# Error Thresholds in Genetic Algorithms

**Gabriela Ochoa** gabro@ldc.usb.ve
Departmento de Computacion, Universidad Simon Bolivar, Po. Box 89000, Caracas 1080-A, Venezuela

**Abstract**

The *error threshold* of replication is an important notion of the *quasispecies* evolution model; it is a critical mutation rate (error rate) beyond which structures obtained by an evolutionary process are destroyed more frequently than selection can reproduce them. With mutation rates above this critical value, an error catastrophe occurs and the genomic information is irretrievably lost. Therefore, studying the factors that alter this magnitude has important implications in the study of evolution. Here we use a genetic algorithm, instead of the quasispecies model, as the underlying model of evolution, and explore the existence of error thresholds on complex landscapes. We also study the effect of modifying the most prominent evolutionary parameters on the magnitude of this critical value. Error thresholds were found to depend mainly on the selection pressure and genotype length. Our empirical study verifies the occurrence of error thresholds in evolving populations of bit strings using a genetic algorithm. In this way, the notion of error threshold is brought from molecular evolution to evolutionary computation.

**Keywords**

Genetic algorithms, error thresholds, quasispecies model, optimal mutation rates, evolutionary parameters, NK landscapes, Royal Staircase functions, multiple knapsack problems.

## 1 Introduction

Quasispecies theory was derived by Eigen and Schuster (1979) to describe the dynamics of replicating nucleic acid molecules under the influence of mutation and selection. The theory was originally developed in the context of pre-biotic evolution (studies of the origin of life), but in a wider sense it describes any population of reproducing organisms. The Quasispecies model has been acknowledged as extremely useful in studying the behavior of populations evolving on a given landscape (Peliti, 2002), and has become a standard model to describe molecular and viral evolution (Kamp et al., 2003). A quasispecies is defined as the stationary population distribution of replicating macromolecules under mutation and selection. The most prominent feature of quasispecies is the existence of an error threshold of replication. If replication were error free, no mutants would arise and evolution would stop. On the other hand, evolution would also be impossible if the error rate of replication were too high (since selection would not be able to maintain the genetic information in the population). The notion of error threshold imposes an upper bound for the mutation rate, beyond this critical value an error catastrophe occurs and the genetic information is irrecoverably lost.

The notion of error threshold is 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, no new individuals would arise

©200X by the Massachusetts Institute of Technology 

and the search process would stagnate. On the other hand, with an excessively high mutation rate, the evolutionary process would degenerate 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 the other evolutionary parameters and the characteristics of the problem at hand. It can, however, be postulated that a mutation rate close to the error threshold would be optimal for the problem under study, because it would maximize the search done through mutation subject to the constraint of not losing information already gained. Some biological evidence supports the idea that evolution is effective close to the error threshold; certain viruses (such as the HIV virus), which are very efficient evolving entities, seem to operate very close to their error threshold (Nowak and Schuster, 1992; Bonhoeffer and Stadler, 1993). Moreover, the existence of a relationship between error thresholds and optimal mutation rates has been suggested before in the evolutionary computation community (Hesser and Männer, 1991; Kauffman, 1993). Also, this idea is implicitly suggested by Harvey (1991) , who presents a GA framework (SAGA) specially designed for evolving genetically converged populations of variable length genotypes. Finally, in (Ochoa et al., 1999) we explicitly tested, using an empirical approach, the hypothesized relationship between error threshold, our results suggested that these notions are indeed correlated.

In (Ochoa and Harvey, 1998) we demonstrated empirically the existence of error thresholds on two simple landscapes (isolated peak, and plateau) using a standard GA; it was also shown that recombination, in those landscapes, shifted error thresholds toward lower values. The present study extends those findings by studying more complex landscapes. The division between simple and complex is somewhat arbitrary. The isolated peak landscape is an extreme uncorrelated landscape, the plateau is less extreme but still highly uncorrelated. This work, on the other hand, explores correlated landscapes, and study the effect of modifying the most prominent evolutionary parameters on the magnitude of error thresholds. This study complements a previous contribution where consensus sequence plots (see section 3) and error thresholds were presented as tools for visualizing the structure of fitness landscapes (Ochoa, 2000)

Section 2 describes the test problems used: both abstract landscapes and real-world applications. Section 3 describes the *consensus sequence* plots. These plots, borrowed and adapted from theoretical biology, constitute an empirical approach for locating error thresholds on general landscapes. Section 4 uses consensus sequence plots, on two fixed abstract problems, to explore the effect of changing various evolutionary parameters on the magnitude of error thresholds. The closing empirical section (section 5), explores whether error thresholds may be identified on real-world applications.

## 2 Test Problems

Two groups of test problems were considered. First, two families of abstract fitness landscapes: Royal Staircase functions, and $NK$ landscapes. This selection is consistent with our belief that ruggedness and neutrality are two important landscape features found in real-world applications. The Royal Staircase family of functions is a very simple class of functions that allows neutrality to be modelled and tuned. The $NK$ family of landscapes is a problem-independent model for constructing landscapes that can gradually be tuned from smooth to rugged.

Second, in order to study whether the issues explored here carried over from abstract landscapes to real-world applications, two real-world domains were selected: a combinatorial optimization problem — the Multiple Knapsack problem, and an en-

gineering problem — the design of an optimal aircraft Wing-Box. This selection was somewhat arbitrary, but again is consistent with the following criteria. First, both are complex problems: the Wing-Box is an engineering design problem based on real data and constraints, and the Multiple Knapsack is a highly constrained combinatorial optimization problem known to be $NP$-hard. Second, both problems were available and relatively easy to implement, and third, both have a natural bit string encoding which was a requirement for the study of error thresholds.

### 2.1 Royal Staircase Functions

The *Royal Staircase* [1] class of functions was proposed by Nimwegen and Crutchfield for analyzing epochal evolutionary search. They justify their particular choice of fitness function both in terms of biological motivations and artificial evolution issues. Although simple, Royal Staircase functions capture some essential elements found on complex problems, namely, highly degenerate genotype-to-phenotype maps, and the existence of extended neutral networks (i.e. sets of equal-fitness sequences that can reach each other via elementary genetic variation steps such as point mutation). The working hypothesis is that many real search problems have genotype search spaces which decompose into a number of such neutral networks. This idea is supported by observations in problem domains as diverse as molecular folding (Schuster, 1994), evolvable hardware (Harvey and Thompson, 1996; Vassilev et al., 1999), and evolutionary robotics (Harvey et al., 1997). One symptom of evolutionary search in the presence of neutral networks is the existence of long periods of 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 these essential elements in a simplified form (van Nimwegen and Crutchfield, 1998). A formal definition of the Royal Staircase class of functions is given below.

1. Genotypes are specified by binary strings $s = s_1 s_2 \ldots s_L$, $s_i \in \{0, 1\}$, of length $L = NK$, where $N$ is the number of blocks and $K$ the number of bits per block.

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 $N$ and $K$ determines a particular problem or fitness landscape.

### 2.2 $NK$ Landscapes

The $NK$ family of landscapes was introduced by Kauffman (1989) in order to have a problem-independent model for constructing fitness landscapes that can gradually be tuned from smooth to rugged. In the $NK$ model, $N$ refers to the number of genes in the genotype (i.e. the string length) and $K$, to the number of genes that influence a particular gene [2]. In other words, the fitness contribution of each gene is determined by the gene itself plus $K$ other genes in the genotype. According to Kauffman, most

[1] These functions are related to the more familiar *Royal Road* functions (Mitchell et al., 1992).

[2] Notice that the meaning of parameters $N$ and $K$ differs from their meaning on the Royal Staircase class of functions.

properties of this model are independent of the alphabet size $A$, hence the simplest case of $A = 2$ (i.e. bit strings) is considered here.

By increasing the value of $K$ from 0 to $N - 1$, $NK$ landscapes can be tuned from smooth to rugged. When $K$ is small, neighboring strings will have small differences in fitness, because the bits that are different in the two strings will influence the contribution of only few bits in each string. The extreme case of $K = 0$ yields a single-peaked and smooth 'Fujiyama' landscape. When $K$ is large, on the other hand, neighboring strings will have large differences in fitness, because the differing bits of the two strings will influence the fitness of a large number of bits in each string. When $K$ assumes its largest possible value ($K = N - 1$), the fitness landscape will be completely random or "uncorrelated", because changing the value of only one bit changes the fitness contribution of all bits in the string, so the overall fitness of neighboring strings will be totally different. The $NK$ landscape, however, was invented not to explore the two extreme landscapes, but to have a model which allows the construction of an ordered family of tunable correlated landscapes.

### 2.3 Multiple Knapsack

The combinatorial optimization problem described here, called the 1/0 multiple knapsack problem, follows the specifications given by Khuri et al.(Khuri et al., 1994). This problem is a generalization of the 0/1 simple Knapsack problem where a single knapsack of capacity $C$, and $n$ objects are given. Each object has a weight $w_i$ and a profit $p_i$. The objective is to fill the knapsack with objects producing the maximum profit $P$. In other words, to find a vector $x = (x_1, x_2, \ldots, x_n)$ where $x_i \in \{0, 1\}$, such that $\sum_{i=1}^{n} w_i x_i \leq C$ and for which $P(x) = \sum_{i=1}^{n} p_i x_i$ is maximized. The multiple version consists of $m$ knapsacks of capacities $c_1, c_2, \ldots, c_m$ and $n$ objects with profits $p_1, p_2, \ldots, p_n$. Each object has $m$ possible weights: object $i$ weighs $w_{ij}$ when considered for inclusion in knapsack $j$ ($1 \leq j \leq m$). Again, the objective is to find a vector $x = (x_1, x_2, \ldots, x_n)$ that guarantees that no knapsack is overfilled: $\sum_{i=1}^{n} w_{ij} x_i \leq c_j$ for $j = 1, 2, \ldots, m$; and that yields maximum profit $P(x) = \sum_{i=1}^{n} p_i x_i$. This problem leads naturally to a binary encoding. Each string $x_1 x_2 \ldots x_n$ represents a potential solution. If the $ith$ position has the value 1 then the $ith$ object is in all knapsacks; otherwise, it is not. Notice that a string may represent an infeasible solution. A vector $x = (x_1, x_2, \ldots, x_n)$ that overfills at least one of the knapsacks; i.e., for which $\sum_{i=1}^{n} w_{ij} x_i > c_j$ for some $1 \leq j \leq m$, is an infeasible string. Rather than discarding infeasible strings, the approach suggested by Khuri et al. (Khuri et al., 1994) is to allow infeasible strings to join the population. A penalty term reduces the fitness of infeasible strings. The farther away from feasibility, the higher the penalty term of a string.

The multiple-knapsack problem instance selected here contains 30 sacks and 60 objects (that is a string length of 60); its known optimum is 7772. This instance is available online from the OR-library (Beasley, 1990), where it is identified as "Sento 1".

### 2.4 Wing-Box Problem

For the Wing-Box problem (McIlhagga et al., 1996) an industrial partner, British Aerospace, provided data from a real Airbus wing box. A common problem when designing aircraft structures, is to define structures of minimum weight that can withstand a given load. Figure 1 sketches the elements of a wing relevant to this problem. The wing is supported at regular intervals by slid ribs which run parallel to the aircraft's fuselage. On the upper part of the wing, thin metal panels cover the gap separating adjacent ribs. The objective is to find the number of panels and the thickness

of each of these panels while minimizing the mass of the wing and ensuring that none of the panels buckle under maximum operational stresses. For the experiments in this paper, the number of panels was set to 50. For encoding an individual, 13 bits were required for the first panel, and 3 for each of the others 49 panels, that is $13 + 3 \times 49 = 160$ bits.

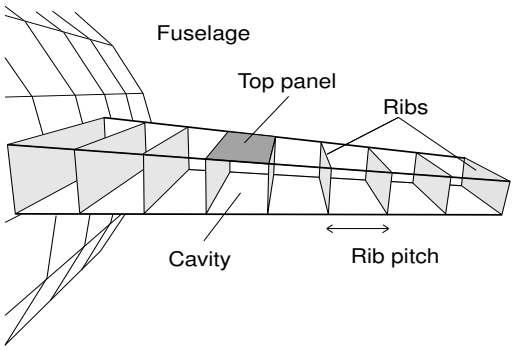


Figure 1: Relevant elements of a wing. Wing dimensions are fixed. The variable elements are the number of ribs and the thickness of the top panels.

## 3 Consensus Sequence Plots

The work of Bonhoeffer and Stadler (1993) studied the evolution of quasispecies on two correlated fitness landscapes, the Sherrington Kirkpatrick spin glass and the Graph Bipartitioning landscape. The authors described an empirical approach for locating thresholds on complex landscapes. In this section, this approach is borrowed and adapted. Instead of the quasispecies model, a GA is used as the underlying model of evolution. The resulting method can be applied for identifying error thresholds in GAs running on general complex landscapes. The approach is to calculate and plot the consensus sequence at equilibrium for a range of mutation rates. The consensus sequence in a population is defined as the sequence of predominant symbols (bits) in each position; it is plotted as follows: if the majority of individuals has a '1' or '0' in a position $i$ the field is plotted white or black, respectively. The field is plotted grey if the position is undecided. Figure 2, shows an hypothetical population and calculates its consensus sequence. The plot shown in Figure 2 will correspond to a single line in a full consensus sequence plot where the per bit mutation rate is varied (see Figure **??** for an example of a full plot). The *equilibrium state* is reached when the proportion of different sequences in the population is stationary. This happens when evolution is simulated for a large enough number of generations. In practice, it is considered that the equilibrium is reached when several parameters of the population (e.g. the maximal and average fitness) reach equilibrium. According to Bonhoeffer and Stadler (1993) the error threshold may be approached from *below* or *above*, with both methods producing similar results.

### 3.1 Approaching the error threshold from below

To approach the error threshold from below, the simulation starts with a homogeneous population at the global optimum. This approach requires knowing the optimal string

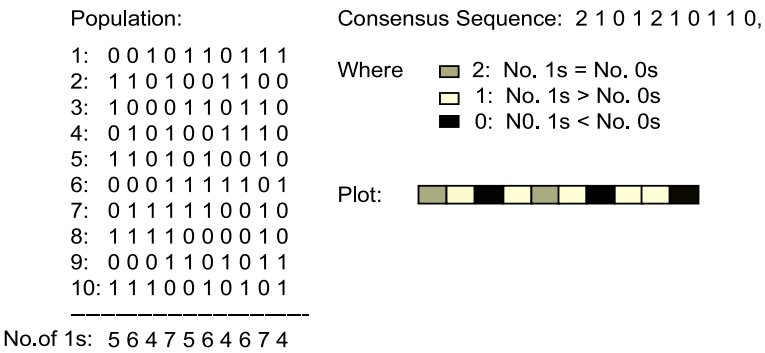


Figure 2: Calculating and plotting the consensus sequence of a population.

beforehand. Then, the population is allowed to reach equilibrium at a constant mutation rate of 0.0. Afterwards, the mutation rate is increased by a fixed, small step and the computation is continued with the current population. This process is repeated until a predefined maximum for the mutation rate is reached. The plot summarizes a single run, there is no averaging of multiple runs.

### 3.2 Approaching the error threshold from above

To approach the error threshold from above, the simulation starts with a random population. Then the population is allowed to reach equilibrium at a constant predefined maximum for the mutation rate[3]. Afterwards, the mutation rate is decreased by a fixed small step and the computation continues with the current population. This process is repeated until the mutation rate is 0.0. Notice that, in this case, it is not necessary to know the optimal string. Hence, in principle, this approach can be used for locating the error threshold on any complex landscape. Again the plot summarizes a single run, there is no averaging of multiple runs.

For both approaches, the consensus sequence in the population is calculated and plotted at the end of each simulation cycle for each mutation step. The error threshold is characterized by the loss of the consensus sequence, i.e. the genetic information of the population. Beyond the error threshold the consensus sequence is no longer constant in time (see Figure 4).

## 4 Error Thresholds and Evolutionary Parameters

This section uses consensus sequence plots to explore the effect of modifying the values of various evolutionary parameters on the magnitude of error thresholds. Unless otherwise stated, experiments use a generational GA with fitness proportional selection, a population of 100 members, and no recombination, i.e., asexual reproduction. Table 1 summarizes these default settings. Two instances of landscapes were selected as default test problems: a Royal Staircase function with number of blocks $N = 3$, and block size $K = 10$; and a $NK$ landscape with string length $N = 24$, and degree of epistatic interaction $K = 10$. Table 2 summarizes the default test problems used in most experiments. Further details on the experiments and departures from the default setting are given in the respective subsections.

[3]This value has to be high enough to be above the error threshold for the landscape under study.

| | |
|---|---|
| Population replacement | Generational |
| Selection Scheme | Proportional |
| Population size | 100 |
| Recombination rate | 0.0 (Asexual) |
| Generations (per mutation rate) | 10,000 |

Table 1: GA default parameters used in the experiments.

| *Landscape* | *Setting* | *String length* |
|---|---|---|
| $NK$ | $N = 24$, $K = 10$ | 24 |
| Royal Staircase | $N = 3$, $K = 10$ | 30 |

Table 2: Default test problems used in the experiments.

A group of preliminary studies carried out in (**?**), confirmed that: (i) error thresholds approached from below and above produce similar results, (ii) the error threshold magnitude is independent of the particular initial population; and (iii) the error threshold is similar for different instances of an $NK$ landscape with fixed $N$ and $K$. Hence, the approach followed here for estimating error thresholds is to approach them from above, that is from a random population, and using a fixed random seed for generating the initial population in all cases; and for the $NK$ landscape, selecting a single landscape instance. Notice approaching error thresholds from above is a more general method, given that it does not require knowing the optimal string beforehand.

### 4.1 Genotype Length

The analytical expression of the error threshold on a single peak landscape: $p = \frac{ln(\sigma)}{L}$ suggests that it decreases in proportion to the string length ($L$). The following experiments explore whether this is also the case on correlated landscapes. On both test problems, the plots shown below illustrate the existence of a stable consensus sequence for mutation rates below the error threshold. The consensus sequence is the string of all 1's for the Royal Staircase, and one particular local optima for the $NK$ Landscape. Moreover, for the Royal Staircase different error thresholds for each fitness level or step can also be observed. Figure 3 compares error thresholds on Royal Staircase functions of increasing length. The number of blocks $N = 3$ is kept constant, while the block size is increased from 10 to 12 and 14. Results on the Royal Staircase function suggest that error thresholds (for all fitness levels or steps) decrease as a function of the genotype length. In other words, the longer the genotype the lower the error threshold. This effect is more noticeable for the first and second step transitions.

Figure 4 compares consensus sequence plots on $NK$ landscapes of increasing genotype length. The parameter $K = 10$ (degree of epistatic interactions) is kept constant, while the string length is varied from 24 to 20 and 28. Results on the $NK$ landscape confirm that even small increases in genotype length, decrease the magnitude of the error threshold. The effect is more noticeable when increasing the genotype length from 20 to 24 than from 24 to 28. It should be noticed that the error threshold, if expressed as mutations per string, slightly decreases with each increase in string length, being 0.6 for L = 20, 0.5 for L = 24, and 0.4 for L = 28. These differences may be due to differences in the overall landscape ruggedness. In other words, although $K$ is the same for all landscapes, the string length ($N$) varies, which in turns modifies the over-

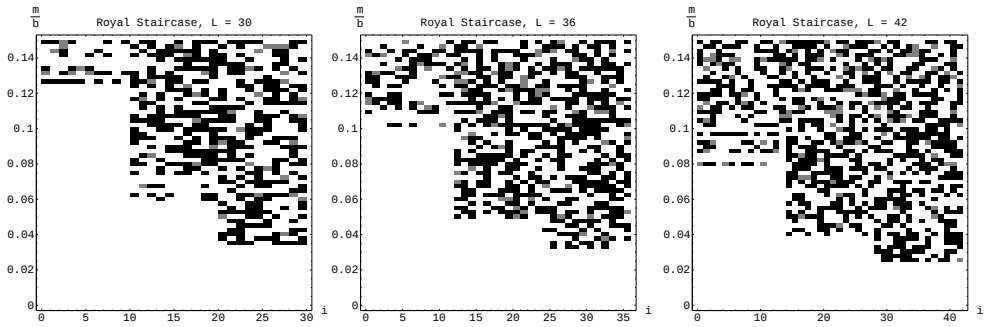


Figure 3: Error thresholds and genotype length. Consensus sequence plots on Royal Staircase functions with $N = 3$ and $K$ = 10, 12, and 14 (i.e. string lengths of 30, 36 and 42). The vertical axis shows per bit mutation rates ($m/b$).

all $NK$ landscape structure. This is just a suggested explanation, these results deserve further investigation.

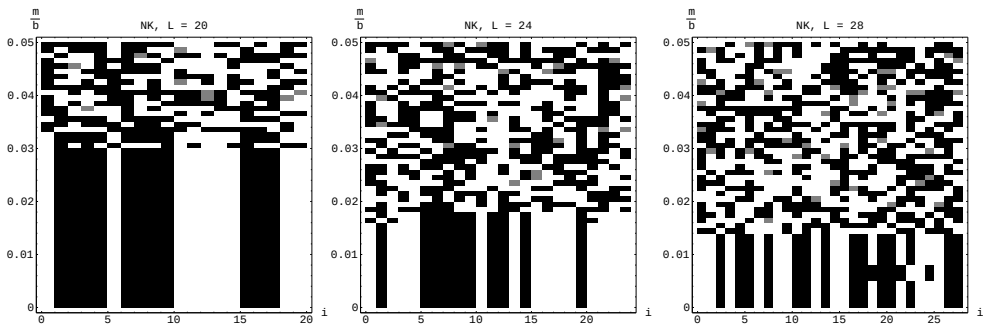


Figure 4: threshold and genotype length. Consensus sequence plots on $NK$ landscapes with $K = 10$ and $N = L$ = 20, 24, and 28. The vertical axis shows per bit mutation rates ($m/b$).

From now on, given that error thresholds were shown to depend on the length of genotypes, mutation rates will be expressed as mutations per genotype ($m/L$ where $L$ is the string length) instead of as mutations per bit. Expressing mutation rates per genotype will be more informative when looking for general principles about parameter interactions, since heuristic such as a mutation rate of $1/L$ can be identified.

### 4.2 Selection Pressure

The analytical expression of the error threshold on a single peak landscape (Equation **??**), suggests that it increases in direct proportion to the strength of selection. The following set of experiments explores whether this is also the case on correlated landscapes. These experiments use tournament selection because this selection scheme allows explicit control over the selection pressure. A common tournament size is 2, but selection pressure increases steadily for growing tournament sizes. Figure 5 shows the effect of increasing tournament sizes on the error threshold on both the Royal Staircase and $NK$ landscapes. For both landscapes, the plot using fitness proportional selection

is also included for the sake of comparison.

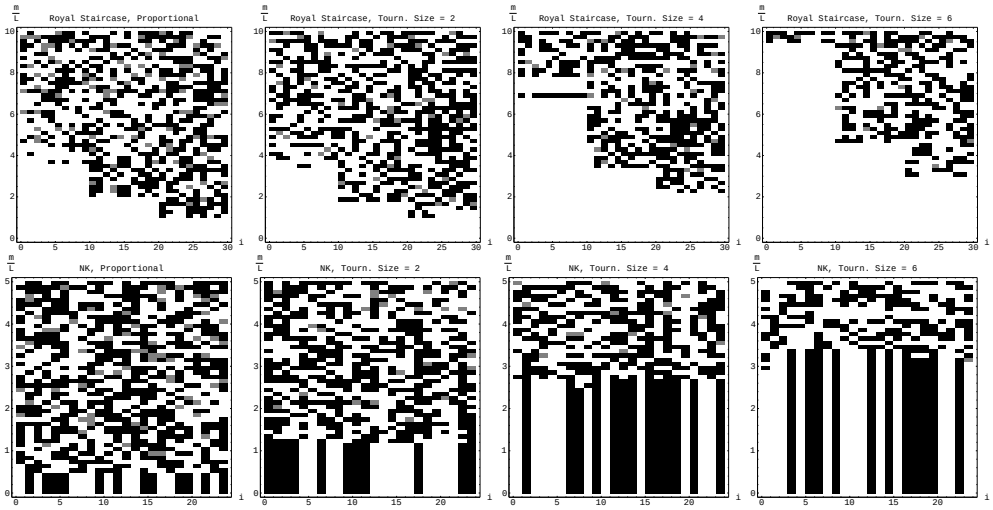


Figure 5: Error thresholds and selection pressure. Consensus sequence plots on the selected test problems for tournament sizes of 2, 4, and 6. The plots using proportional selection are also included for the sake of comparison. The vertical axis shows mutation rates per genotype ($m/L$).

Results on both landscapes show that the strength of selection has a pronounced effect on the error threshold. For increasing tournament sizes (increasing selection pressures) there is a noticeable increase in the magnitude of the error threshold. On the Royal Staircase, the effect is more noticeable for the first and second step transitions. Notice that on the $NK$ landscape (Figure 5, bottom), the error threshold for proportional selection is much lower than for tournament selection.

### 4.3 Population Size

This section explores the effect of modifying the population size on the magnitude of error thresholds. The work by Nowak and Schuster (1989) , extended the calculations of the error threshold on a single peak landscape from infinite to finite populations. In (Ochoa and Harvey, 1998) we show a reformulation of Nowak and Schuster's analytical expression, which explicitly approximates the extent of the reduction in the error threshold as we move from infinite to finite populations. The expression is an infinite series in which successive terms get smaller; here, only the first few are shown ($p_M$ is the critical rate for a population of size $M$):

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

Thus, according to the expression, the error threshold increases with the size of the population given that the second term (the 2nd. greatest of the series) is subtracting and the population size appears in the denominator.

#### 4.3.1 Preliminary Study

As a preliminary study, we compared theoretical error thresholds on a single peak landscape (calculated using Equation 1), with empirical error thresholds estimated using

consensus sequence plots, on the same landscape for various population sizes. Figure 6 shows results of this comparison. The empirical error thresholds were estimated using consensus sequence plots starting from below on a single run. The GA was allowed to run 10,000 generations for each mutation rate (that is, each line of the plot). An acceptable agreement between theory and practice was found. It can be noticed, however, that the difference increases with the size of the population. This may be due to difficulties in reaching the steady-state distribution of the population for higher population sizes. In other words, reaching the steady-state for large populations may require an impractically large number of generations. Differences may also be due to distinct models of evolution. That is, Equation 1 was derived using the quasispecies model, whereas empirical error thresholds were estimated using a GA as the underlying model of evolution.

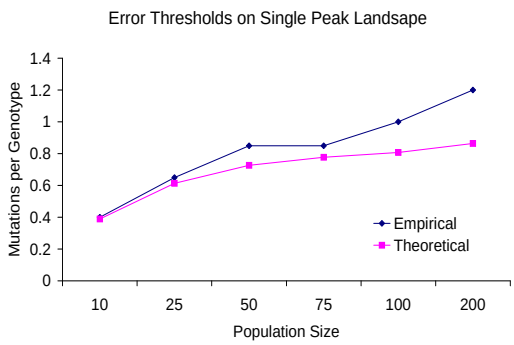


Figure 6: Comparing empirical vs. theoretical error thresholds on a single peak landscape for various population sizes.

#### 4.3.2 Main Study

The preliminary study discussed above suggests that the error threshold increases with increasing population size on a single peak landscape. The next step would be, then, to explore whether the same effect is observed on correlated landscapes. Figure 7 shows the consensus sequence plots for population sizes of 10, 20, 50, and 100; on the two default test problems.

Results on the Royal Staircase function show that error thresholds (for all fitness levels or steps) increase with increasing population size. The effect is more marked on small populations (sizes 10 and 20), and on error thresholds for the first and second step. Results on the $NK$ landscape confirm the increase on error thresholds with increasing population size. Again differences are more noticeable for small populations, and tend to stabilize for larger populations (sizes 50 and 100).

### 4.4 Steady State Population Replacement

This set of experiments compares error thresholds using generational and steady-state population replacement. In both cases tournament selection (with tournament size of 2) was used. Three types of steady-state GAs were implemented:

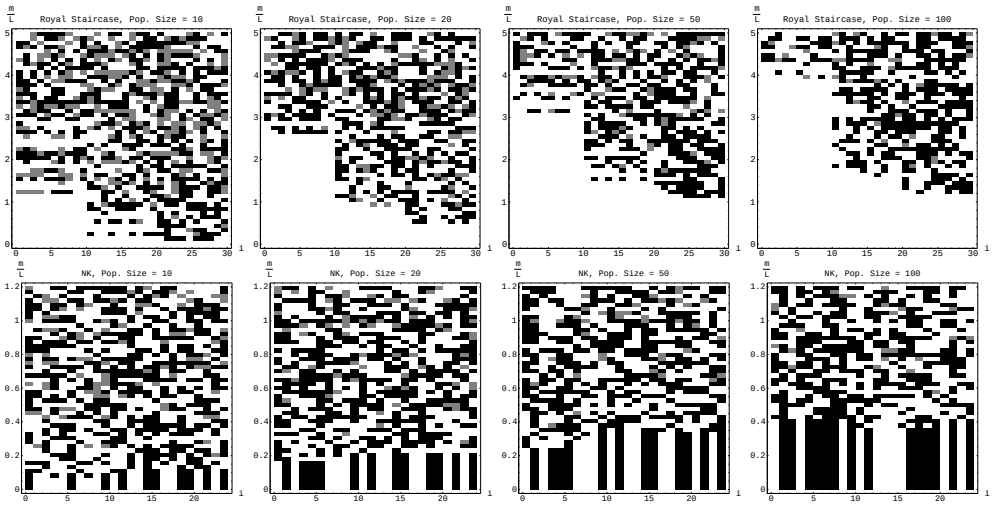


Figure 7: Error thresholds and population size. Consensus sequence plots on the selected test problems for population sizes of 10, 20, 50, and 100. The vertical axis shows mutation rates per genotype ($m/L$).

1. Using tournament selection for parents, and random selection for individuals that are to be replaced

2. Using random selection for parents, and inverse tournament selection for individuals that are to be replaced

3. Using tournament selection for parents, and inverse tournament selection for individuals that are to be replaced

Figure 8 shows the consensus sequence plots for generational replacement and the three types of steady-state replacement discussed above, on the two default test problems. Results on both test problems suggest that error thresholds depend upon the type of steady-state GA used. For type 1, the error threshold is similar to that of generational replacement, although slightly lower. On the other hand, for the other two types of replacement, which include inverse tournament selection for individuals that are to be replaced, the error threshold is noticeably higher (being highest for type 3). This last result is to be expected given that this method imposes the highest selection pressure of the three, since there is selection on parents and individuals that are to be replaced (recall from Section 4.2 that error thresholds are higher for higher selection pressures). Following this line of reasoning, results suggest that inverse tournament selection on individuals that are to be replaced, imposes a higher selection pressure than tournament selection on parents. This suggestion is supported by results in evolutionary strategies (Bäck, 1996), which points out that *extinctive* selection (i.e. a selection scheme that definitely excludes some individuals from being selected) imposes a much higher selection pressure as compared to *preservative* selection (i.e. a selection scheme that always assign selection probabilities greater than zero to all individuals). The presence of implicit elitism on steady-state replacement of types 2 and 3, may also accounts for the observed differences on the error thresholds.

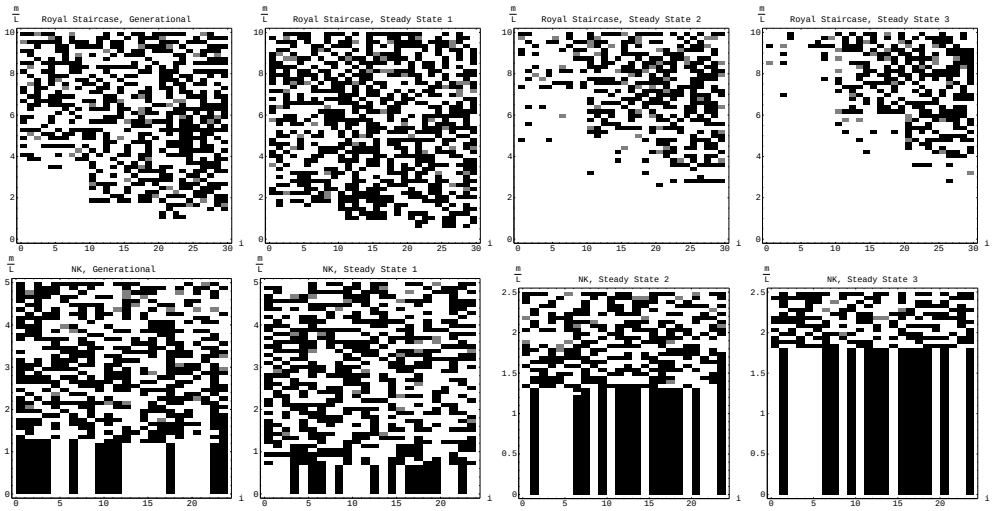


Figure 8: Error thresholds and population replacement. Consensus sequence plots on the selected test problems for generational and steady-state population replacement. Three types of steady-state replacement were tested: (1) applying tournament selection on parents and selecting individuals that are to be replaced at random, (2) selecting parents at random and applying inverse tournament selection on individuals that are to be replaced, (3) applying tournament selection on both parents and individuals that are to be replaced. The vertical axis shows mutation rates per genotype ($m/L$).

### 4.5 Recombination

The work of Bonhoeffer and Stadler (1996), and our replication of this work using GAs (Ochoa and Harvey, 1998), suggest that recombination shifts the error threshold toward lower values on the single peak and plateau landscapes. The following set of experiments explores whether this is also the case on correlated landscapes. Two types of recombination were considered: 2-point and uniform, both with a rate of 1.0. Figure 9 shows the effect of recombination on the Royal Staircase (top) and $NK$ landscape (bottom) using 2-point and uniform crossover. For both landscapes the consensus sequence plot without recombination (i.e. crossover rate = 0.0) is included for the sake of comparison.

On the Royal Staircase function (Figure 9, top) error thresholds for all the steps are lower when recombination is used. The plots with no recombination and 2-point recombination are qualitatively similar, whereas the plot using uniform recombination is different in that the transitions for the three steps are closer to one another. On the $NK$ landscape with $K$ = 10 (Figure 9, bottom), there is no noticeable difference in the magnitude of the error threshold with and without recombination. Results from (Ochoa and Harvey, 1998) suggest that the effect of recombination on the error threshold is related to the ruggedness of the landscape. Hence, an extra set of experiments explores the effect of recombination on a $NK$ landscape with increased ruggedness ($N$ = 24 and $K$ = 12). On this new $NK$ landscape (Figure 10) the error threshold is lower when recombination is used. Results are similar for 2-point and uniform recombination.

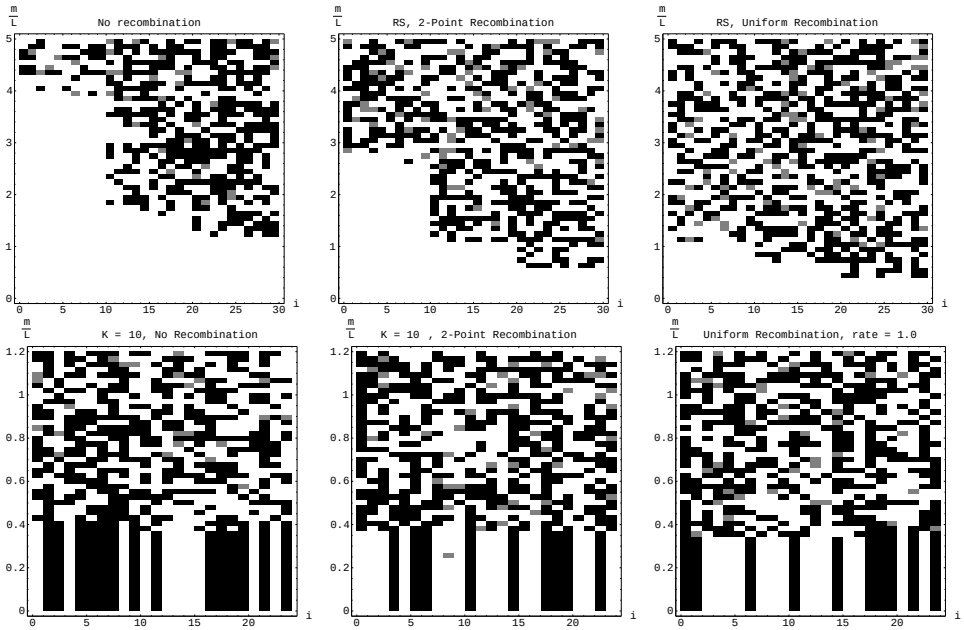


Figure 9: Error thresholds and recombination. Consensus sequence plots on the selected test problems with and without recombination. Both two-point and uniform recombination (with a rate of 1.0) were tested. The vertical axis shows mutation rates per genotype ($m/L$).

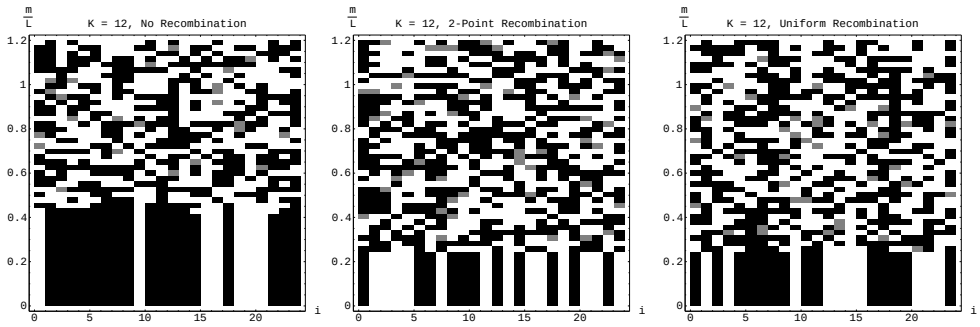


Figure 10: Error thresholds and recombination. Consensus sequence plots on a $NK$ landscape with increased ruggedness ($K$ = 12). Both two-point and uniform recombination (with a rate of 1.0) are tested. The vertical axis shows mutation rates per genotype ($m/L$).

### 4.6 Discussion

This section explored the effect of changing the values of various evolutionary parameters on the magnitude of error thresholds. A few instances of Royal Staircase and $NK$ landscapes were used as test problems. The effect of these various evolutionary parameters are summarized below:

- **Genotype length:** Results suggest that error thresholds decrease as a function of the genotype length. In other words, the longer the genotype the lower the error threshold.

- **Selection Pressure:** Results suggest that the strength of selection has a pronounced effect on the error threshold. For increasing selection pressures there is a noticeable increase in the magnitude of error thresholds. Depending on the fitness function, the use of proportional selection may produce much smaller error thresholds as compared to tournament selection.

- **Population Size:** Results show that error thresholds increase with increasing population size. This effect is more marked on small populations (smaller than 50). Differences on the error thresholds stabilize for larger populations; error thresholds for population sizes of 50, 100 and larger are quite similar.

- **Steady State Population Replacement:** Results suggest that error thresholds depend upon the type of steady-state GA used. When using tournament selection for parents and random selection for individuals that are to be replaced, the error threshold is similar to that of generational replacement. On the other hand, when the steady-state GA includes inverse tournament selection for individuals that are to be replaced (which is known to impose a higher selection pressure), the error threshold is noticeably higher. These results suggest that the magnitude of the error thresholds depend more on the selection pressure than on the type of replacement. That is, inverse tournament selection on individuals that are to be replaced, imposes a higher selection pressure, which in turn explains the higher error threshold. Also the implicit elitism on some types of steady-state GA accounts for the differences observed on the error threshold magnitudes.

- **Recombination:** For discontinuous functions (Royal Staircase), and very rugged landscapes ($NK$ landscapes with $K > 10$) error thresholds are slightly lower when recombination is used, of the order of 0.2 mutations per genotype. Similar results were obtained for uniform and two-point recombination. However, this effect of recombination was not observed on less rugged landscapes and real-world domains ((Ochoa, 2000) and Section 5).

## 5 Error Thresholds in Real-World Domains

This closing empirical section explores whether error thresholds can be observed in real-world applications. All the experiments used a generational GA with tournament selection (tournament size of 2), and a population of size 100. The GA was run in two modes: (i) using mutation only (*Asexual*), and (ii) using both mutation and recombination (*Sexual*). The recombination operator used was two-point recombination with a rate of 1.0. The mutation rate range explored was from 0.0 to 5.0 mutations per genotype with a step of 0.1. Each simulation cycle lasted 15,000 generations. Error thresholds were approached from above, that is, starting from a random population and a high mutation rate.

### 5.1 Wing-Box Problem

Figure 11 shows results on the Wing-Box problem for asexual and sexual GAs. The plots show the existence of a stable consensus sequence for mutation rates below the error threshold. The error threshold is visualized as the transition from a stable consensus sequence to a random sequence of bits. For most of the bits, the transition occurs around 1.5 mutations per genotype. An exception is the portion of bits from 11 to 16, which are randomized even for low mutation rates. These bits correspond to the less significant digits of the thickness of the first panel, and the relative thickness of the

second panel. Given the characteristics of the problem, these bits are neutral in that changes to them are not reflected in the overall fitness of the wing-structure. Also, there is a region from bit 75 to bit 125 where the consensus sequence seems more stable for higher mutation rates. Finally, there are no clear differences between the asexual and sexual GA regarding the magnitude of error thresholds.

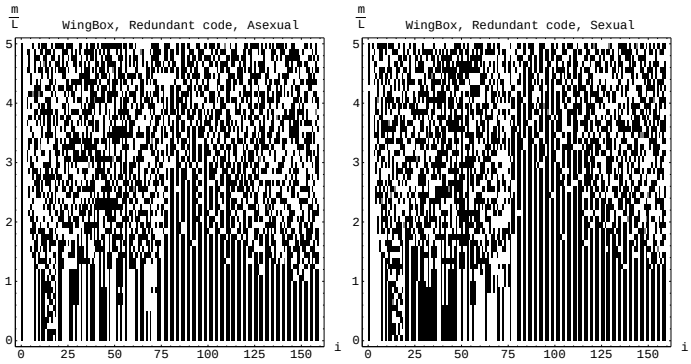


Figure 11: Error thresholds on the Wing-Box problem for both asexual and sexual GAs. The vertical axis shows mutation rates per genotype ($m/L$).

### 5.2 Multiple Knapsack Problem

Four multiple-knapsack instances, taken from the literature, were selected as test problems. Problem sizes ranged from 50 to 105 objects and from 2 to 30 knapsacks. Table 3 summarizes the problem instances tested. These (and several other) problems are available online from the OR-library (Beasley, 1990).

| Instance | $Objects$ | $Sacks$ |
|---|---|---|
| Weish 12 | 50 | 5 |
| Sento 1 | 60 | 30 |
| Weing 7 | 105 | 2 |
| Weish 30 | 90 | 5 |

Table 3: Multiple Knapsack problem instances tested.

Figure 12 shows the consensus sequence plots on the four Knapsack instances selected, asexual (top) and sexual (bottom). Results on the four instances confirm the existence of error thresholds on this real-world application. The error threshold is again visualized as the transition from a stable consensus sequence to a more randomized sequence of bits. The transition in all the instances occurs at approximately 1.0 – 1.2 mutations per genotype. Notice that in all instances there are some regions of the genotype where the consensus sequence is more stable for mutation rates beyond the error threshold. There are no clear differences between the GA with and without recombination regarding the magnitude of error thresholds. However, the transitions looks sharper, and thus the consensus sequences less stable, for the GA without recombination (asexual). This may be due to the use of two-point recombination. It is known that two-point recombination is a less disruptive operator than mutation alone or uniform recombination.

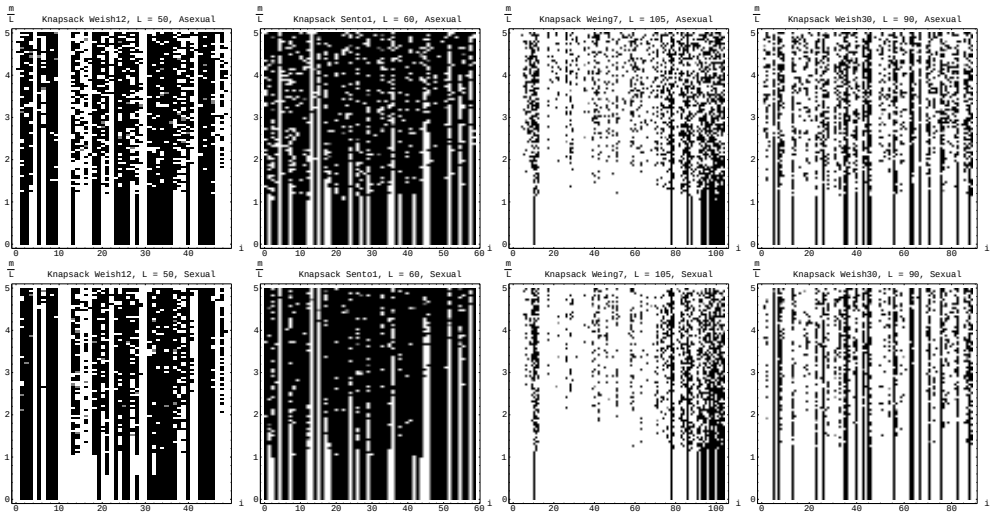


Figure 12: Error thresholds on four instances of the Multiple Knapsack problem. Results for asexual (top) and sexual (bottom) GAs are presented. The vertical axis shows mutation rates per genotype ($m/L$).

### 5.3 Discussion

This closing empirical section explored error thresholds on real-world domains. Results show that error thresholds can also be found on these two complex real-world applications. It should be noticed, however, that other real-world applications might have very different characteristics. No major differences were noticed in the magnitude of error thresholds on GAs with and without recombination. In all scenarios, for the particular GA selected: tournament selection (tournament size of 2), population size of 100, and generational replacement, the error threshold was located at approximately 1.0 – 1.5 mutations per genotype.

## 6 Conclusions

This study verifies the occurrence of error thresholds in evolving populations of bit strings using a GA (with and without recombination). Error thresholds were observed on several landscapes, including real-world domains. In this way, the notion of error threshold, (already introduced for very simple landscapes in (Ochoa and Harvey, 1998)) is brought to evolutionary computation.

We also described the *consensus sequence plots*, firstly introduced in (Ochoa, 2000). These plots, borrowed and adapted from theoretical biology (Bonhoeffer and Stadler, 1993), are new to the evolutionary computation community. They represent a novel way to visualize the structure of fitness landscapes, since features such as the presence of steps or discontinuities can be noticed. Moreover, the degree of ruggedness in a landscape was revealed by these plots. Consensus sequence plots may also serve as a tool to differentiate critical (and less critical) areas in the genotype, which may have practical implications when tackling real-world problems. First, it may be possible to infer important knowledge about an applied problem. Second, it may be possible to refine the genotype representations and optimal schedules for mutation rates. This may be possible on some classes of problems, as for instance the Wing-Box and Knapsack prob-

lems, where producing the consensus sequence plot took few hours (on a standard Sun SPARC Station). However, consensus sequence plots are computationally expensive and may be infeasible for other present-day challenging problems.

The major lesson learned from this study is that error thresholds depend mainly on the selection pressure and the genotype length, regardless of the landscape under study, as long as the landscape is rugged. This knowledge may suggest useful heuristics for setting near-optimal mutation rates. In particular, the suggestion of setting a mutation rate of $1/L$ (one mutation per genotype), is supported by the experiments in this study, but only on rugged landscapes, population sizes greater than 50, and selection schemes imposing a selection pressure similar to that of tournament selection with a tournament size of 2. This figure (a mutation rate of $1/L$) has appeared several times in the evolutionary computation literature. The earliest appearance we can trace back was due to Bremerman (1966) as quoted by Bäck" (1996) . Also, De Jong (1975) suggested this value as quoted by Hesser and Männer (1991) . The work of Mühlenbein (1992) states that $p_m = 1/L$ is optimal for general unimodal functions. This setting has also produced good results for several NP-hard combinatorial optimization problems such as the multiple knapsack problem (Khuri et al., 1994), the minimum vertex cover problem (Khuri and Bäck, 1994), and the maximum independent set problem (Bäck and Khuri, 1994). The work of Smith and Fogarty found $1/L$ as the best fixed setting for the mutation rate, giving results comparable to their best self-adaptive method. Other authors have found a dependence of effective mutation rates upon the string length $L$, although they had not explicitly suggested $p_m = 1/L$ (Schaffer et al., 1989; Hesser and Männer, 1991; Bäck, 1992; Bäck, 1993). Finally, The $1/L$ heuristic is most probably applicable on landscapes with little or no redundancy. As suggested by Harvey and Thompson (1996) , in the presence of redundancy or 'junk' this heuristic should be adjusted so as to give an expected 1 mutation per *non-redundant* part of the genotype.

## 7 Acknowledgements:

Thanks are due to Dr. Inman Harvey who introduced me to the notion of error thresholds and neutral networks, and who has been rather encouraging and stimulating both as a PhD advisor, and later as a collaborator.

## References

Bäck, T. (1992). The interaction of mutation rate, selection, and self-adaption within a genetic algorithm. In und R. Manderick, B. M., editor, *Parallel Problem Solving from Nature II*, pages 85–94, Amsterdam. Elsevier.

Bäck, T. (1993). Optimal mutation rates in genetic search. In Forrest, S., editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 2–8, San Mateo, CA. Morgan Kaufmann.

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*. The Clarendon Press Oxford University Press, New York.

Bäck, T. and Khuri, S. (1994). An evolutionary heuristic for the maximum independent set problem. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, pages 531–535. IEEE Press.

Beasley, J. E. (1990). OR-library: Distributing test problems by electronic mail. *Journal of the Operational Research Society*, 41(11):1069–1072. Also available at http://www.ms.ic.ac.uk/info.html.

Bonhoeffer, S. and Stadler, P. (1993). Error thresholds on correlated fitness landscapes. *Journal of Theoretical Biology*, 164:359–372.

Bremerman, H., Rogson, M., and Salaff, S. (1966). Global properties of evolution processes. In *Natural Automata and Useful Simulations*, pages 3–41. Spartan.

Eigen, M. and Schuster, P. (1979). *The Hypercycle: A Principle of Natural Self-Organization*. Springer-Verlag.

Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary robotics: The Sussex approach. *Robotics and Autonomous Systems*, 20:205–224.

Harvey, I. and Thompson, A. (1996). Through the labyrinth evolution finds a way: A silicon ridge. In *Proceedings of the First International Conference on Evolvable Systems: From Biology to Hardware*. Springer-Verlag.

Hesser, J. and Männer, R. (1991). Towards an optimal mutation probability for genetic algorithms. In Schwefel, H. P. and Männer, R., editors, *Parallel Problem Solving from Nature*. Springer-Verlag, Lecture Notes in Computer Science Vol. 496.

Kamp, C., Wilke, C., Adami, C., and Bornholdt, S. (2003). Viral evolution under the pressure of an adaptive inmune system: Optimal mutation rates for viral escape. *Complexity*, 8(2):28–33.

Kauffman, S. (1989). Adaptation on rugged fitness landscapes. In Stein, D., editor, *Lectures in the Sciences of Complexity*, pages 527–618. Addison-Wesley, Reading, MA.

Kauffman, S. (1993). *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press.

Khuri, S. and Bäck, T. (1994). An evolutionary heuristic for the minimum vertex cover problem. In Hopf, J., editor, *Genetic Algorithms within the Framework of Evolutionary Computation*, pages 86–90, Saarbrücken, Germany. Max-Planck-Institut für Informatik.

Khuri, S., Bäck, T., and Heitkötter, J. (1994). The zero/one multiple knapsack problem and genetic algorithms. In Deaton, E., Oppenheim, D., Urban, J., and Berghel, H., editors, *Proceedings of the 1994 ACM Symposium of Applied Computation*, pages 188–193. ACM Press.

McIlhagga, M., Husbands, P., and Ives, R. (1996). A comparison of search techniques on a wing-box optimisation problem. *Lecture Notes in Computer Science*, 1141.

Mitchell, M., Forrest, S., and Holland, J. H. (1992). The Royal Road for genetic algorithms: Fitness landscapes and GA performance. In Varela, F. J. and Bourgine, P., editors, *Proceedings of the First European Conference on Artificial Life. Toward a Practice of Autonomous Systems*, pages 245–254, Paris, France. MIT Press, Cambridge, MA.

Mühlenbein, H. (1992). How genetic algorithms really work: I. mutation and hillclimbing. In Männer, B. and Manderick, R., editors, *Parallel Problem Solving from Nature II*, pages 15–25. North-Holland.

Nowak, M. and Schuster, P. (1992). What is a quasispecies? *TREE*, 7:118–121.

Ochoa, G. (2000). Consensus sequence plots and error thresholds: Tools for visualising the structure of fitness landscapes. In Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, H.-P. S., editor, *Parallel Problem Solving from Nature - PPSN VI 6th International Conference*, pages 129–138, Paris, France. Springer Verlag. LNCS 1917.

Ochoa, G. and Harvey, I. (1998). Recombination and error thresholds in finite populations. In Banzhaf, W. and Reeves, C., editors, *Foundations of Genetic Algorithms 5*. Morgan Kauffman.

Ochoa, G., Harvey, I., and Buxton, H. (1999). Error thresholds and their relation to optimal mutation rates. In Floreano, D., Nicoud, J.-D., and Mondada, F., editors, *Proceedings of the 5th European Conference on Advances in Artificial Life (ECAL-99)*, volume 1674 of *LNAI*, pages 54–63, Berlin. Springer.

Peliti, L. (2002). Quasispecies evolution in general mean-field landscapes. *Europhysics Letter*, 57(5):745–751.

Schaffer, J., Caruana, R., Eshelman, L., and Das, R. (1989). A study of control parameters affecting online performance of genetic algorithms for function optimization. In Schaffer, J. D., editor, *Proceedings of the Third International Conference on Genetic Algorithms*, San Mateo CA. Morgan Kaufmann.

Schuster, P. (1994). Extended molecular evolutionary biology: Artificial life bridging the gap between chemestry and biology. *Artificial Life*, 1:39–60.

Smith, J. E. and Fogarty, T. C. (1996). Self adaptation of mutation rates in a steady state genetic algorithm. In *Proceedings of the IEEE Conference on Evolutionary Computation*, pages 318–323, New York. IEEE Press.

van Nimwegen, E. and Crutchfield, J. P. (1998). Optimizing epochal evolutionary search: Population-size dependent theory. Technical Report Preprint 98-06-046, Santa Fe Institute.

Vassilev, V. K., Miller, J. F., and Fogarty, T. C. (1999). Digital circuit evolution and fitness landscapes. In *1999 Congress on Evolutionary Computation*, pages 1299–1306, Piscataway, NJ. IEEE Service Center.