Evolution of Bit Strings II:
A Simple Model of Co-Evolution

Harald Freund*
Robert Wolter†
Physics Department, University of Wuppertal,
Gauss-Straße 20, 42097 Wuppertal, Germany

Abstract. In this paper we present the results of simple co-evolutionary models that simulate the temporal development of a population of interacting bit strings. Each bit string is decoded into functional groups called "genes." The individuals participate in a procedure similar to the Darwinian principle, that is, random interaction (in which the genes determine the fitness of an individual) followed by deterministic selection including random errors during reproduction. We use three selection models: *selection by sorting and replacing, tournament selection*, and *energy flow*.

We compare the results for each model to those of a random-selection model to ensure that "emergent" (non-trivial) features are really caused by evolutionary processes. The influence of selection is clearly evident when comparing the age distributions of individuals and species to those of the random-selection case. While the age distribution of the individuals shows a systematic deviation from an exponential behavior in the selection case, the species ages tend to follow a power law when applying non-random selection. The species show a dynamic of "punctuated equilibria" that is similar to the behavior of natural systems.

1. Introduction

In a recent paper [6] we introduced a computational model that was designed to simulate an open-ended evolution (evolution without end). We defined a system consisting of a population of co-evolving interacting individuals, namely bit strings, which are subjected to an evolutionary algorithm based on the Darwinian principles of random interaction and non-random selection. Because of the interaction we have only *frequency-dependent fitness* (see [13]) in our model, that is, there is no *global* optimization toward a predefined goal.

---
*Electronic mail address: `freund@wpts0.physik.uni-wuppertal.de`
†Electronic mail address: `wolter@wpts0.physik.uni-wuppertal.de`

The current population defines whether an individual is fit, and thus selection of the currently fittest individuals takes place.

We study the temporal development of systems with different parameters and compare the results of the dynamics to those of a random-selection model. This comparison proves to be a very helpful tool in deciding which features are non-trivial or "emergent," in contrast to those that are just caused by predefined rules.

2. Short description of the model

In this section we summarize the main features of our model. For details see [6].

We choose individuals with a minimal length of 32 bits because we want to have a huge state space (of bit sequences) so that many different possibilities for evolutionary development are present. The bit strings are decoded into functional units called "genes" using a binary decoding tree that ensures the decoding of every possible bit sequence without ambiguity. There are 20 different genes with lengths of 3 to 8 bits. We implemented this second level to create a hierarchical structure in the highly dimensional state space of the possible bit sequences. We were guided by natural systems that can be described using different hierarchical levels. This second level endows the system with an inherent distinction between genotype and phenotype.

We define an interaction procedure for the individuals that takes place on the gene level. In an *elementary interaction*, two randomly chosen genes of two interacting individuals are used to change the "fitness" or score of the individuals. In each time step, each individual participates in a fixed number of elementary interactions. The fittest individuals of the current population are selected to produce offspring. The contents of an offspring is changed with a low probability using the genetic operators "mutation," "crossover," and "cut and splice." Using these basic ingredients for our model, a single time step of a simulation consists of the following sequence:

1. Interaction phase

2. Selection phase

3. Reproduction phase

2.1 Decoding

One of the results described in [6] gave strong evidence that the use of start-codons (which signal the beginning of a gene in the bit string) is almost unavoidable if a safe and flexible information storage mechanism is wanted. For all simulations presented here we used the same startcodon ('110'), which already proved useful for the binary decoding tree presented in [6].

One consequence of startcodons is the possibility of having "junk" information that is not decoded. The average fraction n_b of decoded bits per
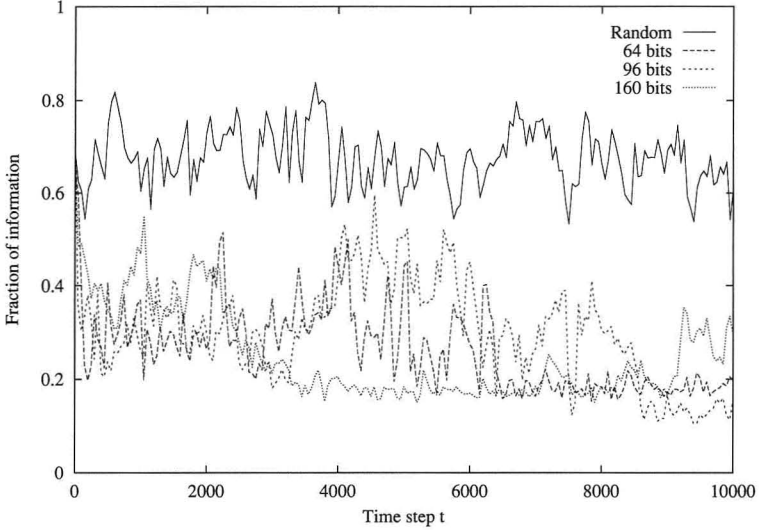
Figure 1: Temporal evolution of the average normalized information length for bit strings of three different bit lengths. For the random-selection model the average normalized information length is approximately 2/3.

individual starts at $n_b \simeq 2/3$ after a random initialization (which is approximately 10 to 14 genes in an individual of 96 bits) and decreases to $n_b \simeq 0.3$ during several hundred time steps (see Figure 1). This phenomenon is typical of all simulations presented here and can be explained as an effort by the population to escape from the strong selection pressure. An individual that contains only a few information bits very likely produces an offspring that belongs to the same species as the parent. If, on the other hand, the information length exceeds some threshold, the offspring more likely belongs to a different species that is in most cases less fit than the original one. This causes all individuals with too much (and thus too easily changed) information to be eliminated quickly, which is similar to the *information error catastrophe* [5]. We cannot really give a lower bound (as can be seen in Figure 1), but we can argue that each individual needs at least one gene. If an individual has no gene, its score is zero because it cannot interact. But if it has a "good" gene, its average score will be positive (individuals with "bad" genes disappear quickly). Therefore, an individual *must* have at least one gene to have a positive score.

2.2 Interaction

There are many different ways to define an interaction procedure (see, for example, [11, 12]). We chose a binary-interaction algorithm that contains deterministic and random parts. The deterministic part is defined by a 20×20

interaction matrix A that contains random numbers of a specified interval. Each entry $A_{g,g'}$ of this matrix represents the interaction between two genes g and g'. For example, the element $A_{5,5}$ gives the interaction strength when gene E encounters gene E. This matrix is fixed for the entire simulation, so it represents some kind of "frozen randomness." The random part consists of the choice of (a) two individuals i and k, or (b) two interacting genes in these individuals, l (from i) and m (from k). A single interaction changes the fitnesses (scores) $f(i)$ and $f(k)$ in the following way:

$$f(i) \leftarrow f(i) + A_{m,l},$$
$$f(k) \leftarrow f(k) + A_{l,m}. \tag{1}$$

Each individual of the population participates in α interactions per time step.

2.3   Selection

In the following section we describe three different selection algorithms that choose the fittest individuals of a population to produce offspring after the interaction phase.

1. *Selection by sorting and replacing.* The simplest selection procedure is to sort the individuals of the population in accordance with their fitness values, and then replicate the fittest individuals and have the offspring of the fittest replace the worst. We implemented versions that choose the best half (or the best quarter) of a population to replicate and replace the worst half (quarter). In this way a constant portion $p_e = (1/2)(1/4)$ of the population is replaced in each time step, which causes a continuous change in the population.

2. *Tournament selection.* Goldberg et al. [7] introduced an alternative method of selection that takes into account the frequency dependence of fitness. This procedure, called tournament selection, consists of the comparison (competition) of all individuals within a subgroup of the population. The population is divided into groups of equal size (at least two, but other numbers are also possible). The one with the highest fitness value survives and may replicate if it wins more than one tournament. All the individuals that do not win a single tournament are eliminated. If the tournament size is s, each individual usually participates in s tournaments.

   In the simplest version of tournament selection, the groups consist of two individuals; the better replicates and the worse is eliminated. This way individuals with a fitness below average still have a chance to "win" when competing against even worse opponents. On the other hand, a good individual can be "beaten" by a better one, although its fitness value might be above the average of the current population. Tournament selection allows two possibilities for the fraction p_e of eliminated individuals per time step:

- a constant fraction $p_e = 1/2$, as in "sorting and replacing," if only one tournament round is performed,

- an average constant fraction $p_e = 1/3$ with local fluctuations, if each individual participates in two tournaments (for more details see Appendix C).

3. *Energy flow.* We also developed another selection procedure [15] that seems more "natural" than the previous ones. Organisms usually need some renewable resources that are available in their environment and that for simplicity we call "energy." Therefore we also implemented an energy-based procedure for selection (see also [2, 14]).

One simple way to do this is to create an energy source E_s for the individuals that is filled at the beginning of each time step with a constant energy. The individuals take energy from this source during the interaction phase (as long as it contains enough energy). The interacting genes determine the amount of energy that the two individuals get from the source.

The individuals replicate when their energy levels exceed some threshold value E_{thr} and give some of the energy to their offspring. At the beginning (or end) of each time step the energy of an individual is decreased by an amount E_c called "energy costs per time step." An individual is eliminated if its energy falls below this value.

Some fluctuations enter the scenario because the previously fixed values for population size, number of replicating individuals per time step, and so forth are no longer constants. The system acts more stochastically, so one has to take great care in assigning initial conditions and boundary conditions. For example, the entire population might die during a single time step if the individuals do not save enough energy to pay their energy costs per time step. This could happen if individuals replicate too often during a single time step and give too much energy to their offspring. An easy way to avoid this behavior is to decrement the energy of an individual by E_c first, then decide whether enough energy remains to produce offspring.

In an energy-flow model the entire scenario is more sensitive to implementation errors. On the other hand, the system might have more capability for evolution than the other models because the "natural fluctuations" may create different pathways in the state space.

2.4   Reproduction

When replicating, the original individual (parent) produces one copy (offspring), which gets all values (genes, energy cost, threshold, etc.) from its parent. The age and offspring counters of the new individual are set to zero (so the individual is like a newborn, not produced by cell division). During the replication process copying errors of the genetic material appear with low probability. We implemented mutation (single-bit flip) and crossover (in a

few cases for comparison also cut and splice), because they seem to be the most significant errors that appear in natural replication processes [3, 5].

3. Randomness

Randomness and determinism are the cooperating counterparts of an evolutionary system. While the deterministic rules define the framework for evolutionary processes (they can be viewed as the combination of all influences of the fixed part of the environment), the random parts of a model simulate realistic boundary conditions like thermal fluctuations, for example. As we already mentioned in [6], the interaction matrix should enable all possible kinds of interaction between individuals. This can be implemented most easily by choosing random numbers for the matrix elements.

In natural environments randomness is caused, for example, by chemico-physical fluctuations and by the stochastic behavior of other individuals. This can be simulated well by implementing the random interaction scheme described in Section 2.2.

Thus far we have explained only the randomness used in our models. Now we introduce an additional tool called *random interaction models*. They prove very useful in distinguishing between evolutionary processes and behavior caused by the predefined rules of the system. We use the term evolutionary processes in contrast to optimization processes. In our view optimization processes have a global predefined aim, whereas evolutionary processes are driven by *local* optimization but include no *global* fitness function.

We used two different versions of a random model that showed similar results:

- In the first version of random selection the fitness is changed by adding random numbers instead of matrix elements decoded by the genes. So the random matrix is "switched off" while the other parts of the model stay the same.

- In the second version we chose random genes instead of the decoded genes for interaction, and the interaction matrix remained the same as in the non-random model.

Using random models we obtain the results of a "zero experiment" of the considered evolutionary model. This is reminiscent of the "flat fitness landscape" of Derrida and Peliti [4]. A random model can be defined easily for *every* evolutionary model and thus should be used as a *standard reference model*.

4. The concept of species

The dynamics of our system lead to successful groups of genes. Sequences with the same gene content but with different arrangements (like $ABABAA$ and $AAAABB$ (A_4B_2)) are equivalent during the interaction phase because on average they get the same score. In addition, A_4B_2 has on average the

same properties as A_2B (AAB). These groups are thus naturally considered as belonging to the same species. A species is consequently defined as the group of individuals containing the same ratio of gene numbers as described above.

Whereas the lifetime of an individual is restricted to a few time steps (depending on the parameters of the model; compare Appendix A and [6]), the lifetime of a species may be much longer (up to thousands of time steps). A necessary condition for this is that individuals produce with high probability *identical* offspring (i.e., offspring belonging to the same species). An analytical treatment of the species age behavior is possible for the random selection model; details are given in Appendix B.

5. Simulations

We simulated the temporal development of bit string populations with rather small population sizes (64 to 512, in few cases up to 10^4). In most runs we chose a constant bit length (e.g., 96 bits) per individual. In some cases where we used "cut and splice" during reproduction we allowed changes in bit length. The interaction rate was varied from $\alpha = 5$ to $\alpha = 50$. We defined an average mutation probability p_m per bit, so short bit strings had the same mutation rate as longer ones. We varied the value from $p_m = 10^{-2}$ to $p_m = 10^{-4}$ and found that $p_m = 10^{-3}$ was most useful for our purpose. This value is very close to the natural mutation probability (see [5]).

We implemented different versions of the three basic selection algorithms introduced in Section 2.3. The results are described in the following subsections.

5.1 General results

One general result valid for all implementations concerns the age distribution of individuals. As explained in Appendix A, we expect an exponential age distribution if elimination of individuals is purely a random process. This is indeed the case for random selection models, as can be seen in Figure 2 (squares), but if selection is present this distribution is deformed. Some individuals that are fitter than the others survive longer, and others produced by mutation lead to worse individuals that are quickly eliminated (see Figure 2, diamonds and crosses).

The age distribution for individuals is deformed slightly, but asymptotically it is still exponential. One reason for this is that the population is usually very homogeneous; that is, nearly all individuals have the same genetic contents, so they are equally fit and elimination is in fact a random process. Only at the beginning of a simulation and when there is genetic takeover do we have a wide range of fitness.

We obtain qualitatively different results when we look at the age distribution of species. In Appendix B we show that, for a purely random model, we also expect an asymptotically exponential age distribution. Figure 3 reflects
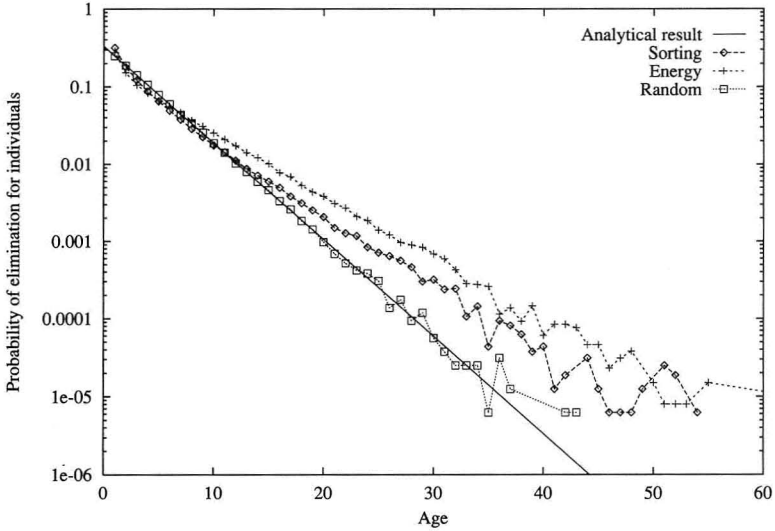
Figure 2: Comparison of the age distribution for the random model ($\square$) and two selection cases (sorting $\diamond$, energy flow $+$). The straight line corresponds to equation (2) with $p = 1/4$.
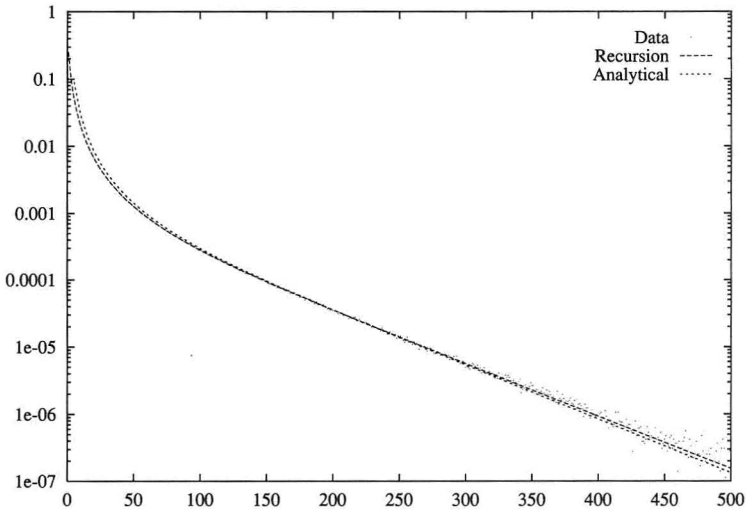

Figure 3: Comparison of the age distribution for species between a simulation (dots) and two analytical functions (see Appendix B for details). Data are well described by these exponential distributions.
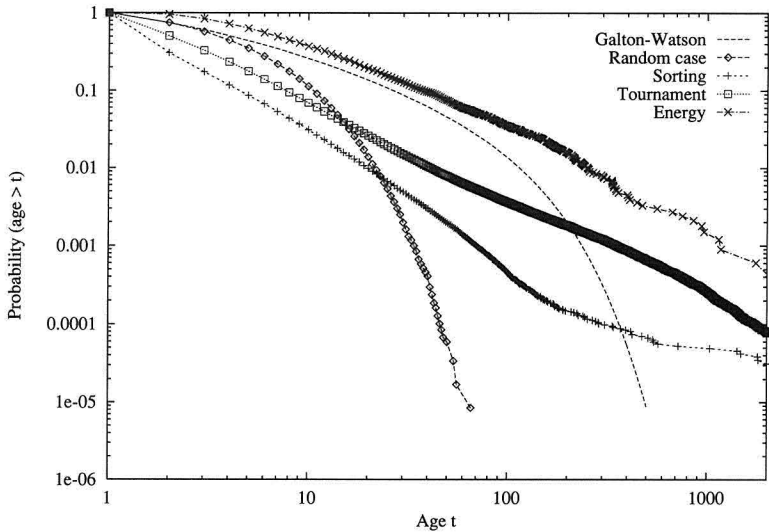
Figure 4: Comparison of species age distribution to the random-selection case on a double logarithmic plot. To make things clearer we show on the y-axis the probability that a species survives at least t time steps. The line labeled "Galton-Watson" accords to the line labeled "Recursion" of Figure 3, corresponding to an infinite population. We see an exponential distribution with an effective cut-off age $t_{\text{cut-off}} \simeq 500$. In a finite population this age decreases sharply, as can be seen for the data labeled "random case" ($\diamond$) that are from a simulation with a population of 64 bit strings ($t_{\text{cut-off}} \leq 100$). In contrast to this we see that, with selection, the distribution can best be described by a power law, even for small populations (64 and 128 bit strings).

our assumption that using the theory of branching processes is valid for the random selection model.Therefore we have only one time scale present that is related to the probability that an offspring belongs to the same species as its parent. In contrast to this we see a completely different behavior if selection is present (see Figure 4). The data are better described by a power law; from this we can conclude that there is more than one time scale.

5.2 Sorting and replacing

In our implementation of this model we sort the population in accordance with the fitness of the individuals and let the best half (or quarter) replicate and replace the worst half (or quarter).

The results of our simulations (of 64 individuals) can be summarized as follows. The strong selection pressure causes an optimization procedure that chooses the successful genes depending on the values of the interaction
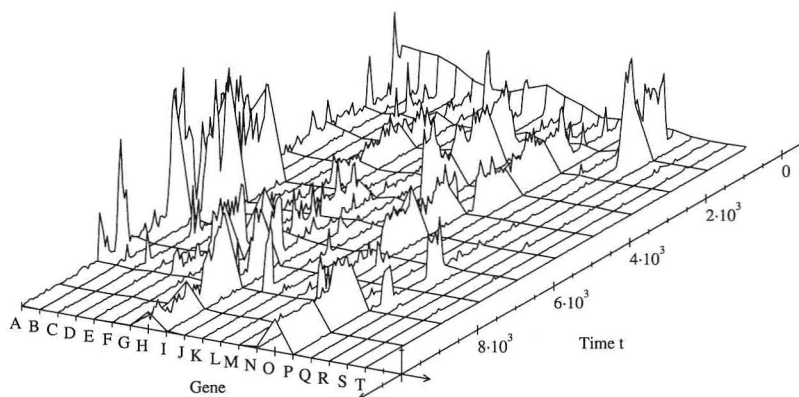
Figure 5: Occupation number of genes as a landscape over the time-genes plane. Only 2 to 5 genes have large occupation numbers, all other genes are negligible. The dominating group of genes changes several times during this simulation (*genetic takeover*).

matrix (see Figure 5). This leads to a "poorness" in the species landscape, in which only one quasispecies (a "wildtype" with few mutants) survives. The system gets "stuck" in a local optimum from which it cannot escape for long periods of time. After such a period of metastability very sudden changes appear until the next state of intermediate stability is reached (see Figure 5; $t \simeq 1500, 3900, \ldots, 8000$). This dynamics of punctuated equilibria is typical for our models, and can also be seen when we look at the *species traces* in Figure 6. This figure shows all species that were in the simulation for at least 20 time steps. The species are numbered in order of appearance; the later a species appears for the first time the higher its number. The number of individuals belonging to that species is not taken into consideration. In some phases many new species come up until a takeover takes place (e.g., $t \in [3900, 6000]$) whereas in other phases there is little change (e.g., $t \in [1500, 3900]$ and $t \in [8000, 10000]$). Usually the strongest species has 5 (in cases of change) to 30 (in stable phases) members and there are approximately 30 species, but only 2 to 3 of them have more than 4 members.

In Figure 6 there are also species traces for the random model. In contrast to the selection case, a nearly constant flux of new species occurs and, moreover, species do not reappear. Although it seems that a single (or few) species take over the whole population, this is not true. There are very few species that survived at least 20 time steps, but on average there are more than 50 species present in each time step. The strongest species usually has no more than 3 or 4 members. From this we conclude that in the random case the dynamics is similar to a random walk in the highly dimensional phase space of possible species.

When looking at the details of the evolutionary process of this model, it becomes evident that there are very few pathways for mutations that can
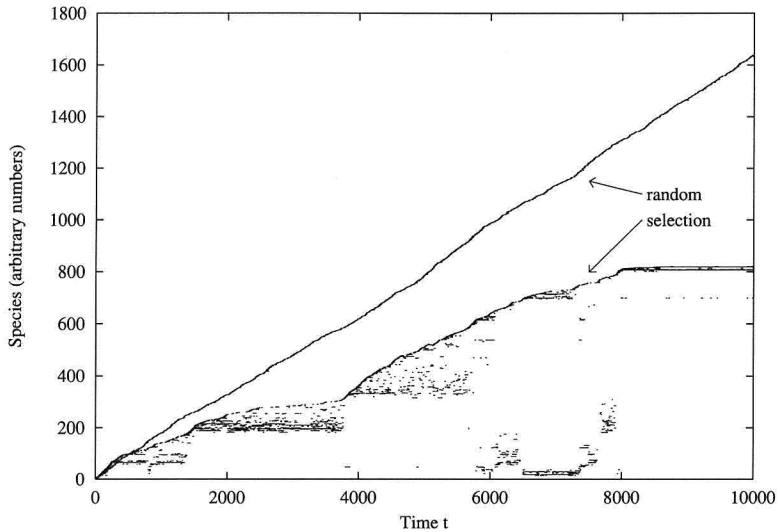
Figure 6: Species traces for the selection case compared using sorting on the random case. All species that were in the simulation at least once for more than 20 time steps are shown. The species are labeled according to their first appearance in the simulation. In the random case there is approximately a straight line that corresponds to a constant flux of new species.

create new successful individuals. Most of the possible mutations create genes that are eliminated during the following time steps so that this material cannot enter the population.

5.3 Tournament selection

The "sorting and replacing" algorithm leads to a poor species landscape because the selection pressure is so strong that individuals that are only slightly less fit than the average drop out immediately. For these individuals there is no chance to prove more useful in future generations, which contradicts the properties of natural evolutionary systems [10]. This is improved in tournament selection. We simulated two different versions of tournament selection models: one is a single round model, so $p_e = 1/2$; the other includes two rounds in each selection phase, so the time average $\langle p_e \rangle = 1/3$ (see Appendix C).

Although for the sorting model we see a distinct deviation from the exponential age distribution for individuals, this is not true for tournament selection (see Figure 7).There is only a very slight but systematic deviation from equation (2) for higher ages ($t > 25$) which is caused by selection. For very high ages ($t > 35$), we see stronger deviations, reflecting the small number of individuals that reach a very old age (the smallest entry corresponds
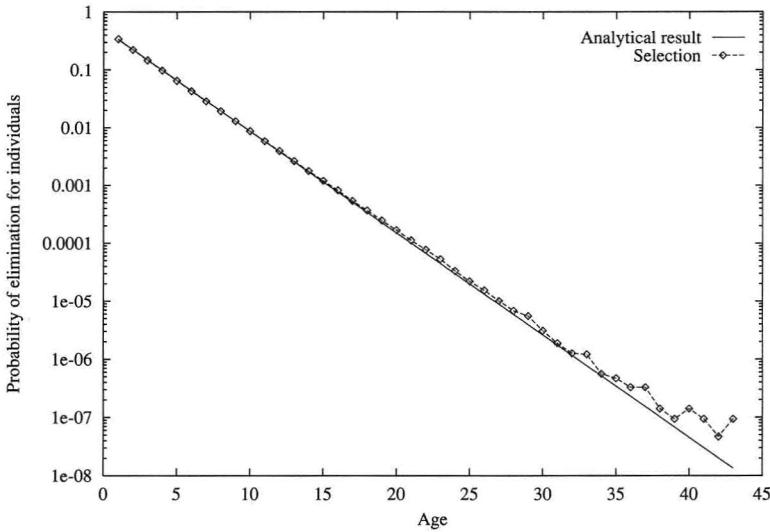
Figure 7: Comparison of the age distribution for tournament selection to equation (2) with $p = 1/3$.

to one individual during a simulation of 64 individuals over 10^6 time steps). These small deviations do not imply that tournament selection corresponds to a random model, as is evident from Figure 4. There we also see that the age distribution for species is best described by a power law.

Although the selection pressure seems less strong compared to the sorting and replacing model, a very similar situation appears. Once again only one quasispecies survives and the phenomenon of punctuated equilibria appears (see Figure 8). The number of different species in the population is comparable to the number in the previous model. Figure 8 shows all species that appeared at least once for more than 50 time steps.

5.4   Energy flow

We implemented different versions of an energy-flow model [15] and compared the results to those of the two other selection algorithms. For the first version, we chose fixed values for (a) energy source per time step, (b) energy cost per time step, and (c) threshold for replication, to keep the results comparable to the other models. The dynamics of this version lead to constant averages for different variables (e.g., population size, number of replicating individuals per time step) and thus ensures that the age distributions for individuals and species are well defined for long simulations (see Figures 2 and 4).

The species landscape for the energy-flow model shows more variability in comparison to the previous models. Figure 9 shows the species traces for all species that were in the simulation more than 50 time steps. A much
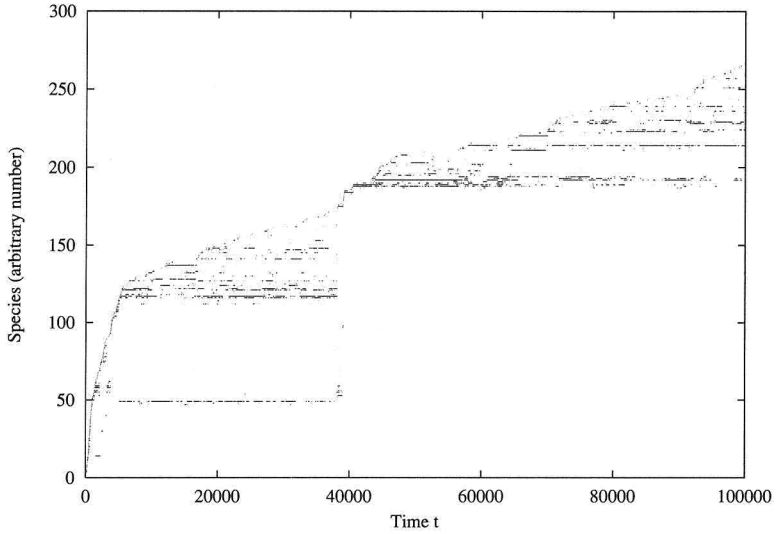
Figure 8: Species traces for the tournament selection model. All species are shown that were at least 50 time steps in the simulation. The species are labeled according to their first appearance in the simulation.
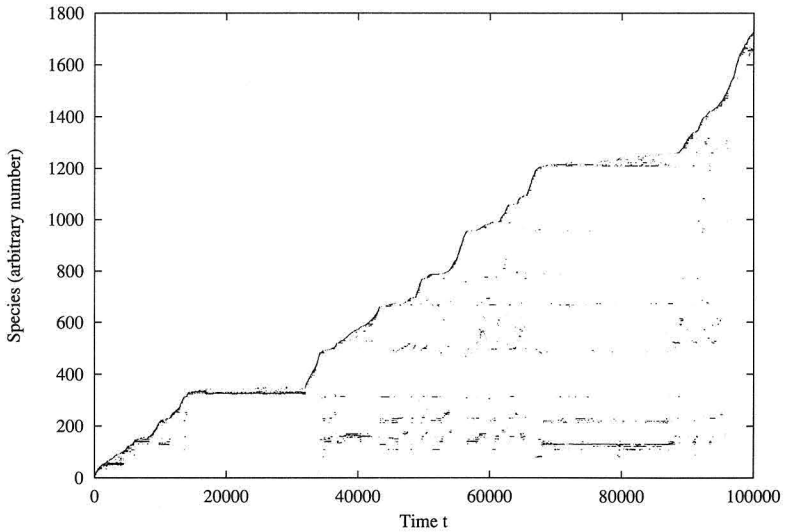

Figure 9: Species traces for the energy model. All species are shown that were in the simulation at least 50 time steps. The species are labeled according to their first appearance in the simulation.

larger number of species, namely 1734, endured more than 50 times steps, in contrast to only 267 species in Figure 8.

But the main results are the same as before: Very few species dominate the dynamics, and a tendency toward short individuals (measured in information contents) is seen.

6. Conclusions

Although in our models there is no real open-ended evolution, some features of natural evolutionary systems do appear, such as punctuated equilibria [8] and the need for "soft" selection [10] and startcodons. Our populations show a stochastic behavior of metastable states interrupted by short phases of genetic takeover. The principle of stochasticity combined with deterministic selection (i.e., Darwinian evolution) seems to cause this metastable behavior, so there is no need for external catastrophes (such as asteroid impacts, as mentioned by [1]).

Another important result for our evolutionary models is the influence of selection on the age distribution of individuals and species. While the age distribution of individuals shows (sometimes only slight) deviation from an exponential distribution when comparing a specific model to a random version of this model, the species age distribution shows more significant changes. In fact, we can give an explicit analytic formula only for the random-selection case, which can be interpreted as a branching process (see Appendix B). For random models the species age distribution converges asymptotically to an exponential distribution, while for nonrandom selection models a power law appears for a wide range of species ages. While for individuals there exists only one time scale (related to the elimination probability p), for species there are several time scales that can lead to the observed power law. This could be a first hint at open-ended evolution.

Nevertheless our models have some deficiencies. Although we see some evolutionary process occurring, it seems that in the long run we have only an optimization process. This is very likely caused by the simple linear interaction scheme described in Section 2.2. This optimization leads to only sparse diversity in the species landscape and contradicts the results of natural evolution. The linear part of the dynamics (i.e., the interaction matrix) thus dominates the stochastic parts (the random choice of individuals and genes). Consequently one has to change the interaction to some nonlinear procedure to have a chance of open-ended evolution. But it is not at all clear how this nonlinear interaction could be implemented in a natural way.

When using energy flow, the dynamics lead to more diversity of species (compare Figures 8 and 9), but in principle the successful gene groups stay the same.

Appendix A   Elimination probability of the individuals

The elimination probability distribution of the individuals can be calculated for the case of a constant fraction $p_e \in [0, 1]$ of replaced individuals per time

step. The expectation value d_t for the probability of an individual being eliminated after a lifetime of t time steps is given by

$$d_t = p(1-p)^{t-1}. \tag{2}$$

This can be rewritten as an exponential distribution

$$d_t = a \times \exp(-bt), \quad a = \frac{p}{1-p} \geq 0, \quad b = -\ln(1-p) \geq 0. \tag{3}$$

The average lifetime $\tau = \sum t d_t$ is then given by p^{-1}.

Appendix B   Lifetimes of species

If we consider a random model we can use the theory of branching processes [9] to estimate the species lifetime distribution. In its simplest form it is called the Galton-Watson branching process and can be formulated as follows:

> Let $p_0, p_1, p_2, \ldots$ be the respective probabilities that an individual has $0, 1, 2, \ldots$ offspring, let each offspring have the same probability for offspring of its own, and so on. What is the probability that this line is extinct after n generations, and more generally what is the probability for any given number of descendants in this line in any given generation?

Theory tells us that if the average number of offspring m is less than 1, each line goes extinct with probability 1. Because our population size is constant (at least on average) and we use mutations as a source of change, we indeed have $m < 1$. So each species definitely becomes extinct. We are now interested in the age distribution of species.

Branching processes are analyzed using the generating function of the probability distribution $\{p_k\}$ of offspring

$$f(s) = \sum_{k=0}^{\infty} p_k s^k \quad \text{with } 0 \leq s \leq 1 \tag{4}$$

and their iterates

$$f_1(s) = f(s) \text{ and } f_{n+1}(s) = f(f_n(s)) = f_n(f(s)) \text{ for } n = 1, 2, \ldots \tag{5}$$

The power series of f_n has as coefficients $p_{n,k}$ the probabilities that after n generations a species contains k members, iff in generation 0 there was only 1 member. From this follows that

$$Q_n = p_{n,0} - p_{n-1,0} = f_n(0) - f_{n-1}(0) \tag{6}$$

is the probability that a species goes extinct after exactly n generations (for details see [9]).

For our models we have the following features. In the selection phase of each time step a part P_0 of the individuals is eliminated, a part P_1 survives, and a part P_2 survives and creates one descendant. For the following calculations we are interested only in those offspring that belong to the same

species where M labels the probability for identical offspring. This changes our probabilities P_i to P_i':

$$P_0' = P_0$$
$$P_1' = 1 - (1 + M)P_0 \tag{7}$$
$$P_2' = MP_0.$$

If we take our dynamics literally and regard each time step as one generation, we get $f(s) = P_0' + P_1's + P_2's^2$ and $m = f'(0) = 1 - P_0' + P_0'M$. Performing the recursion needed in equation (6), we get results that show an exponentially decreasing age distribution for species (see Figure 3, the line labeled "recursion").

A disadvantage of this recursion is that we do not obtain an analytical result. But if the probabilities $\{p_k\}$ form a geometric series of the form

$$p_k = bc^{k-1} \quad \text{with } k = 1, 2, \ldots \text{ and } 0 < b, c \text{ and } b < 1 - c \tag{8}$$

we get the following analytical form for the iterates:

$$f_n(s) = 1 - m^n \left(\frac{1 - s_0}{m^n - s_0} \right) + \frac{m^n \left(\frac{1-s_0}{m^n-s_0} \right)^2 s}{1 - \left(\frac{m^n-1}{m^n-s_0} \right) s}, \quad n = 1, 2, \ldots \tag{9}$$

with $m = b/(1 - c)^2$ and $s_0 = (1 - b - c)/c(1 - c)$.

Starting from equation (7) we can easily calculate the probabilities $\{p_k\}$ as defined in equation (8). For example, the probability p_0 that an individual has no offspring contains two parts, one that the individual does not survive the first time step (i.e., P_0') and a second that it survives the first time step but has no offspring (i.e., $P_1'p_0$):

$$p_0 = P_0' + P_1'p_0. \tag{10}$$

This gives $p_0 = 1/(1 + M)$. In a similar fashion we get a recursion for the p_k,

$$p_k = P_1'p_k + P_2'p_{k-1}, \tag{11}$$

which leads to

$$b = \frac{M}{(1 + M)^2} \tag{12}$$

$$c = \frac{M}{1 + M} \tag{13}$$

as defined in equation (8), which obeys $b < 1 - c$ as demanded.

Inserting all this into equation (6) we get

$$Q_n = \frac{M^n(1 - M)^2}{(1 - M^n)(1 - M^{n+1})} \propto M^n, \tag{14}$$

which is for $M < 1$ asymptotically an exponential distribution.

At this stage we have to make some remarks regarding our approximation. We implicitly made the assumption that we have non-overlapping generations, which considering our dynamics is definitely not true. In addition, in our simulations we measure time in time steps and not in generations. This poses no problem if we rescale with the average lifetime τ as defined in Appendix A:

$$Q(t) = \frac{1}{\tau} Q_{t/\tau}. \tag{15}$$

Simulations of the random model with a large population (10^4 individuals) over long times (2×10^5 time steps; see the data dots in Figure 3) show that the results, described by equation (15), are not as good for $t < 50$ (see Figure 3, the "analytical" line) as for large values of t, although both lines show the same exponential behavior.

Appendix C   Elimination probability for tournament selection

If tournament selection is performed by choosing a tournament size of $s = 2$ and making two rounds, the best individual produces $m = 2$ offspring, the median individual produces $m = 1$ offspring, and the worst individual produces $m = 0$ offspring. These properties are similar for the sorting and replacing model, but for tournament selection they are only valid on a statistical basis.

The probability p for elimination is a straightforward calculation. After an interaction phase each individual can be characterized by a number $x \in [0, 1]$ that is equivalent to the probability that another randomly chosen individual has a lower score. The best individual has $x = 1$, whereas the worst has $x = 0$. Consider the probabilities $q_i(x)$ of an individual producing i offspring:

$$q_0(x) = (1 - x)^2, \quad q_1(x) = 2x(1 - x), \quad q_2(x) = x^2. \tag{16}$$

The average probability p_i of producing i offspring is given by the average of $q_i(x)$ over the entire population,

$$p_i = \int_0^1 q_i(x)\, dx, \tag{17}$$

from which follows $p_0 = 1/3$.

Both the conditions $\sum_{i=0}^{s} p_i = 1$ and $\sum_{i=0}^{s} i p_i = m$ must be valid. Because we consider an average constant population size, we get $p_2 = p_0$, which can be checked easily. Generalization of these results to a different tournament size is straightforward.

Acknowledgments

References

[1] W. Alvarez, F. Asaro, H. V. Michel, and L. W. Alvarez, "Iridium Anomaly Approximately Synchronous with Terminal Eocene Extinctions," *Science*, **216** (1982) 886–888.

[2] A. M. Assad and N. Packard, "Emergent Colonization in an Artificial Ecology," Preprint (1991).

[3] F. Crick, *Life Itself: Its Origin and Nature* (New York, Simon and Schuster, 1981).

[4] B. Derrida and L. Peliti, "Evolution in a Flat Fitness Landscape," *Bulletin of Mathematical Biology*, **53** (1991) 355–382.

[5] M. Eigen and P. Schuster, *The Hypercycle* (Berlin, Springer-Verlag, 1979).

[6] H. Freund and R. Wolter, "Evolution of Bit Strings: Some Preliminary Results," *Complex Systems*, **5** (1991) 279–298.

[7] D. E. Goldberg, B. Korb, and K. Deb, "Messy Genetic Algorithms: Motivation, Analysis, and First Results," *Complex Systems*, **3** (1989) 493–539.

[8] S. J. Gould, *Wonderful Life: The Burgess Shale and the Nature of History* (New York, W.W. Norton, 1989).

[9] T.E. Harris, *The Theory of Branching Processes*, (Berlin, Springer-Verlag, 1963).

[10] M. Kimura, "The Neutral Theory of Molecular Evolution," *Scientific American*, **241** (November 1979) 94–104.

[11] C. G. Langton (editor), *Artificial Life*, (Reading, Mass., Addison-Wesley, 1989).

[12] C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (editors), *Artificial Life II*, (Reading, Mass., Addison-Wesley, 1992).

[13] J. Maynard-Smith, *Evolution and the Theory of Games* (Cambridge, Cambridge University Press, 1982).

[14] N. Packard, "Intrinsic Adaption in a Simple Model for Evolution," pages 141–155 in *Artificial Life*, edited by C. G. Langton (Reading, Mass., Addison-Wesley, 1989).

[15] R. Wolter, "Selbstorganisierende Evolutionsmodelle," Diplomarbeit, University of Wuppertal (1992).