# The Dynamics of a Genetic Algorithm on a Model Hard Optimization Problem

**Alex Rogers**
**Adam Prügel-Bennett**
*Image, Speech, and Intelligent Systems Research Group,*
*Department of Electronics and Computer Science,*
*University of Southampton,*
*Highfield, Southampton, SO17 1BJ, England*

A model of a hard optimization problem suggested in the literature is considered. The dynamics of a genetic algorithm (GA) using ranking selection, mutation, and uniform crossover are completely modeled on this problem and generalized to any symmetrical concave function of unitation. Full finite population effects are taken into account allowing a novel analytical comparison of roulette wheel and stochastic universal sampling. Closed form expressions are derived for the equilibrium population distribution of this model. The first passage time to move from a local to a global minimum in a two-potential well landscape is calculated. A comparison is made with a stochastic hill climber and a GA without crossover. The GA with crossover is shown to perform orders of magnitude faster giving some insights into the nature of GA search and the crossover operator.

## 1. Introduction

Much of the theoretical modeling of genetic algorithms (GAs) has necessarily been performed on simple models of the GA on a simple fitness landscape. Examples of these are functions of unitation such as one-max [7, 8, 9, 14] and the royal road functions of Mitchell, Holland, and Forest [6]. Whilst giving some insights into the behavior of the GA, they are not representative of the sort of optimization problems to which GA are often applied. GA are shown to perform poorly on these model landscapes compared to other optimization techniques such as stochastic hill climbing, whilst there remains a body of empirical evidence that GA often out perform or at least perform comparatively well with other optimization methods, on real-world problems.

In order to gain some insight into the performance of GA on real-world optimization problems, it is necessary to consider harder problem spaces. Characterizing the hardness of a problem has been an active area of research and has led to measures of problem difficulty such as Fitness Distance Correlation [5] and Epistasis Variance [4]. There is

however much debate as to the interpretation of these measures and their applicability to real-world fitness landscapes.

Despite these arguments, there is some consensus as to what features make a problem space hard to search. There may be many local minima. These local minima may be separated by a potential barrier from better solutions, resulting in the need for nonlocal search steps. If these local minima occupy the majority of the search space it may take a long time to generate the moves necessary to fall into the basin of attraction of the global minimum.

In this paper we analyze a previously published model of a hard optimization problem [12] which addresses the last two instances. The problem is known as the "Basin with a Barrier" fitness landscape and is a function of unitation. It has some of the features of a hard optimization problem but is still amenable to analysis.

The landscape consists of a large local minimum separated from the global minimum by a potential barrier. We consider a series of $L$ spins whose value may be 1 or $-1$ and consider the total magnetization $M$ of the string,

$$M = \sum_{i=1}^{L} S_i. \tag{1}$$

The potential or fitness, which we are trying to minimize, is a function of this magnetization and is given generically as,

$$V(M) = \begin{cases} (M - M_l)^2 + V_l & \text{if} \quad M \le M_b \\ 0 & \text{if} \quad M > M_b. \end{cases} \tag{2}$$

The entropy of the system $S$ is such that the number of states in the global minimum is much smaller than that in the local minimum whilst the maximum entropy state is some distance from both the local and global minima. Figure 1 shows the landscape schematically.

We propose that whilst being a toy problem, this model holds some of the features seen in combinatorial optimization problems such as the

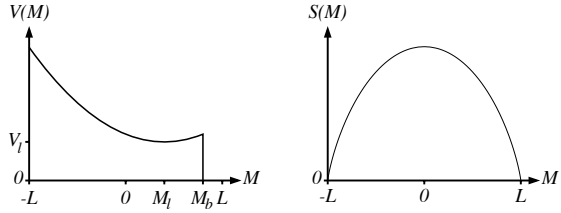


**Figure 1.** Diagram of potential and entropy for the Basin with a Barrier fitness landscape.

traveling salesman. We expect random search to produce poor solutions most of the time and we expect these poor solutions to occupy the majority of the problem space. We expect good solutions to be near one another in problem space but they may be separated by nonlocal moves.

Shapiro and Prügel-Bennett [12] analyzed this landscape using a statistical mechanics formalism of a GA under Boltzmann selection and uniform crossover. They showed that in the infinite population limit there is a phase in which the population will move from the local minimum to the global minimum from any initial configuration. The analysis suggests that this can occur orders of magnitude faster than a stochastic hill climber can find the global minimum. In order to obtain qualitative results, much of the complexities of modeling the dynamics of the GA were omitted and the fit between theoretical and simulation results was poor.

In this paper we extend the analysis and consider the aspect not touched on by the earlier paper, the time required for a GA to find one solution in the global minimum, or the first passage time. We extend the formalism used in the earlier paper to ranking selection as opposed to Boltzmann selection. Apart from its more common use in real GA, the use of ranking selection leads to significant gains. We can solve the dynamics of ranking selection on the one-max fitness landscape by considering just the mean and variance of the population distribution. As the ranking of population members for selection removes the need to know the exact structure of the population and the fitness landscape, we can extend this analysis of selection to any symmetrical concave function of unitation. By considering finite population effects on selection, we are able to analytically compare and model the two most common selection strategies, roulette wheel and stochastic universal sampling (SUS), or Baker selection [2]. Finally, by considering the correlation of the population through selection, the effect of crossover can be accurately modeled.

The result of this is an accurate model of the dynamics of the GA on any symmetrical concave function of unitation. We are able to use this to predict the first passage time to find a solution in the global minimum of the Basin with a Barrier fitness landscape. This time is shown to be orders of magnitude faster than that for a stochastic hill climber and that of the GA without crossover. This finding gives some insight into the search power of crossover.

In section 2 we present the model GA and in section 3 discuss the formalism used to model the dynamics. Section 4 details the effect of ranking selection including finite population effects and a comparison between roulette wheel selection and SUS. Section 5 details the modeling of the other genetic operators, mutation and crossover. Section 6 reviews the accuracy of the model compared to simulation results and

section 7 discusses the calculation of first passage time for the GA and a stochastic hill climber on the Basin with a Barrier problem. The results are discussed in section 8. In the Appendix we present, for completeness, the analysis of the effect of crossover and mutation operators.

## 2. The model genetic algorithm

The model GA consists of a population of $P$ individuals. Each individual consists of a string of $L$ bits whose value may be 1 or $-1$. The magnetization of any individual $\mu$ is given by the sum of its bits,

$$M_\mu = \sum_{i=1}^{L} S_i^\mu \qquad \text{where } S_i^\mu = \{-1, 1\}. \tag{3}$$

The potential which we are trying to minimize is given by the generic expression from section 1. The population is initialized with random strings.

We consider a generational GA and use ranking selection [1] to select the mating pool from the initial population. The mutation operator is then applied whereby each bit has a small probability of mutation,

$$S_i^\mu \to -S_i^\mu \qquad \text{with probability } \gamma. \tag{4}$$

We then apply the crossover operator to the mating pool to produce the next generation. Population members are randomly paired and uniform crossover [13] applied whereby bits are randomly drawn from each parent,

$$S_i^\mu = \chi_i^\mu S_i^\alpha + \left(1 - \chi_i^\mu\right) S_i^\beta, \tag{5}$$

where

$$\chi_i^\mu = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2. \end{cases} \tag{6}$$

The complementary offspring are also created and the pair replaces their parents.

## 3. The statistical physics formalism

Rather than model the exact structure of the population, we use a statistical mechanics technique and model several macroscopic variables of the ensemble average of an infinite number of finite populations. In this way we model the mean behavior of an evolving finite population. The formalism used here was first developed by Prügel-Bennett and Shapiro in an analysis of a GA using Boltzmann selection [8, 9]. The macroscopic variables used are the cumulants of the population magnetization

distribution. With Boltzmann selection it was found that at least four cumulants were needed to accurately model the dynamics of the GA due to finite population effects.

When modeling ranking selection with uniform crossover we find that the higher order cumulants are sufficiently small that just the first two cumulants give a good degree of accuracy. For a finite population these are given by

$$\begin{aligned} \kappa_1 &= \langle M \rangle \\ \kappa_2 &= \langle M^2 \rangle - \langle M \rangle^2, \end{aligned} \tag{7}$$

where $\langle \ldots \rangle$ represents the average over the population. The cumulants of the ensemble we denote as $K_1$ and $K_2$. Due to the well-known finite sampling effect, the cumulants of any particular finite population of size $P$ drawn from the ensemble will differ slightly from those of the ensemble. The two terms are related by the expressions

$$\begin{aligned} \kappa_1 &= K_1 \\ \kappa_2 &= \left(1 - \frac{1}{P}\right) K_2. \end{aligned} \tag{8}$$

In section 6, when crossover is considered, another macroscopic variable is introduced, the correlation of the population due to the replication of individuals through selection. Thus in the final analysis we use just three macroscopic variables to fully describe the state of the ensemble distribution.

## 4. Selection

Ranking selection [1] is a commonly used form of GA selection as it prevents a single highly fit individual from taking over the population and causing premature convergence. The expected number of times that the population member of rank $i$ will be represented in the next generation is controlled by the parameter MAX and is given by

$$E_i = \mathrm{MAX} - 2\,(\mathrm{MAX} - 1)\,\frac{i-1}{P-1}. \tag{9}$$

The fittest population member is expected to be represented MAX times and the least fit (2 – MAX) times. MAX may take any value between 1 and 2.

The effect of ranking selection on a population is found by first considering the infinite population case.

### 4.1 Infinite population model

We first consider the simpler case of the one-max landscape and an infinite population size. Whilst mutation and crossover operate on the

magnetization distribution we must consider the function of unitation when calculating the effect of selection. In the case of one-max the ranking of an individual with magnetization $M$ is solely related to its position within the population. The expression for the expected number of occurrences of any population member given in equation (9) can be rewritten in terms of the population magnetization distribution. We can write this as

$$E_M = (2 - \mathrm{MAX}) + 2\,(\mathrm{MAX} - 1) \int_{-\infty}^{M} \rho\left(M'\right) \mathrm{d}M', \tag{10}$$

where $\rho(M')$ describes the magnetization distribution. The first and second moments of the population distribution after selection are found by integrating the weighting over the ensemble distribution. This distribution is described as a gaussian with mean $K_1$ and variance $K_2$. The resulting moments are

$$\begin{aligned}
\langle M \rangle &= \int_{-\infty}^{\infty} M E_M \rho\,(M)\,\mathrm{d}M \\
&= K_1 + (\mathrm{MAX} - 1)\sqrt{\frac{K_2}{\pi}} \\
\langle M^2 \rangle &= \int_{-\infty}^{\infty} M^2 E_M \rho\,(M)\,\mathrm{d}M \\
&= K_2 + K_1^2 + 2\,(\mathrm{MAX} - 1)\,K_1 \sqrt{\frac{K_2}{\pi}}.
\end{aligned} \tag{11}$$

Thus using the expressions in equation (7), the cumulants after selection are given by

$$\begin{aligned}
\langle K_1 \rangle_s &= K_1 + (\mathrm{MAX} - 1)\sqrt{\frac{K_2}{\pi}} \\
\langle K_2 \rangle_s &= \left[1 - \frac{(\mathrm{MAX} - 1)^2}{\pi}\right] K_2,
\end{aligned} \tag{12}$$

where $\langle \ldots \rangle_s$ represents the average over all ways of performing selection. Clearly the variance is decreased by a factor determined by the selection pressure whilst the mean increases by a factor dependent on $\sqrt{K_2}$, a measure of the width of the distribution. The analysis of ranking selection for an infinite population when MAX = 2 is identical to that of tournament selection which has previously been performed by Blickle and Thiele [3]. Changing the parameter MAX is equivalent to introducing a probabilistic element into tournament selection.

When using ranking selection, the exact function of unitation and population structure need not be considered as only the rank within the population is significant. For a symmetrical concave fitness landscape

with a minimum at $M_l$, the weighting of a population member whose magnetization is $M$ is related to not only its position within the population magnetization distribution but also the position of the minimum $M_l$ and can be written as

$$E_M = \begin{cases} \text{MAX} - 2\,(\text{MAX} - 1)\int_M^{2M_l - M} \rho\,(M')\,\mathrm{d}M' & \text{when } M \le M_l \\ \text{MAX} - 2\,(\text{MAX} - 1)\int_{2M_l - M}^{M} \rho\,(M')\,\mathrm{d}M' & \text{when } M \ge M_l. \end{cases} \tag{13}$$

The calculation of the cumulants after selection is similar to that carried out before but involves rather more algebra. The resulting expressions are

$$\begin{aligned} \langle K_1 \rangle_s &= K_1 + (\text{MAX} - 1)\sqrt{\frac{K_2}{\pi}}\,\text{erf}\left(\frac{M_l - K_1}{\sqrt{K_2}}\right) \\ \langle K_2 \rangle_s &= \left[1 - \frac{2\,(\text{MAX} - 1)}{\pi}\exp\left(-\frac{(M_l - K_1)^2}{K_2}\right)\right. \\ &\qquad \left. -\frac{(\text{MAX} - 1)^2}{\pi}\text{erf}^2\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)\right] K_2. \end{aligned} \tag{14}$$

Here $\text{erf}\,(x)$ represents the standard error function. Clearly when $M_l$ is sufficiently large little of the distribution falls over the minimum and the dynamics will be identical to the case of the one-max landscape. In this case the expressions simplify to those presented earlier.

### 4.2 Finite population effects

The preceding expressions describe the infinite population response to selection. A real GA however has a finite population and the stochastic nature of the selection operator leads to a deviation from the infinite population response. Whilst the first cumulant is unaffected by finite population effects, the effects on the second cumulant is significant and gives rise to genetic drift and performance differences between roulette wheel selection and SUS [2].

To calculate directly the effect of selection on a finite population is not possible. Instead we make an approximation and consider that selection operates on an infinite population from which we sample a finite population. We consider the change in variance to be due to the product of the infinite population result and a factor which describes the stochastic nature of the selection scheme. We find this factor by considering selection from a finite population independent of magnetization. In this way the factor which we calculate is a measure of the genetic drift inherent in the selection scheme and has been used by the authors in another paper as a basis for the comparison of genetic drift in steady state GA with varying generation gaps [11].

If we consider a finite population the second cumulant $\kappa_2$ is given by

$$\kappa_2 = \frac{1}{P}\sum_{\alpha=1}^{P} M_\alpha^2 - \left(\frac{1}{P}\sum_{\alpha=1}^{P} M_\alpha\right)^2. \tag{15}$$

Separating out terms which are not independent gives

$$\kappa_2 = \sum_{\alpha=1}^{P}\left(\frac{1}{P} - \frac{1}{P^2}\right)M_\alpha^2 - \frac{1}{P^2}\sum_{\alpha\neq\beta} M_\alpha M_\beta. \tag{16}$$

We now apply some selection scheme to this population and draw from it a new population of $P$ individuals. In this new population there are now $n_\alpha$ copies of population member $M_\alpha$ and the second cumulant is given by

$$\kappa_2' = \frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha M_\alpha^2 - \left(\frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha M_\alpha\right)^2. \tag{17}$$

We can separate out terms which are independent to give

$$\kappa_2' = \sum_{\alpha=1}^{P}\left(\frac{n_\alpha}{P} - \frac{n_\alpha^2}{P^2}\right)M_\alpha^2 - \sum_{\alpha\neq\beta}\frac{n_\alpha n_\beta}{P^2} M_\alpha M_\beta. \tag{18}$$

Averaging over all ways of performing the sampling gives

$$\kappa_2' = \left(\frac{\langle n\rangle}{P} - \frac{\langle n^2\rangle}{P^2}\right)\sum_{\alpha=1}^{P} M_\alpha^2 - \frac{\langle n_\alpha n_\beta\rangle}{P^2}\sum_{\alpha\neq\beta} M_\alpha M_\beta. \tag{19}$$

Our selection scheme keeps the population size constant so $\langle n\rangle = 1$ and we have the following identity,

$$\left(\sum_{\alpha=1}^{P} n_\alpha\right)^2 = P^2 = \sum_{\alpha=1}^{P} n_\alpha^2 + \sum_{\alpha\neq\beta} n_\alpha n_\beta \tag{20}$$

and thus

$$\langle n_\alpha n_\beta\rangle = \frac{P - \langle n^2\rangle}{P-1}. \tag{21}$$

Using this result gives the expression

$$\kappa_2' = \frac{P - \langle n^2\rangle}{P-1}\left[\sum_{\alpha=1}^{P}\left(\frac{1}{P} - \frac{1}{P^2}\right)M_\alpha^2 - \frac{1}{P^2}\sum_{\alpha\neq\beta} M_\alpha M_\beta\right]. \tag{22}$$

The expression in the square brackets is simply the variance of the initial population and thus

$$\kappa_2' = \frac{P - \langle n^2\rangle}{P-1}\kappa_2. \tag{23}$$

This factor describes the loss in variance when a finite population is sampled by some selection scheme independent of magnetization and is simply dependent on $\langle n^2 \rangle$, that is, a measure of the variance in the number of times any population member is selected. We can combine this result with the infinite population result to give the variance after selection of a finite population

$$\langle K_2 \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} \left[ 1 - \frac{2\,(\mathrm{MAX} - 1)}{\pi} \exp\left( - \frac{\left(M_l - K_1\right)^2}{K_2} \right) - \frac{(\mathrm{MAX} - 1)^2}{\pi} \mathrm{erf}^2 \left( \frac{M_l - K_1}{\sqrt{K_2}} \right) \right] K_2. \tag{24}$$

### 4.3 Calculating $\langle n^2 \rangle$

The finite population effect of the selection scheme calculated above was shown to be entirely dependent on the value of $\langle n^2 \rangle$. Baker [2] noted that whilst any individual with rank $i$ is expected to occur $E_i$ times after selection, the stochastic nature of roulette wheel selection allows anywhere between 0 and $P$ copies to be selected. This is the source of convergence of a finite population due to stochastic effects, that is, genetic drift.

Baker proposed SUS as a selection scheme which limits the range of possible occurrences to either $\lfloor E_i \rfloor$ ($E_i$ rounded down to the nearest integer) or $\lceil E_i \rceil$ ($E_i$ rounded up to the nearest integer). Whilst no arguments were made as to the virtue of doing this in the original paper, it is generally understood that the use of SUS reduces the effects of genetic drift. Intuitively we can see that limiting the range of possible occurrences will reduce $\langle n^2 \rangle$ and hence reduce the loss in variance through stochastic effects. However, using the expression derived above we can calculate this difference and formally compare them.

#### 4.3.1 Roulette wheel selection

Using roulette wheel selection gives rise to a binomial distribution in which we make $m$ trials with a probability of success $p$. The standard result for a binomial distribution is

$$\langle n^2 \rangle = m\,(m - 1)\,p^2 + mp. \tag{25}$$

Assuming independence between each population member, $\langle n^2 \rangle$ can be found by averaging over the weighting for each rank. If the probability of success is $E_i/P$ and we make $P$ trials, the resulting expression is

$$\langle n^2 \rangle = \frac{1}{P} \sum_{i=1}^{P} \left[ P\,(P - 1)\,\frac{E_i^2}{P^2} + P \frac{E_i}{P} \right]. \tag{26}$$

Using the expression for $E_i$ given in equation (9) and performing the summation gives

$$\langle n^2 \rangle = 3 + \frac{(P+1)(\mathrm{MAX}^2 - 2\mathrm{MAX} - 2)}{3P}. \tag{27}$$

#### 4.3.2 Stochastic universal sampling

In SUS, for any value of $E_i$ we have either $\lfloor E_i \rfloor$ or $\lceil E_i \rceil$ copies after selection. The probabilities of either are given by

$$n_i = \begin{cases} \lceil E_i \rceil & \text{with probability } E_i - \lfloor E_i \rfloor \\ \lfloor E_i \rfloor & \text{with probability } \lceil E_i \rceil - E_i. \end{cases} \tag{28}$$

Assuming independence between individuals we can find $\langle n^2 \rangle$ by calculating the distribution averaged over each ranking

$$\langle n^2 \rangle = \frac{1}{P} \sum_{i=1}^{P} \lceil E_i \rceil^2 \left( E_i - \lfloor E_i \rfloor \right) + \lfloor E_i \rfloor^2 \left( \lceil E_i \rceil - E_i \right). \tag{29}$$

Calculating this for most selection schemes is far from trivial as the exact population fitness distribution is required. However for ranking selection $E_i$, and thus $\lfloor E_i \rfloor$ and $\lceil E_i \rceil$, are known independently of the population structure. For $i \leq P/2$, $\lfloor E_i \rfloor = 1$ and $\lceil E_i \rceil = 2$ whilst when $i > P/2$, $\lfloor E_i \rfloor = 0$ and $\lceil E_i \rceil = 1$. Applying these ranges gives

$$\langle n^2 \rangle = \frac{1}{P} \sum_{i=1}^{P/2} 4 \left( E_i - 1 \right) + \left( 2 - E_i \right) + \frac{1}{P} \sum_{i=(P/2)+1}^{P} E_i. \tag{30}$$

Again using the expression for $E_i$ given in equation (9) and performing the summation gives

$$\langle n^2 \rangle = \mathrm{MAX} - \frac{(\mathrm{MAX} - 1)}{2} \frac{P-2}{P-1}. \tag{31}$$

### 4.4 Discussion

For any particular value of MAX, $\langle n^2 \rangle$ for SUS is less than that of roulette wheel selection, showing that the variance in the number of times any population member is selected is less. This results in a smaller reduction of population magnetization variance at each selection step and thus slower convergence of the population through stochastic effects. Under neutral selection where MAX = 1, SUS shows no loss in population variance whilst under roulette wheel selection variance still decreases at each generation by a factor of $(1 - 1/P)$, the rate of genetic drift. This analysis confirms the generally held beliefs about SUS and roulette wheel selection.

Figure 2 shows a comparison of simulation results with these theoretical results. A population of size $P = 2$, 10, and 100 was initially

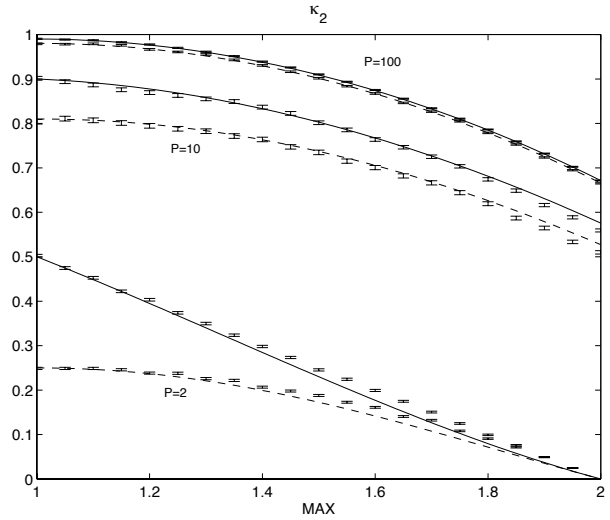


**Figure 2**. Comparison of theoretical and simulation results for $\kappa_2$ after roulette wheel selection (dashed line) and SUS (solid line). Population size $P = 2$, 10, and 100. In each case, SUS is the curve with the least change in variance.

assigned a fitness drawn from a unit gaussian ($K_2 = 1$) and then selection applied once. The variance of the population after selection was calculated and averaged over 1000 runs. Both roulette wheel and SUS are shown with SUS consistently experiencing less loss in variance through selection.

The finite sampling effect described in section 2 is clearly seen in Figure 2 as we are plotting $\kappa_2$, that is, the variance of a finite population. The variance of the initial finite population drawn from the unit gaussian is $(1 - 1/P)$ and not 1 as may be expected.

The theoretical results are in good agreement with the simulation results even at extremes of small population sizes where finite population effects are dominant. Our assumption that we can approximate the effect of a finite population simply by considering the product of the two factors is clearly justified. Our calculation of $\langle n^2 \rangle$ for SUS also assumed independence between population members. This approximation leads to another slight deviation from the simulation data but the fit is generally good.

## 5. Genetic operators

The search power of a GA is determined by how the selection scheme interacts with other genetic operators such as mutation and crossover. To fully understand the dynamics of the GA we must model these operators.

### 5.1 Mutation

When the mutation operator is applied, each bit has a small probability of mutation, $\gamma$. The analysis of the effect of mutation on the cumulants of the ensemble magnetization distribution has previously been performed by Prügel-Bennett and Shapiro [8, 9]. We present the results here and, for completeness, present the derivation in the notation used here in appendix A. The first and second cumulant after mutation are

$$\begin{aligned} \langle K_1 \rangle_m &= \Gamma K_1 \\ \langle K_2 \rangle_m &= \Gamma^2 K_2 + L\left(1 - \Gamma^2\right) \qquad \text{where } \Gamma = 1 - 2\gamma \end{aligned} \tag{32}$$

and $\langle \ldots \rangle_m$ represents the average over all ways of performing mutation.

The effect of mutation can be clearly seen in these expressions. It acts to push the ensemble distribution back to the maximum entropy state, decreasing the mean and increasing the variance. In this way it acts against selection, which is reducing the variance and increasing the mean magnetization.

### 5.2 Crossover

Like mutation, the analysis of uniform crossover has been performed previously by Prügel-Bennett and Shapiro [8, 9]. Details of the derivation are included in appendix B. The effects of crossover on the first two cumulants are

$$\begin{aligned} \langle K_1 \rangle_X &= K_1 \\ \langle K_2 \rangle_X &= \frac{K_2}{2} + \frac{L}{2}\left(1 - q\right), \end{aligned} \tag{33}$$

where $q$ is defined as

$$q = \frac{1}{P\,(P-1)} \sum_{\alpha \neq \beta} \frac{1}{L} \sum_{i=1}^{L} S_i^{\alpha} S_i^{\beta}. \tag{34}$$

We call $q$ the *correlation* of the population, as it describes the similarity of strings. In the maximum entropy state, the correlation of the population is 0. In a population consisting of $P$ identical strings, correlation is equal to 1.

The first cumulant does not change under crossover. This is expected as crossover conserves the states of all the bits. Crossover does change the second cumulant however and forces the variance towards the natural variance of the population which is defined by the value of $q$.

The analysis of higher cumulants show that they are greatly reduced by uniform crossover. It is this feature along with ranking selection which allows us to accurately model the dynamics with just two cumulants.

To fully understand the effect of crossover, we must understand how correlation $q$ changes with selection and mutation. As $S_i^\alpha$ and $S_i^\beta$ are not independent, taking the average of $q$ directly is not trivial. An approach to performing this calculation on a slightly different problem was developed by Magnus Rattray [10]. Here we take a similar approach and describe correlation as a combination of the natural correlation due to the mean magnetization of the population and a contribution due to the existence of multiple copies of individuals within the population, known as the *founder effect*.

We introduce another macroscopic variable which describes the probability that the bits at the same site in two randomly drawn population members originate from the same ancestor. That is,

$$C = \frac{1}{P(P-1)} \sum_\alpha \sum_{\beta \neq \alpha} \frac{1}{L} \sum_i \left[ S_i^\alpha \sim S_i^\beta \right], \tag{35}$$

where

$$\left[ S_i^\alpha \sim S_i^\beta \right] = \begin{cases} 1 & \text{if both bits come from the same ancestor} \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

If the bits originate from the same ancestor, they will be identical and contribute +1 to the correlation. If they originate from different ancestors, we can calculate their contribution to the correlation from the fitness of the two strings. Thus we are able to describe the correlation of the population in terms of this variable and the mean fitness,

$$q = C + (1 - C) \frac{K_1^2}{L^2}. \tag{37}$$

In the initial maximum entropy case $C = 0$ and $q$ is defined by the mean magnetization of the population. If the population consists of identical population members $C = 1$ and the population is perfectly correlated. Selection will act to increase $C$ by introducing multiple copies of individuals whilst mutation will act to reduce $C$. To fully model crossover we must calculate the effect these operators have on $C$.

#### 5.2.1 Selection

If an initial population has a particular value of $C$ its value after selection $C'$ will depend on the probability that a pair drawn from the population for crossover are identical. This is related to the number of copies introduced by selection and the probability that they are drawn from the population together. If $n_\alpha$ is the number of copies of population member $\alpha$, $C'$ is given by

$$C' = \sum_{\alpha=1}^{P} \left[ \frac{n_\alpha}{P} \frac{n_\alpha - 1}{P - 1} + C \frac{n_\alpha}{P} \left( 1 - \frac{n_\alpha - 1}{P - 1} \right) \right]. \tag{38}$$

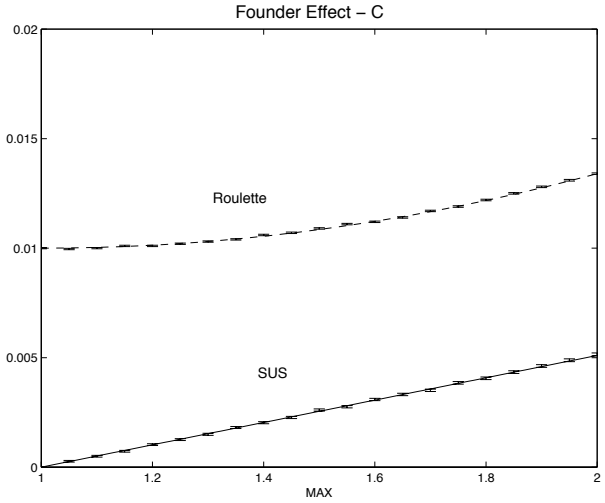


**Figure 3.** Comparison of theoretical and simulation results for $C$ when roulette wheel selection (dashed line) and SUS (solid line) are used with ranking selection. Population size is 100.

Averaging over all $n_\alpha$ and using the fact that population size is constant and $\langle n \rangle = 1$ gives

$$1 - \langle C \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} \big(1 - C\big), \tag{39}$$

where $\langle n^2 \rangle$ is the variance of the selection scheme as calculated in section 3. Indeed, the factor here is exactly the same as the factor describing the loss in population variance after selection.

Figure 3 shows simulation and theoretical results for $C$ after one selection process. An initial population of 100 individuals were created with a fitness assigned arbitrarily. Selection was applied using both SUS and roulette wheel selection and $C$ calculated. Results were averaged over 1000 runs. The difference between the selection schemes is clear. Roulette wheel selection shows a large increase in correlation even at very low selection strengths. Over time this correlation builds up within the population, decreasing the natural variance and thus reducing the effectiveness of crossover in restoring the population variance lost through selection. The GA using roulette wheel selection exhibits a smaller final population variance than one using SUS and thus searches a smaller area of the problem space.

#### 5.2.2 Mutation

The value of $C$ will also change with mutation, as it reduces the degree of replication within the population. By considering the expected value

of individual bits after mutation as derived in appendix A, $\langle S_i \rangle_m = \Gamma S_i$, and the definition of correlation given in equation (34), the correlation after mutation can be shown to be

$$\langle q \rangle_m = \Gamma^2 q. \tag{40}$$

Using the definition of $q$ in equation (37) and the change in the first cumulant due to mutation, $\langle K_1 \rangle_m = \Gamma K_1$. After mutation C is given by

$$\langle C \rangle_m = \frac{\Gamma^2 \left( K_1^2 - L^2 \right)}{\Gamma^2 K_1^2 - L^2} C. \tag{41}$$

## 6. Modeling the dynamics

The result of the analysis is a set of equations describing the change in each macroscopic variable with each operator: selection, mutation, and crossover.

Selection:

$$\langle K_1 \rangle_s = K_1 + (\mathrm{MAX} - 1) \sqrt{\frac{K_2}{\pi}} \mathrm{erf}\left( \frac{M_l - K_1}{\sqrt{K_2}} \right)$$

$$\langle K_2 \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} \left[ 1 - \frac{2\,(\mathrm{MAX} - 1)}{\pi} \exp\left( - \frac{\left( M_l - K_1 \right)^2}{K_2} \right) - \frac{(\mathrm{MAX} - 1)^2}{\pi} \mathrm{erf}^2 \left( \frac{M_l - K_1}{\sqrt{K_2}} \right) \right] K_2$$

$$1 - \langle C \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} \left( 1 - C \right).$$

Mutation:

$$\langle K_1 \rangle_m = \Gamma K_1$$

$$\langle K_2 \rangle_m = \Gamma^2 K_2 + L \left( 1 - \Gamma^2 \right)$$

$$\langle C \rangle_m = \frac{\Gamma^2 \left( K_1^2 - L^2 \right)}{\Gamma^2 K_1^2 - L^2} C.$$

Crossover:

$$\langle K_1 \rangle_x = K_1$$

$$\langle K_2 \rangle_x = \frac{K_2}{2} + \frac{L}{2} \left( 1 - q \right) \qquad \text{where } q = C + \left( 1 - C \right) \frac{K_1^2}{L^2}.$$

By considering an initial population whose bits are assigned randomly ($K_1 = 0$, $K_2 = L$, and $C = 0$) we can iterate these equations to predict the dynamics of the GA. Figure 4 shows the theory predictions compared to

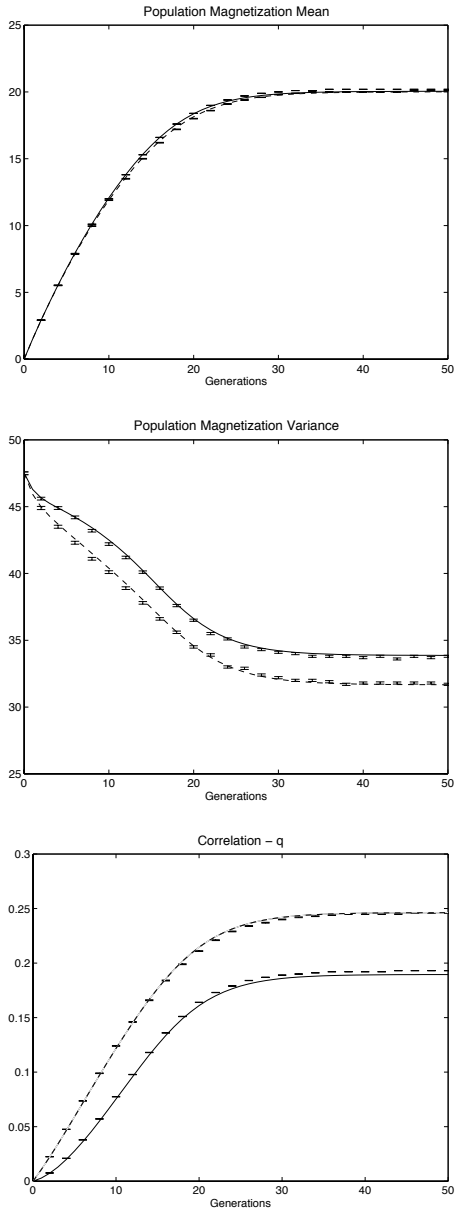


**Figure 4.** Comparison of theoretical and simulation results for the first two cumulants and correlation for a GA with ranking selection, mutation, and crossover. Roulette wheel selection (dashed line) and SUS (solid line) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$, MAX = 1.4, and $M_l = L/2$.

simulation results from repeated runs of a real GA using roulette wheel selection and SUS. The simulation data is averaged over 10 000 runs and uses the parameters $L = 48$, $\gamma = 1/48$, $P = 100$, MAX = 1.4, and $M_l = L/2$. The figures show very good agreement between theory and simulations.

The population evolves until the increase in mean magnetization due to selection is balanced by the decrease due to mutation. The loss in variance due to selection is balanced against the increase due to mutation and crossover. Selection by SUS leads to a final distribution with less correlation and higher population variance than roulette wheel. The mean of both distributions is the same and is close to the minimum at $M_l$. Thus both distributions are centered around the same point but the larger variance in SUS means the GA is continually sampling through crossover a larger area of the problem space.

### 6.1 Equilibrium distribution

To find the equilibrium distribution to which the population evolves was not possible using earlier models due to the dependence on higher order cumulants. The small set of expressions resulting from this model however, enable the equilibrium point to be solved by a set of simultaneous equations. It is not possible to do this exactly analytically but it is easy to do numerically. Figure 5 shows the simulation results and theoretical predictions for the first two cumulants of the equilibrium distribution against changing selection pressure. Simulation results are averaged over 1000 runs.

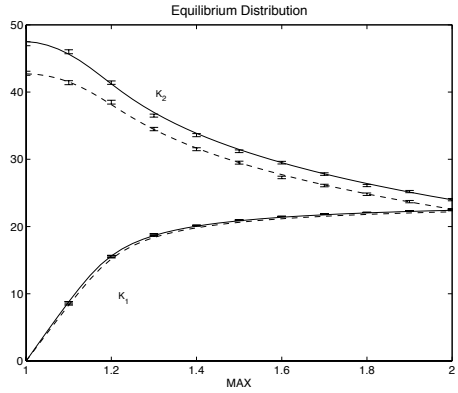


**Figure 5.** Comparison of theoretical numerical solutions and simulation results for equilibrium distribution. Roulette wheel selection (dashed line) and SUS (solid line) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$, and $M_l = L/2$.

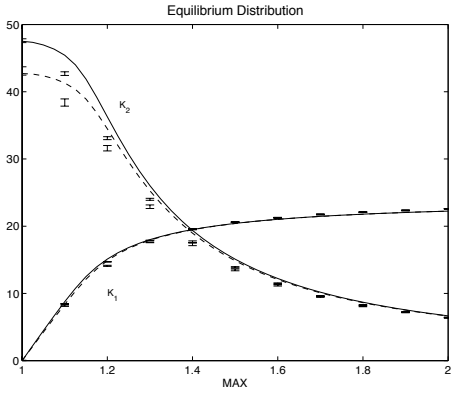


**Figure 6.** Comparison of theoretical numerical solutions and simulation results for equilibrium distribution without crossover. Roulette wheel selection (dashed line) and SUS (solid line) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$, and $M_l = L/2$.

We can also consider the case when just the mutation operator is applied by omitting those macroscopic variables and expressions detailing crossover. Here the model fit is slightly poorer, as without crossover to suppress the higher cumulants, the distribution becomes skewed from a gaussian. However when the distribution is near to the minimum, a gaussian shape is restored and we obtain reasonable agreement without having to calculate higher order cumulant terms. Figure 6 shows the results for the same GA without the crossover operator. We see the same behavior in the mean whilst the variance of the distribution is greatly reduced. Without crossover acting to restore the variance to its natural value, the population evolves very rapidly to a highly converged distribution and samples a much smaller area of the problem space.

### 6.2 A closed form solution

We can attempt an analytical solution by making the assumption that selection is strong enough for the population to approach the minimum and make the approximation that

$$\mathrm{erf}\,(x) \approx x \qquad \text{where } x = \frac{M_l - K_1}{\sqrt{K_2}}.$$

The response of the first cumulant to selection simplifies significantly in this case and is no longer dependent on the variance of the distribution. We can then simply solve for the equilibrium value of $K_1$ to give

$$K_1^* \approx \frac{(\mathrm{MAX} - 1)\,\Gamma M_l}{\sqrt{\pi}\,(1 - \Gamma) + (\mathrm{MAX} - 1)\,\Gamma}. \tag{42}$$

The lack of dependence of the equilibrium mean on the equilibrium variance seen in Figures 5 and 6 is explained. The equilibrium mean will be close to $M_l$. For small mutation rates, the distance from the minimum is independent of string length $L$ and is of the same order as the number of bits per string which are expected to mutate at each generation,

$$M_l - K_1^* \approx \frac{2\gamma\sqrt{\pi}}{\mathrm{MAX} - 1} M_l. \tag{43}$$

The equilibrium correlation is determined by the balance of mutation and selection and the equilibrium mean. We can simply solve for the equilibrium value of $C$

$$C^* = \frac{x(1-r)}{1-xr}, \tag{44}$$

where

$$x = \frac{\Gamma^2\left(K_1^{*2} - L^2\right)}{\Gamma^2 K_1^{*2} - L^2} \qquad \text{and} \qquad r = \frac{P - \langle n^2 \rangle}{P-1}. \tag{45}$$

The equilibrium correlation is thus given by

$$q^* \approx C^* + \left(1 - C^*\right)\frac{K_1^{*2}}{L^2}. \tag{46}$$

To find the equilibrium variance, we sacrifice some accuracy in order to derive a simple result. We assume that the distribution is close enough to the local minimum to make the approximations,

$$e^{-x^2} \approx 1 \quad \text{and} \quad \mathrm{erf}^2(x) \approx 0 \qquad \text{where } x = \frac{M_l - K_1}{\sqrt{K_2}}.$$

Thus the equilibrium variance is given by

$$K_2^* \approx \frac{L\left(1-\Gamma^2\right) + L\left(1-q^*\right)}{2 - \Gamma^2 r\left[1 - \frac{2(\mathrm{MAX}-1)}{\pi}\right]}. \tag{47}$$

For the case when no crossover is applied we may go straight to the result below using the same approximation

$$K_2^* \approx \frac{L\left(1-\Gamma^2\right)}{1 - \Gamma^2 r\left[1 - \frac{2(\mathrm{MAX}-1)}{\pi}\right]}. \tag{48}$$

For all but the largest mutation rates, $(1-\Gamma^2)$ is small compared to the correlation term $(1-q^*)$. Thus the correlation is significant in producing the larger equilibrium variance of the GA with crossover. For all but the smallest population sizes, $C^*$ is small and the correlation is thus defined by the equilibrium mean and thus the position of the minimum, $M_l$.

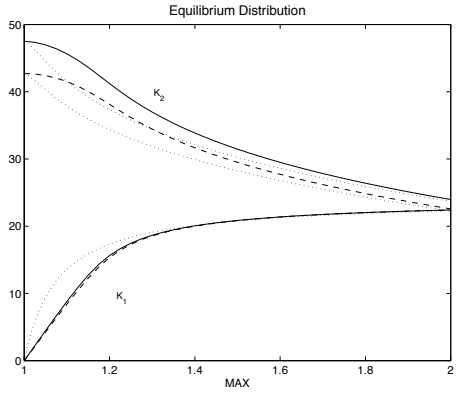


**Figure 7**. Comparison of numerical solution and closed form approximations for the equilibrium distribution of a GA with crossover. Curves are roulette wheel selection (dashed line), SUS (solid line), and closed form approximations (dotted line). GA parameters are as before.

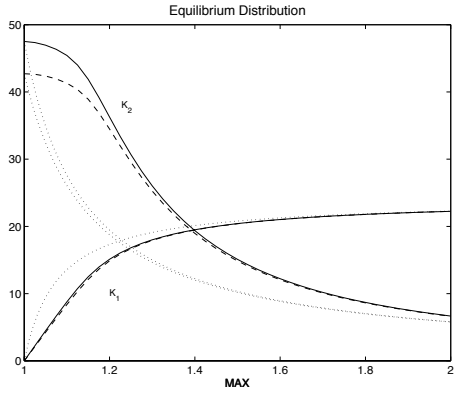


**Figure 8**. Comparison of numerical solution and closed form approximations for the equilibrium distribution of a GA without crossover. Curves are roulette wheel selection (dashed line), SUS (solid line), and closed form approximations (dotted line). GA parameters are as before.

Figures 7 and 8 show the closed form results plotted against the previously shown numerical solutions for the GA with and without crossover.

As expected we see reasonable agreement when selection is strong and the distribution is near the minimum. The accuracy is better for the case with crossover as the larger distribution variance improves the accuracy of the approximations made. The closed form results provide a clear insight into the factors determining the final shape of the distribution.

In section 7 we see that the equilibrium variance is of direct importance to how long it takes to move from the local to the global minimum in the Basin with a Barrier problem.

## 7. Solving the Basin with a Barrier problem

From the analysis of the dynamics of a GA on a concave fitness potential, we can calculate the first passage time of the Basin with a Barrier problem presented earlier. We compare the GA to a stochastic hill climber on a range of problem sizes with $L = 8$, 16, 32, and 48.

### 7.1 Modeling the genetic algorithm

Having solved the dynamics of the GA on the local minimum, we can calculate the time for one population member to exceed the barrier, that is, have magnetization greater than $M_b$, and fall into the global minimum. At each stage in its evolution the ensemble population will be described by the mean and variance of the magnetization. The extremes of the distribution are not well defined but we can naively consider it to be a gaussian and thus the percentage of the ensemble distribution less than the barrier at generation $i$ is,

$$p_i = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{M_b - K_1}{\sqrt{2K_2}}\right)\right]. \tag{49}$$

The probability that at least one member of the population of $P$ jumps over the barrier at each generation is given by $1 - p_i^P$. Thus the expected number of function evaluations for one population member to exceed the barrier is given by

$$n = P \sum_{i=0}^{\infty} \left[ i\left(1 - p_i^P\right) \prod_{j=0}^{i-1} p_j^P \right]. \tag{50}$$

Figure 9 shows the results of the model against simulation results on a logarithmic scale for a typical problem with $M_l = L/2$ and $M_b = 7L/8$. The positions of the barrier and local minimum were chosen to allow easy scaling over many string lengths. The GA is using SUS and crossover. Simulation results are averaged over 1000 runs.

This simple analysis gives reasonable predictions of the first passage time. As the problem size increases, the ensemble distribution lies further from the barrier and the first passage time becomes more dependent on the ill-defined extremes of the distribution. This results in errors when considering a simple gaussian distribution.

At the larger problem sizes, we see an optimum selection strength. This is easy to understand in terms of the end-point distribution plotted in Figure 5. With selection too strong, the population almost reaches

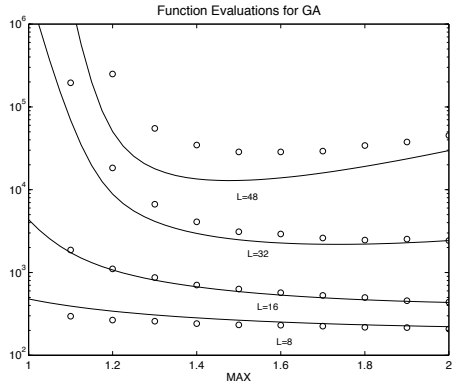


**Figure 9**. Comparison of theoretical and simulation first passage times for a GA with SUS to solve the Basin with a Barrier problem. Simulations are averaged over 1000 runs and the error bars are approximately the size of the symbols. Problem sizes are $L = 8$, 16, 32, and 48.

the center of the local minimum but has a small variance. At the other extreme, weak selection results in a large final variance but the population does not move far from the maximum entropy state. Between these two is an optimum where the extreme of the population distribution has a maximum magnetization.

When the time to solve the problem becomes large, the initial dynamics of the GA become less important and the mean and variance of the equilibrium distribution are most significant. At this limit we can use a simple expansion of the error function, $\mathrm{erf}\,(x) \approx 1 - \mathrm{e}^{-x^2}/x\sqrt{\pi}$, to give an analytical expression for the fraction of the ensemble distribution less than the barrier,

$$p \approx 1 - \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}} \qquad \text{where } x = \frac{M_b - K_1^*}{\sqrt{K_2^*}}. \tag{51}$$

The probability of finding one member in the global minimum is again $1 - p^P$ and the expected time in terms of function evaluations is given by $P/\big(1 - Pp\big)$. Using the previous result gives

$$n \approx x\sqrt{2\pi}\mathrm{e}^{x^2/2}. \tag{52}$$

The most significant term here is the exponential. As shown earlier, the equilibrium mean magnetization $K_1^*$ is constant over the cases considered: roulette wheel selection, SUS, and the GA without mutation. The time to solve the problem is thus strongly dependent on the equilibrium variance

$$n \propto \mathrm{e}^{(M_b - K_1^*)^2/2K_2^*}. \tag{53}$$

For the case considered here, this results in a factor of approximately two in the performance through using SUS as opposed to roulette wheel selection. However the very small equilibrium variance which results when running the GA without crossover has a large influence on the performance. In the case when $L = 48$, the GA without crossover is predicted to perform many orders of magnitude worse. This is verified by experiment in that even after $10^8$ function evaluations, the largest practical limit, the GA does not solve the problem.

### 7.2 Modeling a single stochastic hill climber

Modeling the dynamics of a single stochastic hill climber search method is straightforward. We consider that at each time step a new move is generated by allowing each bit to mutate with probability $1/L$, equivalent to the GA mutation rate. Strictly speaking this allows global and not just local moves. However as $L$ increases, the probability of this becomes small. We model a simple GA where steps which increase fitness are always accepted and steps which will decrease fitness are accepted with some probability $p$. As there is only one barrier in this problem, there is no need to anneal this probability as is done in simulated annealing.

The transition times to reach the global minima are calculated directly from the transition matrix describing the probability of changing from one state to another. A randomly assigned starting string is assumed. Figure 10 shows the results for each problem size on a logarithmic scale. For reasonable size problems there is a clear optimum. When $p = 0$ the hill climber performs simple steepest descent and sits at the local minima waiting for the correct mutation to jump straight into the global minimum. When $p = 1$ the hill climber performs a random walk

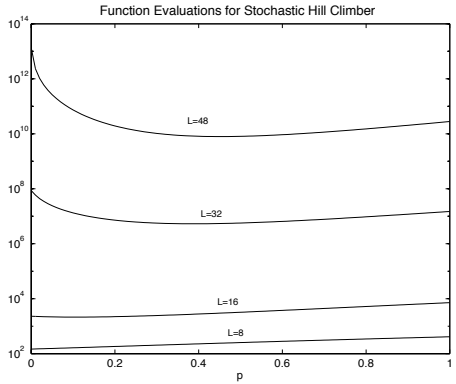


**Figure 10.** Theoretical first passage times for a single stochastic walker to solve the Basin with a Barrier problem with sizes $L = 8$, 16, 32, and 48.

and most time is spent in the maximum entropy area away from both the local and global minima. Between the two an optimum is reached.

As the problem size increases, the time to reach the global minimum increases rapidly. For problem sizes greater than $L = 16$, the single walker is orders of magnitude slower at finding a solution in the global minimum than the time predicted and observed for the GA.

## 8. Discussion

By considering ranking selection we have been able to accurately calculate the effect of selection on any symmetrical concave function of unitation. The inclusion of finite population effects allows a framework by which the effect of roulette wheel selection and SUS may be compared analytically. The comparison confirms the empirically held belief that using SUS rather than roulette wheel selection reduces the effects of genetic drift.

Using and extending results of a previous analysis, we have been able to model the dynamics of the GA when subject to all three operators: selection, mutation, and crossover. A comparison with simulation results shows very good agreement between theory and simulation.

This analysis of the dynamics allows us to look at the time to solve a model of a hard optimization problem, the Basin with a Barrier. We can understand the dynamics in terms of the mean and variance of an ensemble distribution. To find solutions in the global minimum we are looking for population members in the extremes of the distribution. In terms of this search we can understand the effect of crossover and mutation operators.

Mutation will act to increase the variance but at the cost of moving the whole distribution away from the global minimum towards the maximum entropy state.

Crossover acts as a powerful search method in this model. Selection acts on the fitness of the population members to reduce the variance. Crossover acts to increase the variance of the distribution back towards a natural variance defined by the correlation of bits within the population. It does not change the mean and thus searches an area of the problem space without biasing it back towards the maximum entropy state. In the analysis of first passage time it is the significant force in finding solutions in the global minimum.

Whilst there is much analysis which can be performed on this model such as optimum population sizes or parameters, it is interesting to consider the general picture this gives us regarding crossover. The analysis here suggests that, to be effective, crossover should generate new population members with a reasonable degree of fitness correlation with their parents. In this way, selection identifies areas of the problem space which have a good fitness and biases the population to these areas.

Crossover acts on this population, producing a new generation with a larger variance without moving the mean too far from this area. When the fitness of offspring are unrelated to parent fitness, we are effectively doing random search and crossover is unlikely to be the most effective operator in the search.

## Appendix

### A. Analysis of mutation

When the mutation operator is applied, each bit of population member $\mu$ has a small probability of mutation,

$$S_i^\mu \to -S_i^\mu \qquad \text{with probability } \gamma. \tag{A.1}$$

We model the effect of mutation on the ensemble magnetization distribution by first considering the effect on any individual bit. The expected value of any bit after mutation is easily shown to be,

$$\langle S_i^\mu \rangle_m = \Gamma S_i^\mu \qquad \text{where } \Gamma = (1 - 2\gamma), \tag{A.2}$$

and $\langle \ldots \rangle_m$ represents the average over all mutations. Applying this to the expected magnetization of individual $\mu$ after mutation gives,

$$\langle M_\mu \rangle_m = \sum_{i=1}^{L} \langle S_i \rangle_m = \Gamma M_\mu. \tag{A.3}$$

For the second order terms the expression is more complicated as dependent terms must be collected together,

$$\begin{aligned} \langle M_\mu^2 \rangle_m &= \sum_{i \neq j} \langle S_i \rangle_m \langle S_j \rangle_m + \sum_{i=1}^{L} \langle S_i^2 \rangle_m \\ &= \Gamma^2 (M_\mu^2 - L) + L \\ &= \Gamma^2 M_\mu^2 + L(1 - \Gamma^2). \end{aligned} \tag{A.4}$$

Applying these expressions to the definitions of the cumulants and averaging over the ensemble gives,

$$\begin{aligned} \langle K_1 \rangle_m &= \Gamma K_1 \\ \langle K_2 \rangle_m &= \Gamma^2 K_2 + L(1 - \Gamma^2). \end{aligned} \tag{A.5}$$

### B. Analysis of uniform crossover

When we apply uniform crossover [13] the bits of offspring $\mu$ are drawn from each parent $\alpha$ and $\beta$ at random

$$S_i^\mu = \chi_i S_i^\alpha + \left(1 - \chi_i\right) S_i^\beta, \tag{B.1}$$

where

$$\chi_i^\mu = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2. \end{cases} \tag{B.2}$$

We calculate the effect by considering each bit. The expected value of any bit from offspring $\mu$ averaged over all ways of drawing bits from each parent is simply

$$\langle S_i^\mu \rangle_x = \frac{S_i^\alpha}{2} + \frac{S_i^\beta}{2}. \tag{B.3}$$

Clearly then the expected magnetization of offspring $\mu$ produced through crossover is

$$\langle M_\mu \rangle_x = \frac{M_\alpha}{2} + \frac{M_\beta}{2}. \tag{B.4}$$

As $\alpha$ and $\beta$ are drawn independently from the population, we may simply average over the ensemble to give the mean ensemble magnetization after crossover

$$\langle M \rangle_x = \langle M \rangle. \tag{B.5}$$

For second order terms the analysis is slightly more complicated. If we again consider offspring $\mu$ then

$$\begin{aligned} \langle M_\mu^2 \rangle_x &= \sum_{i/=j} \langle S_i \rangle_x \langle S_j \rangle_x + \sum_{i=1}^{L} \langle S_i^2 \rangle_x \\ &= \sum_{i/=j} \langle S_i \rangle_x \langle S_j \rangle_x + L \\ &= \left( \sum_{i=1}^{L} \langle S_i \rangle_x \right)^2 - \sum_{i=1}^{L} \langle S_i \rangle_x^2 + L. \end{aligned} \tag{B.6}$$

Now considering all ways of drawing the parents and the results from equations (B.3) and (B.4) gives

$$\begin{aligned} \langle M_\mu^2 \rangle_x &= \left( \frac{M_\alpha}{2} + \frac{M_\beta}{2} \right)^2 - \sum_{i=1}^{L} \left( \frac{S_i^\alpha}{2} + \frac{S_i^\beta}{2} \right)^2 + L \\ &= \left( \frac{M_\alpha^2}{4} + \frac{M_\beta^2}{4} + \frac{M_\alpha M_\beta}{2} \right) - \sum_{i=1}^{L} \left( \frac{1}{2} + \frac{S_i^\alpha S_i^\beta}{2} \right) + L. \end{aligned} \tag{B.7}$$

Averaging over all ways of drawing $\alpha$ and $\beta$ independently from the population gives the second moment of the ensemble magnetization

$$\langle M^2 \rangle_x = \frac{\langle M^2 \rangle}{2} + \frac{\langle M \rangle^2}{2} + \frac{L}{2}\left(1 - q\right), \tag{B.8}$$

where $q$ is defined as

$$q = \frac{1}{P(P-1)} \sum_{\alpha \neq \beta} \frac{1}{L} \sum_{i=1}^{L} S_i^\alpha S_i^\beta. \tag{B.9}$$

The cumulant terms are simply found from the ensemble moments derived previously

$$\begin{aligned} \langle K_1 \rangle_x &= K_1 \\ \langle K_2 \rangle_x &= \frac{K_2}{2} + \frac{L}{2}\left(1 - q\right). \end{aligned} \tag{B.10}$$

## References

[1] J. Baker, "Adaptive Selection Methods for Genetic Algorithms," in *Proceedings of the First International Conference on Genetic Algorithms*, edited by J. J. Grefenstette (Erlbaum, 1985).

[2] J. Baker, "Reducing Bias and Inefficiency in the Selection Algorithm," in *Proceedings of the Second International Conference on Genetic Algorithms*, edited by J. J. Grefenstette (Erlbaum, 1987).

[3] T. Blickle and L. Theile, "A Mathematical Analysis of Tournament Selection," in *Proceedings of the Sixth International Conference on Genetic Algorithms*, edited by L. Eshelman (Erlbaum, 1995).

[4] Y. Davidor, "Epistasis Variance: A Viewpoint on GA-hardness," in *Foundations of Genetic Algorithms*, edited by G. J. E. Rawlins (Morgan Kaufmann, 1991).

[5] T. Jones and S. Forrest, "Fitness Distance Correlation as a Measure of Problem Difficulty in Genetic Algorithms," in *Proceedings of the Sixth International Conference on Genetic Algorithms*, edited by L. Eshelman (Morgan Kaufmann, 1995).

[6] M. Mitchell, J. Holland, and S. Forrest, "When Will a Genetic Algorithm Outperform Hill Climbing?" in *Advances in Neural Information Processing Systems*, edited by J. Cowan, G. Tesauro, and J. Alspector (Morgan Kauffman, San Francisco, 1994).

[7] A. Nix and M. D. Vose, "Modeling Genetic Algorithms with Markov Chains," *Annals of Mathematics and Artificial Intelligence*, **5** (1991) 79–88.

[8] A. Prügel-Bennett and J. L. Shapiro, "An Analysis of Genetic Algorithms Using Statistical Mechanics," *Physical Review Letters*, **72** (9) (1994) 1305–1309.

[9] A. Prügel-Bennett and J. L. Shapiro, "The Dynamics of a Genetic Algorithm for Simple Random Ising Systems," *Physica D*, **104** (1997) 75–114.

[10] M. Rattray, "The Dynamics of a Genetic Algorithm under Stabilizing Selection," *Complex Systems*, **9** (3) (1995) 213–234.

[11] A. Rogers and A. Prügel-Bennett, "Genetic Drift in Genetic Algorithm Selection Schemes," *IEEE Transactions on Evolutionary Computation*, **3** (4) (1999).

[12] J. L. Shapiro and A. Prügel-Bennett, "Genetic Algorithms Dynamics in Two-Well Potentials with Basins and Barriers," in *Foundations of Genetic Algorithms 4*, edited by R. K. Belew and M. D. Vose (Morgan Kaufmann, San Francisco, 1997).

[13] G. Syswerda, "Uniform Crossover in Genetic Algorithms," in *Proceedings of the Third International Conference on Genetic Algorithms* (Morgan Kaufmann, San Mateo, 1989).

[14] M. D. Vose, "Modelling Simple Genetic Algorithms," in *Foundations of Genetic Algorithms 2*, edited by D. Whitley (Morgan Kaufmann, San Mateo, 1992).