# Bistability of the Needle Function in the Presence of Truncation Selection

Genetic Algorithms Track

**Abstract.** It is possible for a GA to have two stable fixed points on a single-peak fitness landscape. These can correspond to meta-stable finite populations. This phenomenon is called *bistability*, and is only known to happen in the presence of recombination, selection, and mutation. This paper models the bistability phenomenon using an infinite population model of a GA based on gene pool recombination. Fixed points and their stability are explicitly calculated. This is possible since the infinite population model of the gene pool GA is much more tractable than the infinite population model for the standard simple GA. For the needle-in-the-haystack fitness function, the fixed point equations reduce to a single variable polynomial equation, and stability of fixed points can be determined from the derivative of the single variable equation.

## 1 Introduction

Intuitively, one would expect that a GA could have only a single stable fixed point on a single-peak fitness function. However, in this paper we show that a mutation/crossover/selection GA may have two stable fixed points on a single-peak fitness landscape. We call this phenomenon *bistability*.

In practical terms, this means that when the GA is started with a random population, the GA can stagnate without making any progress in the direction of the fitness peak. However, when the GA with the same parameters is started nearer to the fitness peak, the GA will move towards the peak. This behavior is described in more detail in section 5.

In our model, there cannot be bistability without both mutation and recombination. For the more general Vose model, there cannot be bistability without recombination [1] and it is conjectured that there cannot be bistability without mutation. Bistability only occurs for some settings of the parameters where there is the appropriate balance between mutation, recombination, and selection. For our model, the minimum string length for bistability is 4. As the string length increases, the range of mutation rates where bistability occurs increases rapidly.

Bistability is not an isolated phenomenon. While we use gene pool recombination and a needle-in-the-haystack fitness for our model, neither are necessary for bistability. We show that bistability occurs with uniform and one-point crossover, and Wright, Rowe, Poli and Stephens [WRPS03] show that bistability occurs with with fitness functions that show a more gradual increase in fitness as one gets close to the fitness peak This paper extends on [WRPS03] where bistability was analyzed in the case of proportional selection. This paper treats the case of truncation selection.

Bistability was first discovered and investigated by [BBN96] in the context of viral quasispecies and the AIDS virus. Other papers on bistability include [OH97]

[1] this follows from the Peron-Froebenius theorem

and [WRN02]. The last citation has a more complete review of the literature in this area.

Vose [Vos99] and others have developed an elegant dynamical system model for GAs. This model is commonly referred to as the "infinite population model" of a GA. Vose has proved elegant theorems relating fixed points to finite population GA behavior when the population is large. However, except in special cases, only numerical methods can be used to actually find the fixed points. Due to the large dimensionality of the space, numerical methods can be used only for relatively small string lengths. Further, it is difficult to use the applicable numerical methods to achieve understanding of the processes involved.

The model used in this paper makes a "linkage equilibrium" assumption that makes the infinite population model more tractable. Linkage equilibrium has been widely used as a simplifying assumption in approximate models of GAs. These include [SI99] and [PBR01].

We show that for the needle-in-the-haystack fitness functions, fixed points can be found by solving a single variable polynomial equation, and the stability of these fixed points can be determined from this single variable equation. Thus, we can rigorously determine the number of fixed points and their stability, something that has not been possible for the Vose model when there is nontrivial selection, mutation, and recombination. We can determine and plot the ranges of parameters where bistability occurs. The price that we pay for the linkage equilibrium assumption is that we no longer have an exact infinite population model of a standard GA that uses two-parent crossover. Instead, we have an exact model of a GA that uses a special form of recombination called "gene pool recombination". This GA will be described in more detail in section 2

Gene pool recombination is an alternative recombination method for GAs. An individual created by gene pool recombination is chosen from a probability distribution determined directly from the whole population rather than from two parents. Geiringer's theorem [Gei44] shows that gene pool recombination in the infinite population model can be viewed as the limit of repeated applications of two-parent recombination (without selection). For the infinite population model, gene pool recombination takes the population to linkage equilibrium in one step.

Gene pool recombination was proposed in [Sys93]. [MM01], [MM00] have investigated the UMDA algorithm which is a GA that uses gene pool recombination and some kind of selection (proportional, tournament, truncation, or Boltzmann) and no mutation. They have experimentally verified that UMDA can successfully optimize a wide variety of fitness functions but is misled by deceptive problems. Further, gene pool recombination is used in population-based incremental learning (PBIL) [BC95], a machine learning technique.

Our empirical results in section 5 and the empirical results given in [WRPS03] suggest that the gene pool GA model is a reasonable approximation to the two-parent recombination GA for the classes of fitness functions investigated in this paper. Thus, the gene pool GA model can be viewed as an approximate model of the two-parent recombination GA.

## 2 The Gene Pool GA

In this section we give a more precise description of the gene pool GA. Our model is an exact model of this GA in the limit as the population size goes to infinity.

We assume a binary string representation. The string length is $\ell$.

Gene pool recombination uses a sampling process to go from one population (the current population) to another (the new population). Each individual of the new population is created independently of the others, and in fact, each bit of each individual is created independently of the other bits.

The first step in the sampling process is to calculate the relative frequency of a 0 bit at each locus (string position) in the current population. For a given locus, this frequency is just the relative frequency of an order-1 schema whose only defined position is a 0 at this locus. For an individual of the new population, the probability of a 0 at this locus is the same as the relative frequency of a 0 at the same locus of the current population.

Note that expected result of gene pool recombination depends only on the relative frequencies of the order-1 schemata. Thus, the infinite population model can be defined in terms of these frequencies.

Note that since bits of individuals in the new population are chosen independently, there is no expected correlation between the bits at different loci. This is exactly the definition of linkage equilibrium, so in the infinite population limit, gene pool recombination produces a linkage equilibrium population.

Then the steps of the gene pool GA used in this paper are as follows:

1. Choose a random population.
2. Apply gene pool recombination.
3. Apply truncation selection.
4. Apply mutation.
5. Return to step 2 if termination criteria is not met.

## 3 The Infinite Population Model

Our infinite population model is represented in the Walsh basis. (See the appendix for the definition of the Walsh transform.) Since we only need to represent information equivalent to the order-1 schema frequencies, this is not as complicated as it seems. The model uses the order-1 Walsh coefficients, and these coefficients are expressed simply in terms of the order-1 schema frequencies. This is explained below.

Let $\Omega$ be the search space of all length-$\ell$ binary strings. If $j \in \Omega$ is a binary string, $\#j$ represents the number of ones in $j$. Let $\mathcal{L}$ denote the set of strings $j$ with $\#j = 1$.

A population is represented by a vector indexed over $\Omega$. Thus, if $x$ is a population, $x_j$ is the relative frequency of string $j \in \Omega$.

For $k \in \mathcal{L}$, let $x_0^{(k)}$ and $x_1^{(k)}$ denote the schema frequencies of the order-1 schemata whose only defined locus is the locus of the 1 bit in $k$. (The string $k$

can be thought of as a mask. The locus corresponding to $k$ is the locus masked by $k$.) Thus, if $k = 00010$, then $x_0^{(k)}$ is the frequency of the schema $***0*$ and $x_1^{(k)}$ is the frequency of the schema $***1*$.

Let $\widehat{x}_k$ denote the $k$th coefficient of the Walsh transform of $x$. It can be shown [WRPS03] that for any population vector $x$ and any $k \in \mathcal{L}$,

$$\widehat{x}_k = 2x_0^{(k)} - 1 \quad \text{and} \quad \widehat{x}_k = 1 - 2x_1^{(k)} \ . \tag{1}$$

Note that $x_0^{(k)} + x_1^{(k)} = 1$ by definition. For this purposes of this paper, these formulas can be taken as the definition of the order-1 Walsh coefficient $\widehat{x}_k$.

The value of $\widehat{x}_k$ ranges from $-1$ when the frequency of a 0 in the locus masked by $k$ is 0 to $+1$ when the frequency of a 0 in this locus is 1.

Now we can look at modelling the steps of the algorithm given in section 2. There are two important observations to make.

First, the result of the gene pool recombination step depends only on the frequencies of the order-1 schemata and hence of the order-1 Walsh coefficients. Thus, our model only needs to keep track of the order-1 Walsh coefficients.

Second, the expected frequencies of the order-1 schemata do not change in the gene pool recombination step, and thus the infinite population model of gene pool recombination is trivial: the order-1 Walsh coefficients remain unchanged.

So it remains to model mutation and selection.

The Walsh basis formula for mutation is very simple. If the mutation rate is $\mu$, the effect of mutation in the infinite population model is to multiply $\widehat{x}_k$ by $1 - 2\mu$ [WRPS03]. One can see that the effect of mutation is to move the order-1 schema frequencies towards $1/2$. In fact, if the mutation rate is $1/2$, then mutation makes $\widehat{x}_k$ to be zero, which corresponds to schema frequencies of $1/2$.

### 3.1 The needle-in-the-haystack fitness and truncation selection

The needle-in-the-haystack (NEEDLE) fitness function assigns a fitness of 1 to all strings except the all-zeros string. The all-zeros string has a fitness greater than 1. (The exact value is unimportant since we are using truncation selection.)

In truncation selection, a fraction $T$ of the population is kept and the remainder is discarded. For the NEEDLE fitness function, we only need to determine how the frequency of the all-zeros string is increased by selection. In our infinite population model of truncation selection, we assume that the frequency of all other strings is decreased by the same multiplicative factor.

Thus, let $x$ denote the population before selection, and let $x_0$ denote the frequency of the all-zeros string in this population. Let $y$ denote the population after selection. Then it is not hard to see that $y_0 = \min\left(1, \frac{x_0}{T}\right)$ where $T$ is the truncation fraction. For any other string $j \neq 0$, $y_j = x_j \left(\frac{1-y_0}{1-x_0}\right)$.

Recall that $x_1^{(k)}$ is the frequency of the order-1 schema whose value is 1 in the locus masked by $k$. Since the all-zeros string is not a member of this schema, the effect of selection on the frequency of this schema is the same as the effect on any nonzero string. In other words, $y_1^{(k)} = x_1^{(k)} \left(\frac{1-y_0}{1-x_0}\right)$.

This formula can be transformed into a formula on the Walsh coefficients using the second equation of (1).

$$\widehat{y}_k = 1 - (1 - \widehat{x}_k)\left(\frac{1 - y_0}{1 - x_0}\right)$$

Let $G$ denote the mapping the represents the complete model. Thus, if $x$ is the population at the beginning of step 2 of the algorithm, then $G(x)$ is the expected next generation population at the same point in the algorithm. We can get a formula for $G$ by multiplying by $1 - 2\mu$ to include the effect of mutation.

$$\widehat{G(x)}_k = (1 - 2\mu)\left(1 - (1 - \widehat{x}_k)\left(\frac{1 - y_0}{1 - x_0}\right)\right)$$

We are not done since this formula still includes the standard basis quantities $x_0$ and $y_0$, and we want a formula in terms of the order-1 Walsh coefficients. The key to eliminating these quantities is to note that selection occurs right after gene pool recombination in the algorithm, and gene pool recombination takes a population to linkage equilibrium. Thus we can assume that the population $x$ (that selection is applied to) is in linkage equilibrium.

In [WRPS03], it is shown that

$$x_0 = 2^{-\ell} \prod_{j \in \mathcal{L}} (1 + \widehat{x}_j) \ . \tag{2}$$

There are two cases. The first case is when $x_0 \leq T$. In this case, $y_0 = \frac{x_0}{T}$, and some algebraic manipulation shows the first case of the formula below. The second case is when $x_0 > T$, and then $y_0 = 1$. This implies the second case of the formula below.

$$\widehat{G(x)}_k = \begin{cases} (1 - 2\mu)\left(1 - (1 - \widehat{x}_k)\left(\frac{T - \prod_{j \in \mathcal{L}}(1 + \widehat{x}_j)/2}{T(1 - \prod_{j \in \mathcal{L}}(1 + \widehat{x}_j)/2)}\right)\right) & \text{if } x_0 \leq T \\ (1 - 2\mu) & \text{if } x_0 > T \end{cases} \tag{3}$$

We can show that if $x_0 \geq T$, then $\widehat{x}_k = 1 - 2\mu$ is a fixed point since

$$\begin{aligned} x_0 &= 2^{-\ell} \prod_{j \in \mathcal{L}} (1 + \widehat{x}_j) \\ &= 2^{-\ell} \prod_{j \in \mathcal{L}} (1 + 1 - 2\mu) \\ &= (1 - \mu)^{\ell} \\ &> T \ . \end{aligned}$$

In the case where $x_0 \leq T$, formula (3) leads to the fixed point equations:

$$\widehat{x}_k = \frac{(1 - 2\mu)\left(1 - \left(\frac{T - \prod_{j \in \mathcal{L}}(1 + \widehat{x}_j)/2}{T(1 - \prod_{j \in \mathcal{L}}(1 + \widehat{x}_j)/2)}\right)\right)}{1 - (1 - 2\mu)\left(\frac{T - \prod_{j \in \mathcal{L}}(1 + \widehat{x}_j)/2}{T(1 - \prod_{j \in \mathcal{L}}(1 + \widehat{x}_j)/2)}\right)}$$

The right hand side of this equation is the same for all $k$. Therefore, if $\widehat{x}$ is a fixed point of $G$, then $\widehat{x}$ is symmetric in the sense that all $\widehat{x}_k$ for $k \in \mathcal{L}$ are equal.

If the GA model is started with a symmetric population, symmetry will be maintained. In this case, the recurrence equation can be written in terms of the variable $w$,

$$\widehat{G(w)} = \begin{cases} (1-2\mu) & \text{if } T < 2^{-\ell}(1+w)^{\ell} \ ; \\ (1-2\mu)\left(1-(1-w)\frac{T-2^{-\ell}(1+w)^{\ell}}{T(1-2^{-l}(1+w)^{\ell})}\right) & \text{otherwise.} \end{cases}$$

The fixed points occur when $\widehat{G(w)} = w$, or equivalently when equation ((4)) below holds.

**Theorem 1.** *The fixed points of the infinite population model of the gene pool GA algorithm are the solutions to the variable polynomial equation:*

$$w = \begin{cases} (1-2\mu) & \textit{if } T < 2^{-\ell}(1+w)^{\ell} \ ; \\ (1-2\mu)\left(1-(1-w)\frac{T-2^{-\ell}(1+w)^{\ell}}{T(1-2^{-l}(1+w)^{\ell})}\right) & \textit{otherwise.} \end{cases} \tag{4}$$

As we will see, understanding the solutions to this equation is far easier than understanding the solutions to the system of polynomial equations that come from a more general model.

As an example we solve (4) numerically using $\ell$=8, $\mu$=.1, and $T$=.4. We find $w = .02869$ and $w = .7222$. A third equilibrium occurs at $w = 1 - 2\mu = .8$. A graph of $G(w) = w$ is shown in Fig. 1. We see three fixed points; the first and third fixed points are stable and the middle fixed point is unstable relative to symmetric populations. However, stability in the figure does not necessarily imply stability in the space of all (possibly non-symmetric) populations.

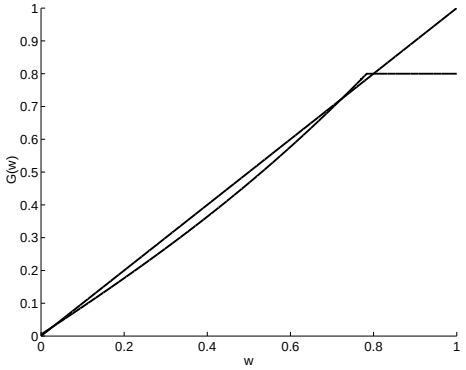


**Fig. 1.** Graph of $\widehat{G(w)}$ and 45° line for $\mu = .1$, $\ell = 8$ and $T = .4$.

We are interested in the stability of the fixed points in the space of all populations. It is well known for a discrete model that if all of the eigenvalues of

$dg_x$ have modulus less than 1, then $x$ is an asymptotically stable fixed point of $g$. The following theorem found in Cripe [Cri03] shows the eigenvalue of $d\widehat{G}$ is equal to the derivative of the single-variable function $\widehat{G}$ defined in (4). Therefore the fixed points can not only be found by solving a single-variable polynomial but additionally their stability can be determined by taking a single variable derivative.

**Theorem 2.** *At a symmetric point $\widehat{x}$ where $\widehat{x}_k = w$ for all $k \in \mathcal{L}$, The largest modulus eigenvalue of $G_{\widehat{x}}$ is equal to $\frac{dG}{dw}$.*

Since the largest modulus eigenvalue of $\widehat{dG(x)}$ is equal to the derivative of the single variable function $G$, the stability of the fixed points in the cube $[-1, 1]^\ell$ is the same as the stability of the fixed points in the one variable space of symmetric populations.

We have shown that we can find the fixed points of the model by solving a single variable polynomial of degree $\ell + 1$ and furthermore, the stability of the fixed points can be determined from this equation.

## 4 Explorations of Parameter Space

In this section we analyze the fixed points for our model to determine the parameter settings where bistability occurs. Due to space constraints the proofs of lemmas are not given when they are straightforward. Proofs are in [Cri03].

We begin to explore the parameter space by finding a relationship between $T$ and $w$ when $w$ is a fixed point. Solving (4) for $T$, in the case $T > 2^{-\ell}(1+w)^\ell$,we find

$$T(w) = \frac{-(1-2\mu)(1-w)}{(w-1+2\mu-\frac{2\mu w}{x_0})} \quad . \tag{5}$$

That is, we define $T(w)$ to be the value of $T$ for which $w$ is a fixed point. Lemmas 1 and 2 show that the fixed points can occur in the region $0 < w < 1 - 2\mu$.

**Lemma 1.** *$x_0 = 2^{-\ell}(1+w)^\ell < T(w)$ for $0 < w < 1 - 2\mu$ and $x_0 > T(w)$ for $1 > w > 1 - 2\mu$.*

*Proof.*

$$T(w) - x_0 = \frac{x_0(1-x_0)(1-2\mu-w)}{-wx_0 + x_0 - 2\mu x_0 + 2\mu w}.$$

If $0 < w < 1 - 2\mu$, the denominator,

$$x_0(1-w) - 2\mu(x_0 - w) > x_0 2\mu - 2\mu x_0 + 2\mu w = 2\mu w > 0.$$

$(1 - 2\mu) > w$ implies $(1 - w - 2\mu) > 0$. The numerator is nonnegative. For $w > (1 - 2\mu)$, the numerator is negative and the denominator is positive. □

**Lemma 2.** *$T(w) > 0$ for $0 < w < 1 - 2\mu$.*

Recall that if $T < (1-\mu)^\ell$, then one fixed point occurs at $w = 1-2\mu$. In order for bistability to exist, there must be two additional fixed points, both less than $1-2\mu$. These are solutions to equation 5. The left hand drawing in Fig. 2 shows the plots of $T(w)$ for various values of $\mu$. Bistability occurs for a fixed value of $T$ if a horizontal line drawn at height $T$ intersects the curve three times. The plot of $T(w)$ ends with a vertical line at $w = 1-2\mu$.

The right hand drawing in Fig. 2 shows the progression from three fixed points to one fixed point when $T$ is increased. A bifurcation occurs when $T$ is approximately .43. When $T = .4$ there are three fixed points, one near zero, one at approximately $w = .5$, and one at the critical value of $w = 1-2\mu = .8$. When $T = .43$ the middle fixed point merges with the fixed point at $w = .8$. When $T$ is further increased to .45, the last fixed point has disappeared completely, leaving only the fixed point near $w = 0$.

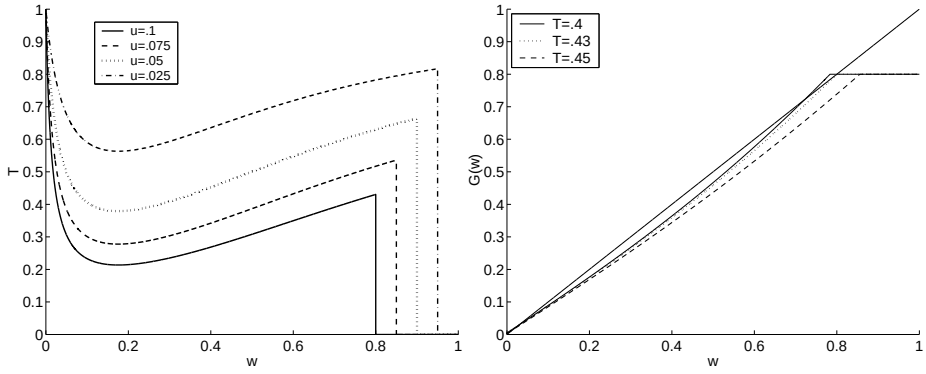


**Fig. 2.**

The $(w, T)$ points where the curve has a zero slope can be critical values. As $T$ increases through such a value, a pair of fixed points can appear or disappear. Therefore it is useful to differentiate (5) and set the result equal to zero. This gives

$$0 = -\ell w^2 + (\ell + x_0 - 1)w + (x_0 - 1) \ . \tag{6}$$

The reader should note that since 6 is independent of $\mu$, the minimums in the left hand side of Fig. 2 all occur at the same value of $w$.

Lemma 3 shows the conditions under which $T$ has a local minimum in the region $0 < w < 1-2\mu$.

**Lemma 3.** $\frac{dT}{dw} = 0$ *has exactly one solution between* 0 *and* $1-2\mu$ *when* $1 - \ell\mu(1-2\mu) - \mu - (1-\mu)^{\ell+1} < 0$.

Let $w_c$ be the critical point of $T$ between 0 and $1-2\mu$. Then $T(w_c)$ is a minimum since the first derivative of $T$ passes from negative to positive. Since $w_c$ is the only critical point, it must be a global minimum in the interval $0 < w < 1-2\mu$.

For the parameter values in the left hand side of Fig. 2, we calculate that $w_c = .1758$. If $.1758 < 1 - 2\mu$, or $\mu > .4121$ then $T$ will not have an interior local minimum. We check that the hypothesis in Lemma 3 is not satisfied. $1 - \ell(1-2\mu) - \mu - (1-\mu)^{\ell+1} > 0$ when $\mu > .4121$. By numerically analyzing the inequality, we find that the minimum string length that satisfies the hypothesis is $\ell = 4$.

**Lemma 4.** *In the case $T < (1-\mu)^\ell$, if $w_c < 1 - 2\mu$, equivalently if*

$$1 - \ell\mu(1-2\mu) - \mu - (1-\mu)^{\ell+1} < 0,$$

*then there exists a value of $T$ that gives bistability.*

*Proof.* By Lemma 3 there exists horizontal lines that will cross the graph of $T(w)$ more than once. Each place of intersection represents a fixed point. Another fixed point of $G$ exists at $w = 1 - 2\mu$. □

We also note that bistability exists if $T(w_c) < T < (1-\mu)^\ell$. It remains to determine the stability of the fixed points.

**Theorem 3.** *If $0 < \mu < 1/2$, then there can be at most three fixed points for $\widehat{G}$. When there are three fixed points, they are stable, unstable and stable when ordered by $w$ values. If there are two fixed points, then a small perturbation of either $T$ or $\mu$ can give one fixed point.*

*Proof.* Since $\frac{dG}{dw} > 0$ then at a fixed point where the graph crosses from above to below the slope must be less than one and therefore must be stable. A fixed point where the graph crosses from below to above must have slope greater than one and is unstable.

Since $\widehat{G}(0) > 0$, when there are three fixed points, the graph must cross the diagonal from above to below, then below to above. □

We have exactly characterized the fixed points for the gene pool model on the NEEDLE fitness function. For fixed values of $T$ and $\mu$ the location of the fixed points can be found using (4).

For a fixed value of $\mu$ the range of values of $T$ which give bistability can be found. For example see the left drawing in Fig. 3. The area between the two curves is the region in $(\mu, T)$ space where bistability occurs. The top curve is $T = (1-\mu)^\ell$ and the lower curve is found by solving $\frac{dT}{dw} = 0$ to find $w_c$ and then taking $T(w_c)$.

For a fixed value of $T$ and $\ell$ the range of $\mu$ which give bistability can be found. This can be seen in the right hand drawing in Fig 3. This figure shows the region in $(\ell, \mu)$ space for which bistability occurs. As noted before, a string length of 4 or more is needed for bistability. The top curve of each pair in this figure is $\mu = 1 - T^{\frac{1}{\ell}}$ The bottom curve of each pair was found for each $\ell$ by solving $T(w_c) = T$ for $\mu$ for $T = .25, .3, .5, .7$.

These results are surprising. Truncation selection with a small value of $T$ is thought of as a very strong selection method. One might think that sufficiently

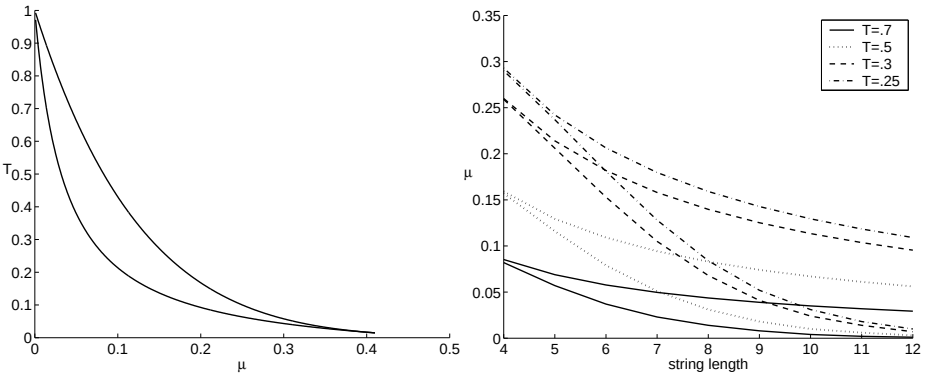


**Fig. 3.**

strong selection would eliminate the possibility of bistability, but these results show bistability even with $T = 0.1$ and a string length of 4.

In [WRPS03] it was shown that the minimum string length for bistability with proportional selection is 6, and this occurs with a very small mutation rate and weak selection. (The extra increment in fitness of the needle string is a strength-of-selection parameter for proportional selection.) Here, the minimum string length for bistability is 4, and it can occur over a wide range of mutation rates for different values of $T$. This includes strong selection and high mutation.

## 5 Empirical Results

In this section we show that bistability can be observed in finite populations GAs with one point crossover and realistic settings for the mutation rate, the crossover rate, and the truncation fraction $T$.

The following procedure can be used to empirically test if a finite population GA with certain parameter settings has bistability for the NEEDLE fitness function. We run the GA with two different initial populations. First, we use an initial population containing only a few (but not zero) copies of the optimal needle string. Second, we use an initial population consisting mostly or all of copies of the needle string. If GA with the first initial population does not move to a population with many copies of the optimal string, and with the second initial population maintains many copies of the optimal string, this indicates that the GA has bistability. (This experiment should be repeated several or many times since anything can happen on a single run of a GA.)

Simulations were performed using a Java program written by the first author and his students. Simulations were performed with truncation selection with $T = 0.5$, string length 8, mutation rate 0.06. (These values are in the middle of the bistability region in the right hand side of Fig. 3.) The initial population was either generated randomly or by choosing strings with a probability of 0.85 of a zero for each bit (biased initialization).

For these parameter settings, the model predicts fixed points at $w = 0.00355$ and $w = 0.8268$. Formula (2) gives the corresponding $x_0$ values of 0.004 and 0.6096. First, we verified the model by doing 100 runs with uniform crossover with crossover rate 1 and a population size $10,000$. The average $x_0$ after 50 generations with random initialization was 0.00661 and with biased initialization was 0.6096. (Additional experiments for other parameters reported in [Cri03] also show very close agreement with the model for uniform crossover and genepool recombination.)

Next we did experiments with weaker crossover and smaller population sizes. For these experiments we used one point crossover with crossover rate 0.8. For population size 500, the average $x_0$ after 50 generations with random initialization was 0.0146 and with biased initialization was 0.6093. Over the 100 runs, there was a minimum of 273 copies of the needle string with biased initialization and a maximum of 32 copies of the needle string with random initialization. In other words, all populations with random initialization were close to the random fixed point, and all populations with biased initialization were close to the needle fixed point.

For population size 100, the average $x_0$ after 50 generations with random initialization was 0.0512 and with biased initialization was 0.613. Over the 100 runs, there was a minimum of 50 copies of the needle string with biased initialization. With random initialization, there were 9 runs with more than 9 copies of the needle string. For these runs, one could say that the GA had escaped the random fixed point and moved toward the needle fixed point. This is a case when a small population size is better.

## 6 Conclusion

In this paper we have shown that an infinite population GPR model closely approximates the finite population two-parent uniform crossover GA when the fitness function exhibits a single peak. Under the gene pool recombination model, the complicated dynamical system of the finite population GA becomes tractable. In particular, we can explicitly calculate the fixed points and determine their stability by examining a single variable polynomial function. Finite population simulations suggest that TPR and GPR produce bistability. The fixed points produced in the simulations closely match those predicted by the model.

Furthermore, we have demonstrated that the infinite population GPR model correctly predicts the presence of bistability in the finite population GA. We have derived explicit formulas that relate the parameter values under which the bistability phenomena occurs.

There is a lesson for the practitioner. When the GA is initialized with a random population, bistability is a phenomenon that should be avoided since it may prevent the GA from finding peaks in the fitness. One way to do this is to keep the strength of recombination down. This can be done by reducing the crossover rate, or by choosing a "weaker" crossover (such as one-point or two-point instead of uniform).

## References


[BBN96] M. C. Boerlijst, S. Bonhoeffer, and M. A. Nowak. Viral quasi-species and recombination. *Proc. Royal Society London B*, 263:1577–1584, 1996.

[BC95] Shumeet Baluja and Rich Caruana. Removing the genetics from the standard genetic algorithm. In A. Prieditis and S. Russel, editors, *The Int. Conf. on Machine Learning 1995*, pages 38–46, San Mateo, CA, 1995. Morgan Kaufmann Publishers.

[Cri03] Greg Cripe. Bistability of the needle function in the presence of truncation selection. Master's thesis, University of Montana, Missoula, MT 59812 USA, 2003.

[Gei44] H. Geiringer. On the probability of linkage in Mendelian heredity. *Annals of Mathematical Statistics*, 15:25–57, 1944.

[MM00] Heinz Mühlenbein and Thilo Mahnig. Evolutionary algorithms: From recombination to search distributions. In L. Kallel, B. Naudts, and A. Rogers, editors, *Theoretical Aspects of Evolutionary Computation*, pages 137–176. Springer Verlag, 2000.

[MM01] Heinz Mühlenbein and Thilo Mahnig. Evolutionary computation and beyond. In Y. Uesaka, P. Kanerva, and H. Asoh, editors, *Foundations of Real-World Intelligence*, pages 123–188. CSLI Publications, Stanford, California, 2001.

[OH97] G. Ochoa and I. Harvey. Recombination and error thresholds in finite populations. In *Foundations of Genetic Algorithms 5*, pages 245–264, San Mateo, 1997. Morgan Kaufmann.

[PBR01] Adam Prügel-Bennett and Alex Rogers. Modelling genetic algorithm dynamics. In *Theoretical Aspects of Evolutionary Computing*. Springer Verlag, 2001.

[SI99] Hideaki Suzuki and Yoh Iwasa. Crossover accelerates evolution in gas with a babel-like fitness landscape: Mathematical analyses. *Evolutionary Computation*, 7(3):275–310, 1999.

[Sys93] G. Syswerda. Simulated crossover in genetic algorithms. In L. Darrel Whitley, editor, *Foundations of Genetic Algorithms 2*, San Mateo, 1993. Morgan Kaufmann.

[Vos99] M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press, Cambridge, MA, 1999.

[WRN02] A. H. Wright, J. E. Rowe, and J. R. Neil. Analysis of the simple genetic algorithm on the single-peak and double-peak landscapes. In *Proceedings of the Congress on Evolutionary Computation (CEC) 2002*, pages 214–219. IEEE Press, 2002.

[WRPS03] A. H. Wright, J. E. Rowe, R. Poli, and C. R. Stephens. Bistability in a gene pool GA with mutation. In *Foundations of genetic algorithms (FOGA-7)*, San Mateo, 2003. Morgan Kaufmann. http://www.cs.umt.edu/u/wright/pubs.htm.


### Appendix

The Walsh matrix $W$ is a $2^\ell$ by $2^\ell$ matrix defined by $W_{i,j} = (-1)^{\#(i \otimes j)}$. The Walsh matrix is symmetric and $W^{-1} = 2^{-\ell} W$. If $x$ is a population vector, then the Walsh transform of $x$ is $Wx$ and is denoted $\hat{x}$. Note that this definition uses a different scaling factor than the Walsh transform given by Vose [Vos99].