

Error Thresholds and their Relation to Optimal Mutation Rates

Gabriela Ochoa, Inman Harvey, and Hilary Buxton

Centre for the Study of Evolution
Centre for Computational Neuroscience and Robotics
COGS – The University of Sussex
Falmer, Brighton BN1 9QH, UK
{gabro,inmanh,hilaryb}@cogs.susx.ac.uk

Abstract. The *error threshold* — a notion from molecular evolution — is the critical mutation rate beyond which structures obtained by the evolutionary process are destroyed more frequently than selection can reproduce them. We argue that this notion is closely related to the more familiar notion of *optimal* mutation rates in Evolutionary Algorithms (EAs). This correspondence has been intuitively perceived before ([9], [11]). However, no previous study, to our knowledge, has been aimed at explicitly testing the hypothesis of such a relationship. Here we propose a methodology for doing so. Results on a restricted range of fitness landscapes suggest that these two notions are indeed correlated. There is not, however, a critically precise optimal mutation rate but rather a range of values producing similar near-optimal performance. When recombination is used, both error thresholds and optimal mutation ranges are lower than in the asexual case. This knowledge may have both theoretical relevance in understanding EA behavior, and practical implications for setting optimal values of evolutionary parameters.

1 Introduction

The *error threshold* — a notion from molecular evolution — is the critical mutation rate beyond which structures obtained by the evolutionary process are destroyed more frequently than selection can reproduce them. With mutation rates above this critical value, an optimal solution would not be stable in the population, i.e., the probability that the population loses these structures is no longer negligible. On the other hand, an optimum mutation rate — a more familiar notion within the EAs community — is the mutation value which solves a specified search or optimization problem with optimal efficiency, that is with the least number of generations or function evaluations.

The notion of error threshold seems to be intuitively related to the idea of an optimal balance between *exploitation* and *exploration* in genetic search. In this sense, we argue that optimal mutation rates are related to error thresholds. The aim of this paper is to test this hypothesis using an empirical approach together with knowledge from molecular evolution theory.

Optimal parameter settings have been the subject of numerous studies within the EA community [2], [6], [17], and particular emphasis has been placed on finding optimal mutation rates [9]. There is, however, no conclusive agreement on what is best, most people use what has worked well in previously reported cases. It is very difficult to formulate *a priori* general principles about parameter settings, in view of the variety of problem types, encodings, and performance criteria possible in different applications. Our hypothesis — that optimal mutation rates are correlated to the notion of error thresholds — promises practical relevance and useful guidelines in finding optimal parameter settings, thus enhancing evolutionary search.

In the remainder of the paper we summarize the knowledge from molecular evolution relevant to our argument: the notions of *quasispecies* and error thresholds; we discuss the relation between error thresholds and optimal mutation rates; and we describe the fitness landscape used for our experiments: the *Royal Staircase* functions. Thereafter, we describe the empirical methodology used to test the hypothesis under study, we present the experimental results obtained, and we discuss the insight gained.

2 Quasispecies And Error Thresholds

The concept of a ‘quasi-species’ was developed in the context of polynucleotide replication, and in particular studies of early RNA evolution [3], [4], [5]. A protein space, [12] or more generally a sequence space, can be modelled as the space of all possible sequences of length ν drawn from a finite alphabet of size A . Each sequence has a fitness value which specifies its replication rate, or expected number of offspring per unit time. The fitnesses of all A^ν possible sequences define a ‘fitness landscape’. When $A = 2$, a binary alphabet, the fitness landscape is equivalent to specifying fitness values at each vertex of a ν -dimensional hypercube; with some mathematical imagination — and some caution — this can be pictured as spread out over a geographical landscape where fitness is analogous to height, and the dynamics of evolution of a population corresponds to movement of the population over such a landscape.

Given an infinite population, and a specified mutation rate governing errors in (asexual) replication, one can determine the stationary sequence distribution reached after any transients from some original distribution have died away [4]. Unless the mutation rate is too large or differences in fitnesses too small, the population will typically cluster around the fittest sequence(s), forming a concentrated cloud; the average Hamming distance between two members of such a distribution drawn at random will be relatively small. Such a clustered distribution is called a ‘quasi-species’. As the mutation rate is increased, the local distribution widens and ultimately loses its hold on the local optimum.

This can be seen at its clearest in an extreme form of a fitness landscape which contains a single peak of fitness $\sigma > 1$, all other sequences having a fitness of 1. With an infinite population there is a phase transition at a particular error rate p , the mutation rate at each of the ν loci in a sequence. In [5], this critical

error rate (the error threshold) is determined analytically (Equation 1), and it is defined as the rate above which the proportion of the infinite population on the peak drops to chance levels.

$$p = \frac{\ln(\sigma)}{\nu} \quad (1)$$

In equation 1, σ represents the selective advantage of the master sequence over the rest of the population, and ν the chromosome length. In the simplest case σ is the ratio of the master sequence reproduction rate (fitness) to the average reproduction rate of the rest.

2.1 Error Thresholds In Finite Populations

In [14] the calculations of an error threshold for *infinite* asexually replicating populations are extended to *finite* populations (we shall call the critical rate p_M for a population of size M). Finite populations lose grip on the solitary spike of superior fitness easily, because of the added hazard of natural fluctuations in this case. In [15], we derived a reformulation of the Nowak and Schuster analytical expression. This new expression (equation 2) explicitly approximates the extent of the reduction in the error threshold as we move from infinite to finite populations. The expression strictly should be an infinite series in which successive terms get smaller; here, we are ignoring all after the first few:

$$p_M = \frac{\ln(\sigma)}{\nu} - \frac{2\sqrt{\sigma-1}}{\nu\sqrt{M}} + \frac{2\ln(\sigma)\sqrt{\sigma-1}}{\nu^2\sqrt{M}} \quad (2)$$

3 Error Thresholds and Optimal Mutation Rates

The notion of error threshold seems to be intuitively related to the idea of an optimal balance between *exploitation* and *exploration* in genetic search. Too low a mutation rate implies too little exploration; in the limit of zero mutation, successive generations of selection remove all variety from the population, and once the population has converged to a single point in genotype space all further exploration ceases. On the other hand, clearly, mutation rates can be too excessive; in the limit where mutation places a randomly chosen allele at every locus on an offspring genotype, then the evolutionary process has degenerated into random search with no exploitation of the information acquired in preceding generations.

Any optimal mutation rate must lie between these two extremes, but its precise position will depend on several factors including, in particular, structure of the fitness landscape. It can, however, be hypothesized that where evolution proceeds through a successive accumulation of information then a mutation rate close to the error threshold is an optimal mutation rate for the landscape under study; since this should maximise the search done through mutation subject to the constraint of not losing information already gained. The main purpose of our paper is to empirically test this hypothesis (section 5).

Some biological evidence supports the relationship between error thresholds and optimal mutation rates. Eigen and Schuster [5] have pointed out that viruses — which are very efficiently evolving entities — live within and close to the error thresholds given by the known rates of nucleotide mutations. This correspondence has also been noticed before in the GA community: Hesser and Manner [9], devised a heuristic formula for optimal setting of mutation rates inspired by previous work on error thresholds [14]; Kauffman [11] (p. 107) also suggest a relationship between these two notions.

4 Royal Staircase Fitness Functions

van Nimwegen and Crutchfield [19] proposed the *Royal Staircase* functions for analyzing epochal evolutionary search. This class of functions are related to the previous *Royal Road* functions [13]. In [19] the authors justify their particular choice of fitness function both in terms of biological motivations and in terms of artificial evolution issues. In short, many biological systems and artificial evolution problems have highly degenerate genotype-to-phenotype maps; that is, the mapping from genetic specification to fitness is a many-to-one function. Consequently, the number of different fitness values that genotypes can take is much smaller than the number of different genotypes. Moreover, due to its high dimensionality, it is possible for the genotype to break into networks of “connected” sets of equal-fitness genotype that can reach each other via elementary genetic variation steps such as point mutation. These connected subsets of iso-fitness genotypes are referred to as “neutral networks” [10].

Our paper is guided by the working hypothesis that many real search problems have genotype search spaces that decompose into a number of such neutral networks. Such neutrality has been observed in problem domains as diverse as molecular folding [18], evolvable hardware [8], and evolutionary robotics [7]. One symptom of evolutionary search where neutral networks are important is that of long periods of (sometimes noisy) fitness stasis (— search along a neutral network) punctuated by occasional fitness leaps (— transitions to a higher neutral network). The Royal Staircase class of fitness functions capture the essential elements discussed above, and are suitable for evaluating our hypothesis. They are defined as follows [19]:

1. Genotypes are specified by binary strings $s = s_1 s_2 \dots s_L, s_i \in \{0, 1\}$, of length $L = NK$.
2. Starting from the first position, the number $I(s)$ of consecutive 1s in a string is counted.
3. The fitness $f(s)$ of string s with $I(s)$ consecutive ones, followed by a zero, is $f(s) = 1 + \lfloor I(s)/K \rfloor$. The fitness is thus an integer between 1 and $N + 1$, corresponding to 1 plus the number of consecutive fully-set blocks starting from the left.
4. The single global optimum is $s = 1^L$; namely, the string of all 1s.

Fixing N (number of blocks) and K (bits per block) determines a particular problem or fitness landscape.

5 Experimental Design

The approach taken here is to independently assess error thresholds and optimal mutation rates, comparing then these two measures.

For the experiments, we used Royal Staircase functions (section 4). The rationale for this choice is two-fold. First, because we agree with their proposers [19] that these functions, despite their simplicity, have some ingredients encountered in evolutionary search problems. Secondly, because Staircase functions have a step feature similar to that of single peak landscapes. Theoretical results on error thresholds are available for single peak landscapes. Error thresholds can be extended to other landscapes, however, a degree of ruggedness is needed.

We used a generational GA with fitness proportional selection and without elitism. Fitness functions were Staircase functions. Specifically, we tested 6 different functions (choices of N and K): $N = 1-3$, $K = 10$; and $N = 4-6$, $K = 5$. Population size was 100, genetic operators were standard bit mutation and two-point crossover with a rate of 0.6. Several mutation rates were tested, from 0.0 to 0.2 expected mutations per bit. The algorithm was run in two modes *Asexual*: using mutation only; and *Sexual*: using both mutation and recombination. Each run lasted a maximum of 5000 generations.

5.1 Empirically Determined Optimal Mutation Rates

For the purpose of this paper, we defined the optimal mutation rate as that which finds the peak (on average) with the least number of generations.

For determining optimal mutation rates as defined above, we ran the GA starting from a random population, and stored the generation number at which the peak was attained for the first time. This measure was averaged over 100 trials for each mutation rate tested.

5.2 Empirically Determined Error Thresholds

The error threshold is the critical mutation rate beyond which structures obtained by the evolutionary process are destroyed more frequently than selection can reproduce them. Aiming at capturing this definition in an algorithm, we designed the following method for empirically estimating error thresholds:

For the selected range of mutation rates:

- Start from a population of all 1s, that is, all members on the peak.
- Run the GA for a maximum of 5,000 generations or until the whole population has completely lost the peak.
- Count how many times, out of 100 trials, the population completely loses the peak.

For low mutation rates, at least one member of the population is on the peak during all the generations, for all the 100 trials. As the mutation rate is increased, there is a point where the population completely loses the peak for

some or all the 100 trials. The error threshold is identified as the mutation rate where this transition occurs. The observations are approximate in that, firstly, precision is limited to the mutation step size used, and secondly, the limit of 5,000 generations was assumed to be sufficiently long for the purpose.

Validating the Empirical Method For validating the empirical method described above, we designed the following experiment: we considered a single peak landscape with $\sigma = 2$, and for distinct string lengths (10, 25, 50, 100, and 200), we calculated the error threshold analytically using equation 2 (valid for finite asexual populations). The population size, M , was 100. Results of these calculations are shown in the second column of table 1. Next, for the the same landscape and settings, we estimated the error thresholds empirically following our method. Results of these estimations are shown in the third column of the table. There is reasonable agreement between analytical and empirical figures, though worst at short string lengths. The empirical figures were always higher than the analytical ones, which is related to the limited number of generations used. The higher the number of generations the lower one should expect the empirical figure to be; the analytical figure assumes an infinite number of generations.

Table 1. Comparing Analytical and Empirical Error Thresholds on a Single Peak Landscape

String Length	Analytical	Empirical
10	0.05	0.11
25	0.02	0.03
50	0.01	0.015
100	0.005	0.006
200	0.003	0.004

6 Results

Figure 1 summarizes experimental results for the six landscapes studied. The curves show the number of generations to reach the global peak as a function of the mutation rate, for asexual and sexual populations. Each data point gives the number of generations for finding the peak averaged over 100 runs. Optimal mutation rates are those which find the peak with the least number of generations. Two general trends may be observed, if one excludes the first result shown. First, there is not a single critically precise optimal mutation rate, but instead a range of mutation values producing near-optimal results. The curves are U-shaped with a flat bottom. Secondly, the curves for sexual populations are shifted to the left, that is, to lower mutation values, when compared to the asexual population curves.

In the plots we indicate the empirically estimated error thresholds for asexual (solid arrows) and sexual (dotted arrows) populations. Error thresholds for the

landscapes studied were found to be within the range of optimal mutation rates, for both asexual and sexual populations. The results support the hypothesized relationship between these two measures.

An exception to these general trends is the plot for $N = 1$, $K = 10$. This scenario is equivalent to a single peak landscape, where a single string has the highest fitness, and all the other strings have the same but lower fitness. The rationale for hypothesising that optimal mutation rates are correlated with optimal mutation rates (Section 3) was that of maximising search subject to the constraint of not losing information already gained. However, this landscape is an extreme case where there is no intermediate step, no accumulation of information which might be lost through excessive mutation. The peak is found randomly without any gradual approach. Here, high mutation rates (close to 1.0) were optimal for both asexual and sexual populations.

Notice that error thresholds for sexual populations were, in all cases, lower than for asexual populations. A similar trend is observed for optimal mutation rates; this is compatible with the hypothesized correlation between error thresholds and optimal mutation rates.

The experiments determining optimal mutation rates showed standard deviations of the same order as the average number of generations measured. Thus, there were large run-to-run variations in the time to reach the optimal string. This observation was also reported by van Nimwegen and Crutchfield who performed similar experiments using the Royal Staircase function [19]. Figure 2 shows standard deviations for the $N = 2$, $K = 10$ landscape.

7 Discussion

Our results suggest that error thresholds and optimal mutation rates are indeed correlated. This empirical evidence supports previous intuitions expecting this correlation [9] [11], [15], [16]. There is not, however, a single critically precise value for the optimal mutation rate, but instead a range of values producing near-optimal performance. The error threshold, on the other hand, is a more precise measure. Hence, mutation rates slightly lower or higher than estimated error thresholds are likely to produce near-optimal results.

The implication of this finding is two-fold. First, theoretically, in helping to understand EAs' behavior, as insights regarding error thresholds will be reflected in our understanding of optimal mutation rates. Second, practically, as heuristics for finding error thresholds will provide useful guidelines for setting optimal mutation rates, thus improving the performance of EAs.

In our experiments, both error thresholds and optimal mutation ranges were lower for sexual compared to asexual populations. This result has been observed before: a recent work from the evolutionary biology literature [1] studied the role of recombination on evolving population of viruses, particularly, the effect of recombination on the magnitude of the error threshold. They report that recombination shifts the error threshold to lower mutation rates. Moreover, in [16], now in the realm of genetic algorithms, we found that recombination shifts

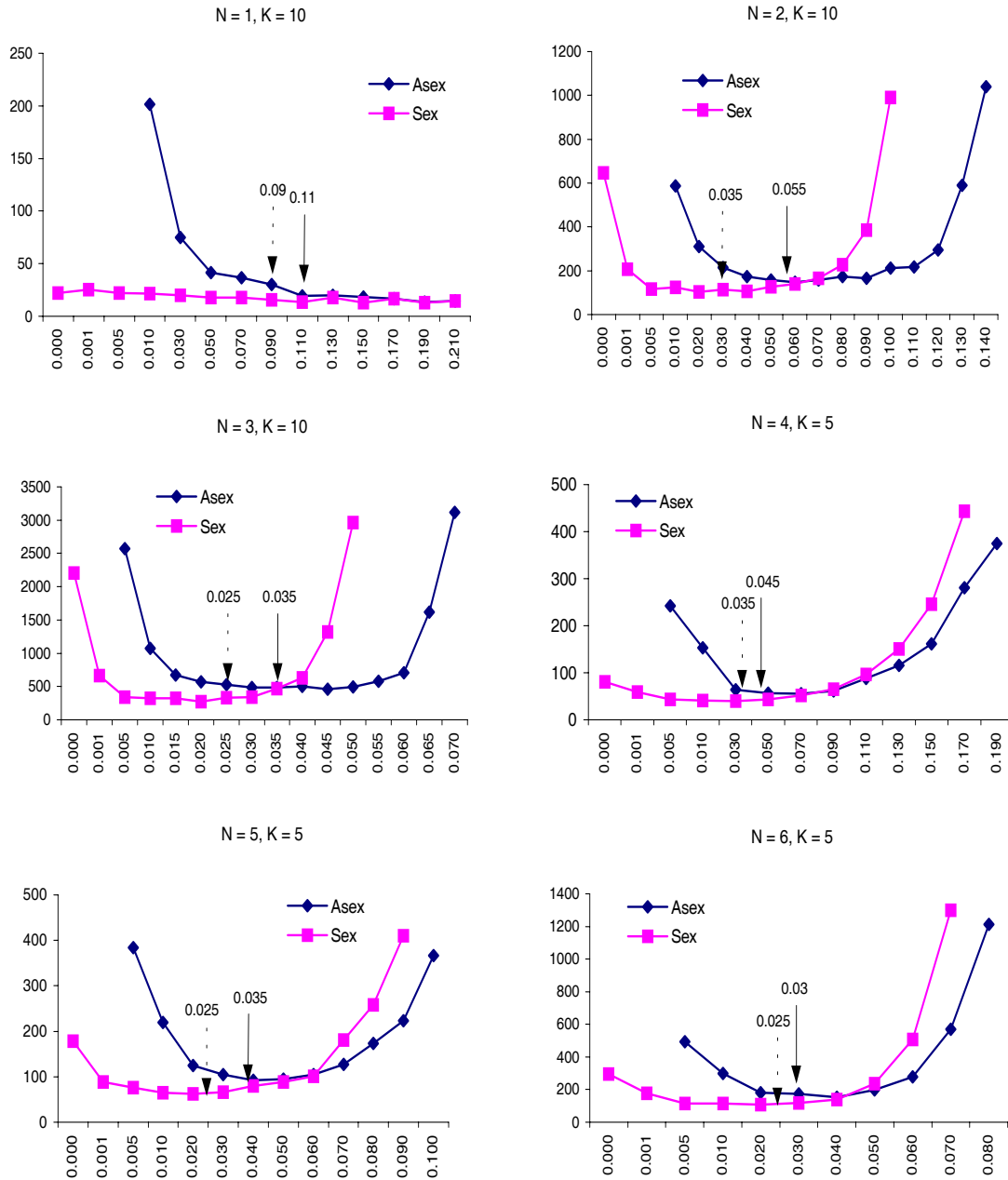


Fig. 1. For the six different landscapes explored, plots show the number of generations for finding the peak (Y axis) as a function of the mutation rate (X axis — note the scale is not linear at the lower end of the axis), for both asexual and sexual populations. Error thresholds are indicated by solid vertical arrows (asexual) and dotted vertical arrows (sexual).

N = 2, K = 10

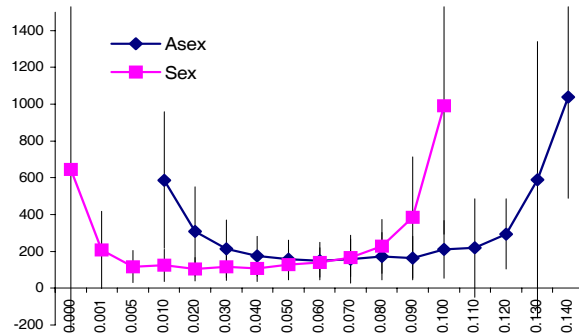


Fig. 2. To demonstrate the significant variance in the results, for one of the landscapes we show error bars for \pm one standard deviation in the number of generations for finding the peak.

optimal mutation rates to lower values. This evidence points indirectly towards the relationship between the notions of error thresholds and optimal mutation rates.

In this paper, we explored a single family of fitness functions, namely Royal Staircase functions. We have found, however, compatible results using the Kauffman NK family of landscapes. Moreover, the methods discussed here can be applied elsewhere. Error thresholds, and hence the implications of our results, hold for landscapes with certain degree of discontinuity. Thus, they are not applicable to smooth monotonic landscapes (or regions). Results will also be modified where there is elitism, either explicit or that implicit in some tournament selection algorithms for steady-state GAs.

Finally, there is a circularity implicit in using the method suggested here for assessing optimal mutation rates in real problems as opposed to toy landscapes — namely that one cannot estimate the error threshold until *after* one has found the global peak. However, knowledge of error thresholds in one region of the landscape, or in one member of a class of landscapes, may be of guidance in assessing error thresholds (and thus optimal mutation rates) elsewhere. For this to be the case one needs to make some assumptions of statistical regularity, but such assumptions are necessary anyway for EAs to be practical at all. So despite this circularity there is the prospect that in real EA applications where the fitness landscape is unknown except to the extent that it is sampled experimentally, the methodology given above allows for experimental assessment of error thresholds and thus guidance for setting optimal mutation rates.

Acknowledgements Thanks to A. Meier for support and critical reading.

References

1. M. C. Boerlijst, S. Bonhoeffer, and M. A. Nowak. Viral quasi-species and recombination. *Proc. R. Soc. London. B*, 263:1577–1584, 1996.
2. K. A. DeJong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, Ann Arbor, MI, 1975.
3. M. Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.
4. M. Eigen, J. McCaskill, and P. Schuster. Molecular quasi-species. *J. Phys. Chem.*, 92:6881–6891, 1988.
5. M. Eigen and P. Schuster. *The Hypercycle: A Principle of Natural Self-Organization*. Springer-Verlag, 1979.
6. J. Grefenstette. Optimisation of control parameters for genetic algorithms. *IEE Trans SMC*, 16(1):122–128, 1986.
7. I. Harvey, P. Husbands, D. Cliff, A. Thompson, and N. Jakobi. Evolutionary robotics: the sussex approach. *Robotics and Autonomous Systems*, 20:205–224, 1997.
8. I. Harvey and A. Thompson. Through the labyrinth evolution finds a way: A silicon ridge. In *Proc. of The First International Conference on Evolvable Systems: From Biology to Hardware (ICES96)*. Springer-Verlag, 1996.
9. J. Hesser and R. Männer. Towards an optimal mutation probability for genetic algorithms. In H.-P. Schwefel and R. Männer, editors, *Parallel Problem Solving from Nature*. Springer-Verlag, Lecture Notes in Computer Science Vol. 496, 1991.
10. M.A. Huynen. Exploring phenotype space through neutral evolution. Technical Report Preprint 95-10-100, Santa Fe Institute, 1995.
11. S. A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.
12. J. Maynard Smith. Natural selection and the concept of a protein space. *Nature*, 225:563–564, 1970.
13. M. Mitchell, S. Forrest, and J. H. Holland. The Royal Road for genetic algorithms: fitness landscapes and GA performance. In F. J. Varela and P. Bourguin, editors, *Proceedings of the First European Conference on Artificial Life. Toward a Practice of Autonomous Systems*, pages 245–254. MIT Press, Cambridge, MA, 1992.
14. M. Nowak and P. Schuster. Error thresholds of replication in finite populations: Mutation frequencies and the onset of Muller’s ratchet. *J. Theor. Biol.*, 137:375–395, 1989.
15. G. Ochoa and I. Harvey. Recombination and error thresholds in finite populations. In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms (FOGA-5)*. Morgan Kaufmann, 1998.
16. G. Ochoa, I. Harvey, and H. Buxton. On recombination and optimal mutation rates. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-99)*, 1999. To appear.
17. J.D. Schaffer, R.A. Caruana, L.J. Eshelman, and R. Das. A study of control parameters affecting online performance of genetic algorithms for function optimization. In J. D. Schaffer, editor, *Proceedings of the 3rd ICGA*. Morgan Kaufmann, 1989.
18. P. Schuster. Extended molecular evolutionary biology: Artificial life bridging the gap between chemistry and biology. *Artificial Life*, 1:39–60, 1994.
19. E. van Nimwegen and J. P. Crutchfield. Optimizing epochal evolutionary search: Population-size dependent theory. Technical Report Preprint 98-10-090, Santa Fe Institute, 1998.