutation* and *selection*. The stochastic nature of these algorithms requires a probabilistic description.

The complete description of a stochastic system is given by the probability distribution $P(x,t)$ of its state $x$ at a time $t$, characterizing the density of searchers of an ensemble, or the probability to find one of them at $(x,t)$. The dynamics of $P(x,t)$ is considered to be an *optimization* if any (or nearly any) initial distribution $P(x,0)$ converges to a stationary distribution $P_0(x)=\lim_{t\to\infty}P(x,t)$ that is concentrated around the global optimum of $F(x)$.

The evolutionary algorithms may be divided according to their physical or biological origin into two basic classes. Simulated annealing algorithms are motivated by thermodynamic systems minimizing the free energy and the stationary distribution $P_0(x)$ is the Boltzmann distribution. Therefore we will call the strategy representing this class of algorithms *Boltzmann strategy* [6]. Nearly all other evolutionary algorithms use a selection scheme adopted from the natural selection of biological systems. The class of algorithms using Darwinian selection will be referred to as *Darwin strategy* [6].

In this article, an optimization problem shall be given by the question of determination of optima (minima and maxima) of a real-valued function $F:X\to\mathbb{R}$, also called *fitness*. The fitness function is a quality measure quantifying the goal of a given problem (e.g., optimizing the costs of building a network). Complex optimization problems have a high-dimensional search space $X$ (e.g., the space of a 24 points network consists of $10^{84}$ graphs, more than the guessed number of particles in the Universe), making the application of classical optimization techniques infeasible. $X$ may be discrete or continuous, but for the analytic description we assume $X=\mathbb{R}^d$ and a continuous time. Then the dynamics of an evolutionary strategy can be described by a differential equation.

At first sight there seems to be no connection between the Boltzmann and Darwin strategy. Nevertheless, the fact is astonishing that both strategies can equally be described by a *Schrödinger equation* [7]. We will show that the spectrum of the Hamiltonian given by the optimization problem determines the properties of the evolutionary strategy. For unimodal fitness landscapes we are able to calculate exact solutions to the optimization dynamics. To compare the results of the strategies with evolutionary algorithms we construct algorithms that nearly exactly realize the dynamics of the Boltzmann and Darwin strategies.

The unified description of both strategies by a Schrödinger equation leads to the fact that the Boltzmann strategy becomes a Darwinian one with a transformed fitness function [8]. The properties of both strategies complement each other on the same fitness landscape. Mixing the strategies generates a different class of Evolutionary Algorithms, the *mixed strategy*, which shows a remarkable improvement in robustness and velocity of optimization [6–9].

## II. BOLTZMANN STRATEGY

The first and simplest *evolutionary strategy* is known from literature as ''simulated annealing.'' This strategy was

 

first discussed in the form of the Metropolis algorithm [10]. Following Kirkpatrick *et al.* in [1], we are interested in the analogy between equilibrium statistical mechanics and this algorithm. If we simplify our investigation to the case of a fixed temperature then, according to [8], the dynamics is given by the Fokker-Planck equation

$$\frac{\partial}{\partial t}P(\vec{x},t)=\nabla D\cdot[\nabla P+\beta\nabla FP]=D\triangle P+D\nabla(\beta P\nabla F), \tag{1}$$

where $D$ is the ''diffusion'' constant, $\beta$ is the reciprocal temperature, and $\vec{x}$ is the state vector.

The ansatz

$$P(\vec{x},t)=\exp\left[-\frac{\beta F(x)}{2}\right]y(\vec{x},t), \tag{2}$$

taken together with the separation of the time and space variables $y(\vec{x},t)=\exp(-\epsilon t)\psi(\vec{x})$, leads to the eigenvalue equation

$$D\triangle\psi(\vec{x})-V(\vec{x})\psi(\vec{x})=-\epsilon\psi(\vec{x}), \tag{3}$$

where $\epsilon$ is the eigenvalue and

$$V(\vec{x})=\frac{\beta^2}{4}D\nabla F\cdot\nabla F-\frac{\beta}{2}D\triangle F \tag{4}$$

is the redefined fitness $V$. This equation is known from quantum mechanics as the stationary Schrödinger equation. On the assumption that the operator in Eq. (3) is bounded, we obtain a discrete spectrum leading to the solution

$$P(\vec{x},t)=\exp\left[-\frac{\beta}{2}F(\vec{x})\right]\sum_{i=0}^{\infty}c_i\psi_i(\vec{x})\exp(-\epsilon_i t). \tag{5}$$

In [8] we discussed the necessary conditions for the convergence of this sum. The expression for the equilibrium distribution

$$P_0(\vec{x})=c_0\exp[-\beta F(\vec{x})] \tag{6}$$

corresponds to the eigenvalue $\epsilon=0$. A complete discussion of this fact including the construction of the Liapunov functional is contained in [8]. Thus the equilibrium distribution is concentrated around the optimum since the minimum of $F(x)$ is related to the maximum of $P_0(x)$. In the limit $t\to\infty$ the distribution $P(\vec{x},t)$ converges to the equilibrium distribution and the strategy successfully terminates at a distribution localized around the optimum. Consequently, the Boltzmann strategy fulfills the minimal requirement to be an evolutionary strategy.

As the first calculable case we consider the unimodal function

$$F(x)=F_{min}+\frac{1}{2}\sum_{i=1}^{d}a_i(\vec{x}_i-\vec{x}_i^{(0)})^2. \tag{7}$$

For $a_i>0$ we get the simple harmonic oscillator that is solved by separation of variables. With respect to the dimension of the search space the eigenfunctions are products of Hermitian polynomials. Apart from a constant, the same result is obtained in the case $a_i<0$. A collection of formulas can be found in the Appendix of [8].

The velocity of an algorithm answers the important question of how fast the algorithm reaches a desired fitness value or finds an optimum. We define a first velocity $v^{(1)}$ on the fitness landscape and a second one $v_k^{(2)}$ in the $k$th direction of the search space. The velocities are given by the time derivative of the mean values of the fitness $F(\vec{x})$ and the vector $x_k$, respectively. With respect to Eq. (1) we obtain

$$v^{(1)}=-\frac{d}{dt}\langle F\rangle=D\beta\langle\nabla F\cdot\nabla F\rangle-D\langle\triangle F\rangle, \tag{8}$$

$$\langle F\rangle=\frac{\int P(x,t)F(x)dx}{\int P(x,t)dx},$$

and

$$v_k^{(2)}=-\frac{d}{dt}\langle x_k\rangle=D\beta\langle\nabla x_k\cdot\nabla F\rangle. \tag{9}$$

The velocity $v^{(1)}$ depends on the curvature and the gradient of the fitness. In the special case (7) we are able to explicitly calculate the velocities with $a_i>0$, i.e.,

$$v^{(1)}=\frac{c_2}{\beta c_0}\epsilon_2\exp(-\epsilon_2 t), \tag{10}$$

$$v_k^{(2)}=\frac{c_1}{\sqrt{\beta a_k}c_0}\epsilon_1\exp(-\epsilon_1 t). \tag{11}$$

It is interesting to note that only the first two eigenvalues are important to the velocities and that both velocities go to zero in the limit $t\to\infty$. The other case $a_i<0$ produces a more complicated but similar result (see [8]). For initial distribution $P(x,0)=\delta(x-x_0)$ in which all searchers are concentrated in one point $x_0$, we calculate the velocity

$$v^{(1)}=\sum_{i=1}^{d}a_iD(\beta a_i x_0^2-1)\exp(-a_i\beta D2t) \tag{12}$$

and the mean fitness $\langle F\rangle$ as well as the squared mean fitness variance $\sigma^2$,

$$\langle F\rangle=F_{min}+\sum_{i=1}^{d}\frac{1}{2\beta}(\beta a_i x_0^2-1)\exp(-a_i\beta D2t), \tag{13}$$

$$\sigma^2:=\langle F^2\rangle-\langle F\rangle^2=\frac{1}{2\,\beta^2}\sum_{i=1}^{d}((2x_0^2\beta a_i-1)\times(e^{-2Da_i\beta t}-e^{-4Da_i\beta t})+1-e^{-2Da_i\beta t}), \tag{14}$$

which later on will be compared with simulations in Sec. IVA.

## III. DARWIN STRATEGY

In the following we will discuss a biological model of evolution known as the Fisher-Eigen model. To illustrate this

model we switch to the chemical reaction picture. A searcher having a fixed fitness can be regarded as a species $A$ and another searcher having a different fitness as species $C$. Thus the process of mutation is the reaction $A \to C$ described by a diffusion term. Then the reproduction process is given by $A \xrightarrow{F} 2A$ with the fitness $F$ of $A$ as a transition rate. The selection pressure is realized by the demand that the number of individuals in the population be constant. The reproduction and selection term was first introduced by Fisher [11] and Eigen [12] to become known as the Fisher-Eigen equation. If we add diffusion to the Fisher-Eigen equation and take into account the minimization of the fitness function then the Darwin strategy is defined by

$$\frac{\partial}{\partial t}P(\vec{x},t)=[\langle F\rangle - F(\vec{x})]P(\vec{x},t)+D\triangle P(\vec{x},t). \tag{15}$$

By using the ansatz

$$P(\vec{x},t)=\exp\left[\int_0^t \langle F\rangle(t')dt'\right]y(\vec{x},t) \tag{16}$$

we obtain

$$\frac{\partial}{\partial t}y(\vec{x},t)=-Hy(\vec{x},t)=[D\triangle - F(\vec{x})]y(\vec{x},t). \tag{17}$$

The separation of time and space variables $y(\vec{x},t)=\psi(\vec{x})\exp(-\epsilon t)$ reduces Eq. (17) to the stationary Schrödinger equation

$$-\epsilon_i\psi_i(\vec{x})=D\triangle\psi_i(\vec{x})-F(\vec{x})\psi_i(\vec{x}), \tag{18}$$

where $\epsilon_i$ are the eigenvalues and $\psi_i(\vec{x})$ are the eigenfunctions leading to the complete solution

$$y(\vec{x},t)=\sum_i a_i e^{-\epsilon_i t}\psi_i(\vec{x}). \tag{19}$$

The eigenvalue $\epsilon_0$ of the Darwin strategy admits a nonzero value in contrast to the Boltzmann strategy. However, with respect to the normalization of the solution we obtain only relative eigenvalues $\epsilon_i-\epsilon_0$, reflecting the fact that a shift of the fitness function should not influence the dynamics. For the quadratic fitness (7) having $a_i>0$ the problem is exactly solvable for any dimension $d$ and the solution is very similar to the Boltzmann strategy. For $a_i<0$ we obtain a different problem known from scattering theory. Now we have to deal with the problem that the operator $H$ is unbounded and thus the spectrum is continuous. We can repair this defect by restricting the search space to a compact subspace. In practice the search space is always bounded and we can choose a natural compact subspace by the demand that the fitness function (7) admits only positive values. This leads to the restriction of the search space to the interval $[-b,b]$ in each direction, where $b=\sqrt{2F_{max}/|a_i|}$ fulfills the positivity condition. Further, we have to introduce the boundary conditions for the operator $H$ defined on this compact search space. The most natural choice is to let the solution vanish on the boundary. Due to the boundary conditions an additional restriction appears and the spectrum of the operator $H$ is discrete.

As in Sec. II we are interested in the velocities of the strategy. In practice we need only the first velocity $v^{(1)}$, which, with respect to the Darwin strategy (15), is given by

$$v^{(1)}=-\frac{d}{dt}\langle F\rangle=\langle F^2\rangle-\langle F\rangle^2-D\langle\triangle F\rangle. \tag{20}$$

The use of an ensemble of searchers has the great advantages that the velocity $v^{(1)}$ reflects the main properties of the strategy and this velocity can easily be calculated during the search process. Furthermore the knowledge of the magnitude $v^{(1)}$ can be used to optimize the search parameters. In the first approximation we may suppose the velocity to be constant $v^{(1)}=v$. For the $D$ parameter we get

$$D=\frac{\langle F^2\rangle-\langle F\rangle^2}{\langle\triangle F\rangle}-\frac{v}{\langle\triangle F\rangle}.$$

Hence a value of $D$ below the critical value

$$D_{crit}=\frac{\langle F^2\rangle-\langle F\rangle^2}{\langle\triangle F\rangle}$$

still guarantees progress. Now we have to discuss the concrete example of a quadratic fitness function (7). Using the general solution (19) together with the condition $a_i>0$, we obtain the expansion

$$v^{(1)}=\left(\sum_{i=1}^{d}\sqrt{\frac{Da_i}{2}}\right)10\frac{c_2}{c_0}\epsilon_2 e^{-(\epsilon_2-\epsilon_0)t}+O(e^{-(\epsilon_4-\epsilon_2)t}). \tag{21}$$

On the basis of this velocity, the Darwin strategy can be compared with the Boltzmann strategy [8].

For the initial distribution $P(x,0)=\delta(x-x_0)$ we obtain simple explicit results for the velocity

$$v^{(1)}=\sum_{i=1}^{d}\sqrt{Da_i}\times\frac{\sqrt{2}a_i(x_0)_i^2\sinh(t\sqrt{2Da_i})-\sqrt{a_iD}\cosh(t\sqrt{2Da_i})}{\cosh^3(t\sqrt{2Da_i})} \tag{22}$$

and for the expectation value of the fitness (7)

$$\langle F\rangle=F_{min}+\sum_{i=1}^{d}\left(\frac{1}{2}\sqrt{2a_iD}\tanh(t\sqrt{2a_iD})-\frac{a_i(x_0)_i^2}{2\cosh^2(t\sqrt{2a_iD})}\right). \tag{23}$$

The comparison with the numerical results will be discussed below.

## IV. SIMULATION OF THE STRATEGIES USING STOCHASTIC ALGORITHMS

### A. Description of the Boltzmann strategy by Langevin equations

We shall begin with the simulation of the Boltzmann strategy (1) by using two different approaches. At first we consider the transition probability

$$p_{xy}=\begin{cases}1 & \text{if } F(y)>F(x)\\ \exp\left(-\dfrac{F(y)-F(x)}{T}\right) & \text{otherwise,}\end{cases} \tag{24}$$

where $T$ is the temperature. By means of the standard theory of master equations [13] we deduce from this expression the dynamics of the Boltzmann strategy (1).

Alternatively, we can reproduce Eq. (1) by the strong coupling between the Fokker-Planck and Langevin equation. For that purpose we simulate Brownian particles in the overdamped case driven by the gradient of the fitness function $F(\vec{x})$ and the mutation is given by Gaussian white noise. All assumptions put together lead to the Langevin equation

$$\frac{d}{dt}\vec{x}=-D\beta\nabla F+\sqrt{2D}\xi(t), \tag{25}$$

where $\xi(t)$ is the noise, $D$ is the diffusion constant, and $\beta$ is the inverse of the temperature. The conditions on the noise are expressed by

$$\langle\xi(t)\rangle=0,$$

$$\langle\xi(t')\xi(t)\rangle=\delta(t'-t).$$

According to [13], the behavior of Eq. (25) can be described by the Fokker-Planck equation (1) in an equivalent way.

We prefer the second ansatz to realize the Boltzmann strategy (1) in the simulation. This simulation of the Langevin equation is a well-known problem in stochastics and we will use the discretization scheme

$$x(t+1)=x(t)+\sqrt{2D\Delta t}\xi(t) -D\beta\frac{F(x(t)+\Delta x)-F(x(t)-\Delta x)}{2\Delta x}\Delta t, \tag{26}$$

where $\Delta t$ is the time step and $\Delta x$ is the box length.

Furthermore, we have to compare the dynamics of this algorithm with the analytical solution to the Boltzmann strategy (1). Two important classes of fitness landscapes are given by uni- and multimodal functions. Thus we consider the parabola $F(x)=x^2$ with one minimum at $x=0$ and the double well $F(x)=0.16+0.0278x^2(0.166\,67x-0.9)\times(0.166\,67x+0.8)$ with two minima at $x\approx-3.4,3.8$. One of the most interesting pieces of information is the time evolution of the probability distribution pictured in Fig. 1 for minimization of the parabola. This figure shows the distribution of the population $N=10\,000$ at different times. Figure 1(a) visualizes the behavior at the temperature $T=0.1$ ($\beta=10$), whereas 1(b) represents the distribution for $T=1$. According to Secs. II and III, the investigation of the mean fitness is key information of the dynamics in order to understand the convergence process to the optimum. Thus we compare the theoretical results and the simulations of the mean fitness $\langle F\rangle$ and the mean fitness variance $\langle F^2\rangle-\langle F\rangle^2$ as shown in Fig. 2. This nearly exact agreement between simulations and theory

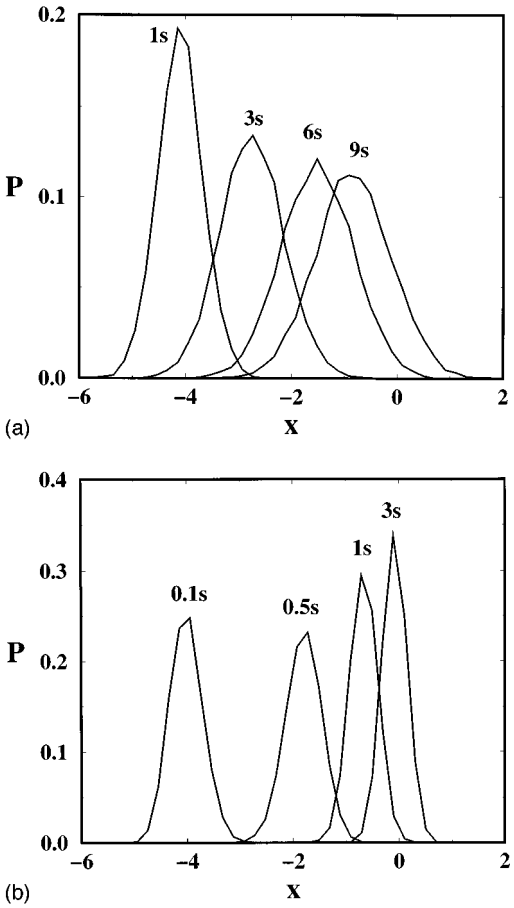


FIG. 1. Time evolution of the distribution $P(x,t)$ of $N=10\,000$ Brownian particles simulating the Boltzmann strategy. Fitness function: parabola, $\Delta t=0.01$, $\Delta x=0.01$, $D=0.1$, and (a) $\beta=10$ and (b) $\beta=1$.

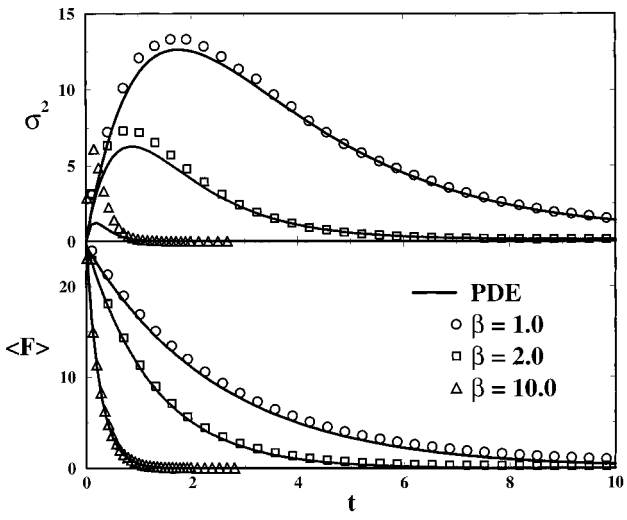


FIG. 2. Theoretical (PDE) and simulated time evolution of mean fitness and variance of the Boltzmann strategy. Fitness function: parabola, $\beta=1,2,10$, and other parameters as in Fig. 1.

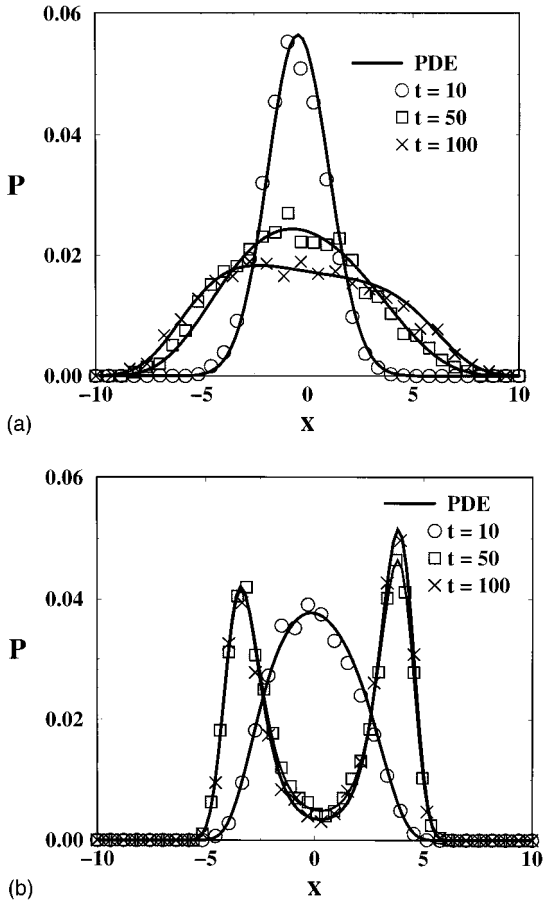


FIG. 3. Time evolution of the distribution $P(x,t)$ of $N=10\,000$ Brownian particles simulating the Boltzmann strategy compared with the theoretical solutions (PDE). Fitness function: double well, $\Delta t=0.1$, $\Delta x=0.01$, $D=0.1$, and (a) $\beta=1$ and (b) $\beta=20$.

can easily be deduced from Eqs. (13) and (14) using the parameters $a=2, D=0.1, x_0=-5$, and the corresponding $\beta$ values. Only for low temperatures do we obtain an aberration for the mean fitness variance explained by the low population size.

The next more complicated type of a fitness landscape is the double well having two minima separated by a maximum. We use the same parameters as before and start the ensemble at the maximum. To compare the simulation with the theoretical solutions we integrate the partial differential equation (PDE) (1) numerically. Figure 3 shows good agreement between the simulated and theoretical distribution $P(x,t)$ for different temperatures. For high temperatures ($\beta=1$) the searchers are spread out without localization around the minima, which leads to an increasing mean fitness (Fig. 4). Low temperatures ($\beta=20$) force the concentration of searchers to the nearest minima (Fig. 3) and the mean fitness decreases monotonically to a stationary value.

This temperature dependence of the behavior of localization in finite times leads to the fact that the Boltzmann strategy carries out an efficient search at low temperatures only. The tendency to get trapped in local minima increases by cooling. The simulated annealing approach addresses this problem by means of a cooling schedule $T(t)$, which carefully decreases temperature. But the usefulness of a certain schedule strongly depends on the concrete form of the fitness landscape [14]. Another way to reduce sticking in local optima is offered by mixing strategies (Sec. V) with the advantage that there is no need to introduce exogen parts of the dynamics such as a cooling schedule.

### B. Description of the Darwin strategy by reaction-diffusion algorithms

In this section we simulate the Darwin strategy (15) using a stochastic algorithm to compare the theoretical model with an algorithm accessible for application. Similar to the discussion at the beginning of Sec. III we introduce species $x$ and $y$ having fitness values $F(x)$ and $F(y)$, respectively. To realize the reproduction and selection processes we consider the transitions

$$(x,y) \overset{|f(x)|}{\rightarrow} \begin{cases} (x,x), & f(x)>0 \\ (y,y), & f(x)<0, \end{cases} \tag{27}$$

with the fitness proportional rate

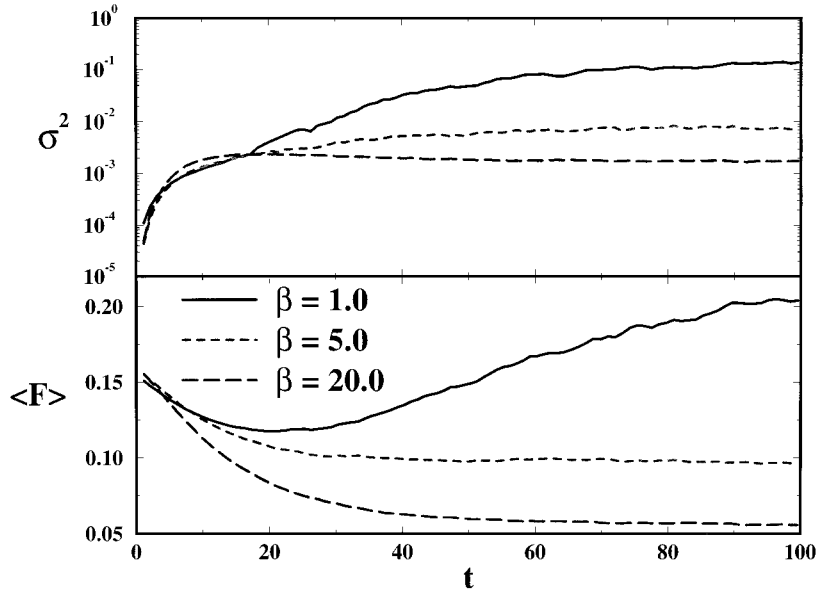


FIG. 4. Simulated time evolution of the mean fitness and variance of the Boltzmann strategy. Fitness function: double well, $\beta=1,5,20$, and other parameter as in Fig. 3.

$$f(x) := \langle F \rangle - F(x), \tag{28}$$

leading to the obvious property of population size conservation. The solution to Eq. (15) is a distribution of searchers over the continuous search space. To compare the algorithm and the partial differential equation (15) we have to consider the dynamics of the density $P(x)$ of the species $x$ with respect to the transitions (27) described by the master equation. In the following we consider only reactions of a pair of species, which is also used in the computer algorithm later on. This two-particle picture of the selection and reproduction processes was suggested by Schimansky-Geier [15] and we thank him for many discussions to clarify the questions. To establish the master equation we have to divide the transitions (27) into two parts, i.e., loss $W^-$ and gain $W^+$. We express the transition rates $W^-$ and $W^+$ by

$$\begin{aligned} W^+(x,\acute{x};y) &:= W((\acute{x},y) \to (x,y)), \\ W^-(\acute{x},x;y) &:= W((x,y) \to (\acute{x},y)), \end{aligned} \tag{29}$$

leading to the master equation for the two species $x$ and $y$,

$$\begin{aligned} \frac{\partial}{\partial t} P(x,y,t) = \int & [W^+(x,\acute{x};y) P(\acute{x},y) \\ & - W^-(\acute{x},x;y) P(x,y)] d\acute{x} \\ & + \int [W^+(y,\acute{y};x) P(x,\acute{y}) \\ & - W^-(\acute{y},y;x) P(x,y)] d\acute{y}. \end{aligned} \tag{30}$$

Now we have to determine the transition rates with respect to the processes of reproduction and selection. For that purpose we introduce intermediate states denoted by $\acute{x}$. By integration over these states we can deduce the correct processes. For the rates we assume

$$\begin{aligned} W^+(x,\acute{x};y) &:= W^+((\acute{x},y) \to (x,y)) \\ &= \delta(x-y)[|f(\acute{x})| \Theta(-f(\acute{x})) + f(y) \Theta(f(y))], \end{aligned} \tag{31}$$

$$\begin{aligned} W^-(\acute{x},x;y) &:= W^-((x,y) \to (\acute{x},y)) \\ &= \delta(\acute{x}-y)[|f(x)| \Theta(-f(x)) + f(y) \Theta(f(y))], \end{aligned}$$

where the other rates can be derived by the interchange of $x$ and $y$. The complete derivation of the Fisher-Eigen equation

$$\frac{\partial}{\partial t} P(x,t) = [\langle F \rangle - F(x)] P(x), \tag{32}$$

with respect to the master equation (30), can be found in the Appendix. By adding a symmetric transition rate $W_D$ with

$$W_D(x,y) = W_D(y,x) = W_D(y;r), \quad r = x - y$$

to the loss and gain terms, we obtain the mutation process leading to the Laplacian term in Eq. (15). According to [13] (p. 214), the master equation

$$\begin{aligned} \frac{\partial}{\partial t} P(x,t) = & \int W_D(x-r;r) P(x-r,t) dr \\ & - P(x,t) \int W_D(x;-r) dr \end{aligned} \tag{33}$$

describes a process with symmetric transition rates. Assuming that only small jumps occur and that the solution $P(x,t)$ also varies slowly with $x$, it is possible to deal with the shift from $x$ to $x-r$ in the first integral in Eq. (33) by means of a Taylor expansion up to second order where the term with the first derivative is neglected by the symmetry of $W_D$. If we set

$$2D = \int r^2 W_D(x;r) dr = \text{const}$$

then the diffusion equation is obtained [13]. With respect to the assumption of sufficiently small steps $r$ we have unified the mutation and the selection process in one equation. But we note that the requirement of small steps $r$ for the mutation process leads to a different realization of mutation and selection in the algorithm. Furthermore, it is to be noted that the diagonal term $W(x,x)$ of the rates can be chosen arbitrarily.

The simulation of the master equation is a well-known problem in the theory of stochastic processes. We use the waiting time distribution of Eq. (30) to establish a stable and effective algorithm of the evolutionary process [16–18]. The main idea of the algorithm is very simple. The dynamics of the process can be split into two parts. At first the population remains unchanged for a certain time, the waiting time $\tau$, at the current state. After $\tau$ time steps the population is turned into a new state. If we know the distribution of the waiting time $\tau$, we only need to simulate the effective changes of the system representing the master equation. Indeed, this is exactly the method of evolutionary algorithms to execute the evolutionary process. But the common simulation scheme of evolutionary algorithms pays no attention to the time scale of the process, resulting in the inability to compare the dynamical behavior of the evolutionary algorithm with analytical results of the Darwin strategy.

In fact, it is easy to extend the common simulation scheme to an algorithm respecting the real time scale of the evolutionary process. Solving the *first passage time* problem of Eq. (30) with initial and boundary conditions

$$P(x,0) = \delta_{x\acute{x}}, \quad P(x,t) = 0, \ x \neq \acute{x}, \tag{34}$$

we obtain the waiting time distribution [19,20]

$$p(\tau) = \frac{\partial}{\partial \tau}[1 - P(x,\tau)] = \frac{1}{\langle \tau \rangle} \exp\left( - \frac{\tau}{\langle \tau \rangle} \right). \tag{35}$$

The waiting time between two changes is exponential distributed and $\langle \tau \rangle$ is the mean waiting time.

The mutation term of the Darwin strategy describes a diffusion process. To simulate such a process we have to guarantee that one mutation step $r$ is small in comparison with the size of $P(x,t) > 0$. The mean value of a diffusion step is $\langle r \rangle = \sqrt{2 D \Delta t}$; thus we must choose the maximal time scale $\Delta t_{max} > \Delta t$ of the mutation step in such a manner that $r$ remains sufficiently small. Diffusion is a continuous process,

whereas selection is discrete. Thus we execute in every step a mutation and selection having the rate $\langle F\rangle - F(x)$. To ensure that the mutation step $r$ is small we set the mean waiting time between two steps of algorithm $\langle\tau\rangle < \Delta t_{max}$. The stochastic dynamics of diffusion is given by a Gaussian mutation

$$x(t') = x(t) + \sqrt{2\,D(t-t')}\,\xi, \tag{36}$$

with the normal distributed random accession $\xi$. The selection process can be modeled by a birth-death process. A randomly chosen individual $x$ dies if the rate $W_s = \langle F\rangle - F(x)$ is negative and will be reduplicated if $W_s$ is positive [21]. To keep the population size $N$ constant another randomly chosen opponent $y$ must be reduplicated or die. This process can be considered as tournament selection having the rate $|W_s|$.

The selection step will be executed if a uniform random number is

$$z < W_s, \quad z \in \left(0, \frac{1}{\langle\tau\rangle}\right),$$

with a mean waiting time $\langle\tau\rangle$ fulfilling

$$\langle\tau\rangle < \min\left(\frac{1}{\max(W_s)}, \Delta t_{max}\right).$$

The choice of the fixed value $\max(W_s)$ instead of $W_s$ synchronizes both time scales $1/\max(W_s)$ and $\Delta t_{max}$. Such a choice is equivalent to the introduction of diagonal terms into the rates that leave the master equation unchanged. Let us summarize one generation of the algorithm by the following scheme using uniform random numbers $z$, the exponentially distributed number $\tau$, and the normally distributed number $\xi$ for $N$ steps:

$$t' = t + \frac{\langle\tau\rangle}{N}\tau,$$

$$i = z_1 \in \{1, \ldots, N\},$$

$$x(t') = x(t) + \sqrt{2\,D(t-t')}\,\xi,$$

$$z_2 \in \left(0, \frac{1}{\langle\tau\rangle}\right) \quad \text{if } z_2 < |W_s = F(x) - \langle F\rangle|,$$

$$j = z_3 \in \{1, \ldots, i-1, i+1, \ldots, N\},$$

$$y = x, \quad W_s > 0$$

$$x = y, \quad W_s < 0.$$

Now we compare the dynamics of this algorithm with the analytical solution to the Darwin strategy (15) for two different fitness functions: the parabola $f(x) = x^2$ and the double well $f(x) = 0.16 + 0.0278x^2(0.166\,67x - 0.9)(0.166\,67x + 0.8)$, which are equivalent to the case of the Boltzmann strategy. Of course we use the same characterization of the strategy explained in Sec. IV A. The time evolution of the probability distribution of the population is represented in Fig. 5. The exact solution to the Darwin strategy (PDE) is indicated by the solid line. The circles, squares, and crosses denote the density of the population ($N = 10\,000$) at different times. Without any doubt, the evolutionary algorithm realizes nearly exactly the dynamics of the Darwin strategy. Also, Fig. 6 emphasizes this fact for the time dependence of the mean fitness of the population. For small population sizes ($N < 1000$) the stochastic fluctuations are still visible and go to zero for larger populations. The Darwin strategy describes the dynamics of the population without taking into account fluctuations, but we can see that a population of $N = 10$ individuals shows already a behavior quite similar to the Darwin strategy in the case of a parabolic fitness function.

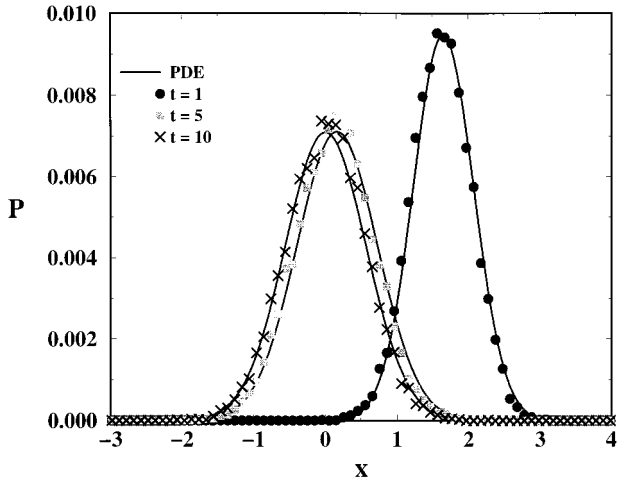


FIG. 5. Time evolution of the distribution $P(x,t)$ of $N = 10\,000$ individuals simulating the Darwin strategy compared with the theoretical solutions (PDE). Fitness function: $x^2$, $D = 0.1$.

For the multimodal fitness landscape of the double well we also found good agreement between the solution to the Darwin strategy (PDE) and the simulation of $N = 10\,000$ individuals (Fig. 7). (The solution to the PDE was found by numerical integration.) The mean fitness (Fig. 8) shows slightly more fluctuations, but the main features of the PDE and the evolutionary algorithm are still the same.

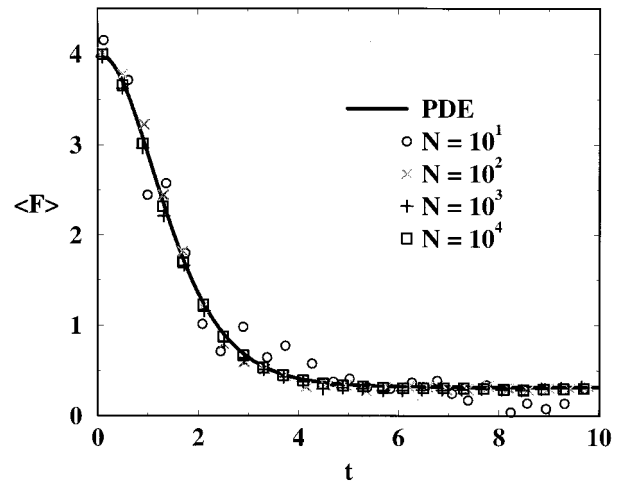


FIG. 6. Theoretical (PDE) and simulated time evolution of the mean fitness of the Darwin strategy. Fitness function: $x^2$.

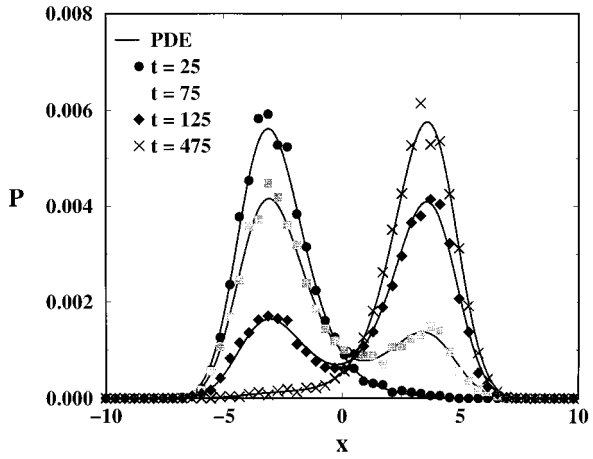


FIG. 7. Time evolution of the distribution $P(x,t)$ of $N=10\,000$ individuals simulating the Darwin strategy compared with the theoretical solutions (PDE). Fitness function: double well, $D=0.1$.

The parabola and double well represent the main features of a general fitness landscape: the existence of local optima and the transition between two optima. For both cases we could find good agreement between the dynamics of the Darwin strategy and its stochastic realization: the evolutionary algorithm. Within the range of this agreement we are able to say that the analytical results found for the Darwin strategy are valid for the evolutionary algorithm too. In this sense the Darwin strategy becomes an important tool for the analytical investigation of evolutionary algorithms.

## V. MIXING OF STRATEGIES

In our paper [8] we analyzed the Boltzmann strategy as well as the Darwin strategy in detail in order to show that with respect to the velocities, the strategies have oppositional behavior. Because of their oppositional advantages and disadvantages it seems desirable to mix them. Furthermore, the dynamic equations defining both strategies contain a common term $D\triangle P$. We define the dynamics of a mixed strategy by

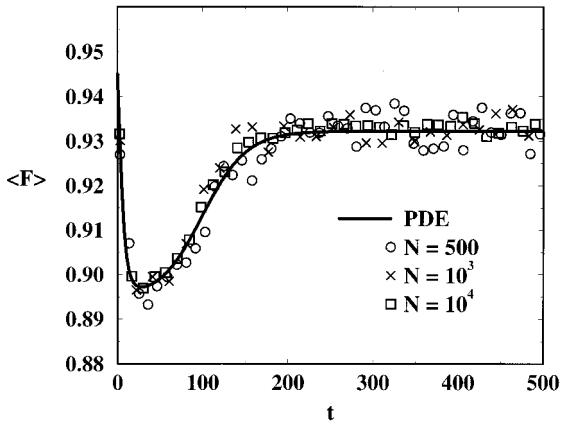


FIG. 8. Theoretical (PDE) and simulated time evolution of the mean fitness of the Darwin strategy. Fitness function: double well.

$$\frac{\partial}{\partial t}P(\vec{x},t)=(D_1+D_2)\triangle P(\vec{x},t)+\beta D_2\nabla(P\nabla F) \quad (37)$$

$$+\gamma[\langle F\rangle-F(\vec{x})]P(\vec{x},t). \quad (38)$$

For $\gamma=0$ this dynamics reduces to a pure Boltzmann strategy and for $\beta=0$ we obtain a Darwin strategy. In the case of $D_2\beta=1$ and $D_1+D_2=D$ the corresponding equation can be interpreted as a mixing of a gradient strategy and a Darwin strategy. We found in the simulations that this mixing is an effective and fast algorithm. In the following we simplify the equation by the settings $D_2=D$ and $D_1=0$. The mixed case may be treated by means of the ansatz

$$P(\vec{x},t)=\exp\left[\gamma\int_0^t\langle F\rangle dt'-\frac{1}{2}\beta F(\vec{x})\right]y(\vec{x},t), \quad (39)$$

leading to

$$\frac{d}{dt}y(\vec{x},t)=[D\triangle-W(\vec{x};\beta,\gamma)]y(\vec{x},t), \quad (40)$$

$$W(\vec{x};\beta,\gamma)=\gamma F-\frac{D\beta}{2}\triangle F+\frac{D\beta^2}{4}(\nabla F)\cdot(\nabla F),$$

$$W(\vec{x};\beta=0,\gamma=1)=F(\vec{x}),$$

$$W(\vec{x};\beta,\gamma=0)=V(\vec{x}),$$

where the explicit solution is given by the same expression (19) with the eigenfunctions of the problem

$$0=D\triangle\psi_i(\vec{x})+[\epsilon_i-W(\vec{x};\beta,\gamma)]\psi_i(\vec{x}). \quad (41)$$

The linearity of the differential equation leads to simple relations between the solutions and velocities. For example, the velocity for the Boltzmann strategy can be added to the velocity of the Darwin strategy with respect to the constant $\gamma$ to obtain the velocity of the mixed strategy

$$v^{(1)}=-\frac{d}{dt}\langle F\rangle=\gamma(\langle F^2\rangle-\langle F\rangle^2)+D\beta\langle\nabla F\cdot\nabla F\rangle-D\langle\triangle F\rangle. \quad (42)$$

In this way the critical value of $D$ is given by

$$D_{crit}=\frac{\gamma(\langle F^2\rangle-\langle F\rangle^2)}{\langle\triangle F\rangle-\beta\langle\nabla F\cdot\nabla F\rangle},$$

which still guarantees progress in the mean for $D$ values below this critical value $D_{crit}$. Now the magnitude $D_{crit}$ can be used to control the strategy during the evolutionary process to adapt the parameters $\beta$ and $\gamma$.

The mixed strategy is a well-suited tool for multicriterion optimization. Let us assume that the fitness function consists of two contradictory parts $F(x)=F_1(x)+F_2(x)$ and that $F_1(x)$ is a more local requirement. One example of such a situation is the network optimization (see [9]), where $F_1(x)$ would be the detour and $F_2(x)$ the total length of the paths. Another example is the ''democratic'' representation introduced by Dittes [22]. A well-designed strategy may use the Boltzmann strategy for local optimization of $F_1(x)$ and the Darwin strategy for global optimization of $F(x)$. By using

the mixed strategy one is able to respect the structure of the local and global requirements of multicriterion problems.

Mixing not only improves the convergence velocity but also reduces the tendency of the Boltzmann strategy to get caught in local minima. Since the Darwin strategy is able to escape from local minima by a tunnelinglike process and the Boltzmann strategy works very well on landscapes with low curvature, the mixed strategy proceeds by an efficient search in both regions of the landscape [8]. To obtain these advantages we have to choose the coefficients $\gamma$ and $\beta$ appropriately. In general, the optimal search requires $\beta>0$ and $\gamma>0$. In summary, adding some amount of the ''complementary'' strategy is in most cases recommended. This has already been found out empirically in an earlier work [7,6].

## VI. SUMMARY AND CONCLUSION

The marvelous fact that optimization strategies adopted from physical and biological systems allow a unified description by a Schrödinger equation makes the assumption of a fundamental meaning of optimization principles appear in a different light. This remarkable relationship between optimization and the Schrödinger equation makes it feasible to study the behavior of both strategies in detail. The differences of the Boltzmann and Darwin strategy can be understood by a transformation (4) of the fitness. Hence an investigation of differences and similarities was simply realized [8]. A main characteristic of the strategy is given by the convergence velocity. For the parabolic case and special initial conditions ($\delta$ function) explicit expressions for the velocity and the mean fitness are obtained.

In Sec. II algorithms for the simulating of the stochastic processes were derived from the dynamical equations. For the Darwin strategy a master equation description was chosen. In the case of the Boltzmann strategy, a Langevin approach turned out to be more suitable. Theoretical results are nearly identical to the numerical simulation. Numeric integration of the PDE was necessary for the double well and the results were also in accordance with the simulations. In summary, the relationship between optimization theory and quantum mechanics has been established. As an important indication of this relationship we observe the well-known result that the stationary distribution of the Darwin strategy is the ground state of a quantum-mechanical system of the fitness $F(x)$. This result, combined with the fact that the Darwin strategy can be simulated by an evolutionary algorithm, provides us also with a tool for calculating the ground states of complex quantum-mechanical systems.

## ACKNOWLEDGMENTS

This work has been supported by the Department of Education and Science (BMBF) in the frame of the EvoAlg project. We thank Professor L. Schimansky-Geier for his helpful discussions.

## APPENDIX A: DERIVATION OF THE FISHER-EIGEN EQUATION

We consider the master equation (30). If we assume transition rates related to the transitions (27) for the gain and loss (31) we obtain the selection and the reproduction term of the Darwin strategy. Together with the useful abbreviations and the one-particle distribution $P(x)$

$$A(x):=|f(x)|\Theta(-f(x)),\quad B(x):=f(x)\Theta(f(x)),$$

$$P(x)=\int P(x,y)dy, \tag{A1}$$

the master equation (30) is obtained as

$$\begin{aligned}\frac{\partial}{\partial t}P(x,y,t)=&\,\delta(x-y)\left(\int A(\acute{x})P(\acute{x},y)d\acute{x}+B(y)P(y)\right)\\&-[A(x)+B(y)]P+\delta(y-x)\\&\times\left(\int A(\acute{y})P(x,\acute{y})d\acute{y}+B(x)P(x)\right)\\&-[A(y)+B(x)]P.\end{aligned} \tag{A2}$$

In this equation the correlation between two searchers is included, whereas the one-particle representation of the Darwin strategy does not realize such a correlation. To this end we neglect the correlation and make the approximation

$$P(x,y)=P(x)P(y); \tag{A3}$$

using $2A-|f|=-f$, $2B-|f|=f$, and $\langle f\rangle=0$, we obtain the Fisher-Eigen equation

$$\frac{\partial}{\partial t}P(x,t)=(\langle 2\,A-|f|\rangle+2\,B-|f|)P(x)=fP(x). \tag{A4}$$

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[2] I. Rechenberg, *Evolutionsstrategien — Optimierung Technischer Systeme nach Prinzipien der Biologischen Information* (Frommann-Verlag, Stuttgart, 1973).

[3] J. H. Holland, *Adaptation in Natural and Artificial Sytstems* (University of Michigan Press, Ann Arbor, 1975).

[4] D. B. Fogel, *Evolutionary Computation — Toward a New Philosophy of Machine Intelligence* (IEEE, Piscataway, 1995).

[5] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection* (MIT Press, Cambridge, MA, 1992).

[6] T. Boseniuk, W. Ebeling, and A. Engel, Phys. Lett. A **125**, 307 (1987).

[7] W. Ebeling and A. Engel, Syst. Anal. Model Simul. **3**, 377 (1986).

[8] T. Asselmeyer and W. Ebeling, BioSystems **41**, 167 (1997).

[9] F. Schweitzer, W. Ebeling, H. Roŝe, and O. Weiss, in *Parallel*

*Problem Solving from Nature-PPSN IV*, edited by H.-M. Voigt, W. Ebeling, H. P. Schwetel, and I. Rechenberg (Springer, Berlin, 1996), p. 940.
[10] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[11] R. A. Fisher, *The Genetical Theory of Natural Selection* (Oxford University Press, Oxford, 1930).
[12] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[13] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
[14] P. J. M. Laarhoven and E. H. C. Aarts, *Simulated Annealing: Theory and Applications* (Reidel, Dordrecht, 1987).
[15] L. Schimansky-Geier (private communication).
[16] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[17] R. Feistel, Wiss. Z. Univ. Rostock, **26**, 663 (1977).
[18] D. T. Gillespie, J. Comput. Phys. **28**, 435 (1978).
[19] R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon & Breach, New York, 1963), Vol. 1.
[20] I. Matheson, D. F. Walls, and C. W. Gardiner, J. Stat. Phys. **12**, 21 (1975).
[21] T. Fricke, Ph.D. thesis, University Aachen, 1994.
[22] F.-M. Dittes, Phys. Rev. Lett. **76**, 4651 (1996).