

Crossover and Diploid Dominance with Deceptive Fitness

Buster Greene

Buster Greene and Assoc.
6920 Roosevelt NE #126
Seattle, WA 98115
bgreene@accessone.com

Abstract. A deterministic approach is used to study the effects of recombination and diploid vs. haploid structure on EP efficiency. The deterministic analysis produces an exact result without resorting to multiple trials, at the cost of assuming an infinite population size. An efficiency limit is demonstrated for the recombination (crossover) rate. New evidence of a reduction is shown for the required growth time of diploid vs. haploid.

Complete diploid dominance is implemented in a manner which can be applied to any scalar EP, GA or GP problem, stationary or otherwise. In so doing, a dual interpretation of inter- and intra-gene fitness evaluation becomes apparent, and has a natural extension to vector fitness criteria. Results are consistent with previous non-deterministic, multiple trials that also used stationary fitness criteria.

1. Introduction

Analyzing EP behavior typically requires the use of multiple runs to account for different initial (generation 0) conditions. By using an infinite population, giving a *deterministic* model, multiple trials are avoided. As a result, exact solutions are obtained using a single run and a given set of parameters. Deterministic approaches were previously used by Bodmer (1967) and Eshel (1970) to study the effects of recombination on evolutionary efficiency. Felsenstein (1974) further summarized these effects with conclusions about whether and in what situations recombination might be helpful or detrimental. Those analyses used a 2-locus model, but without mutation. This provided insight and analytic simplification. Findings presented in this paper provide similar insight into the effects of recombination, but *with* the use of mutation.

The effects of recombination on EP efficiency have long been a subject of interest and controversy. This paper draws on previous in-depth studies of this issue to illustrate some effects in large populations, caused by varying the recombination, or crossover rate. The models chosen utilize a 2-locus model since this is the simplest that permits crossover.

Like mutation and recombination, diploid dominance can also be viewed as a genetic operator, although its utility has been difficult to establish. An approach to diploidy was described by Greene (1996) that follows a specific model known as *complete dominance*. Partial and "complete" dominance are well established mechanisms from biology that provide an effective explanation for dominance (e.g., see Stansfield (1983)). In addition, this implementation of diploidy requires no genotype modification, and has also shown promise with stationary (i.e., ordinary) fitness criteria. Specifically, implementation extends to real (EP) chromosomes or GP parse trees. For simplicity, the analysis used here will use EP for the experimental test bed.

Previous *non-deterministic* studies by Greene (1996) indicated that complete diploid dominance can provide a performance improvement in stationary GA's. That approach considers any *scalar* fitness GA to have a single *dominance* locus, and is identical to the approach used here. The term "dominance locus" refers to a distinct chromosome subset on each diploid homologue (i.e., gene) that functionally maps to a scalar fitness value. The *maximum* of the two resulting scalars is then the diploid individual's fitness. Those findings were extended in Greene (1998) using a multiple-allele, single dominance-locus model that permits deterministic analysis of diploidy under deceptive fitness conditions

2. Methods

2.1. Deterministic Analysis of Recombination

The deterministic model used to study recombination has 2 loci, each with 2 alleles. The 2 alleles at each locus are designated as a or A, at the first locus, and b or B at the second locus following previously mentioned analyses. The combination of alleles in each genotype map to a unique fitness, which can collectively be represented in a fitness matrix. This model, and specifically the use of 2 loci, permits the simplest possible study of the effects of varying crossover, or *recombination* rate. Mutation is introduced just prior to selection and recombination. Models are constructed for both haploid and diploid populations.

Following a 2-locus analysis by Bodmer (1967), the frequencies of the haploid gametes AB, Ab, aB and ab are designated as x_1, x_2, x_3 and x_4 , respectively, as shown in equation 2.1. This model will be used for both the diploid and haploid simulations. The diploid zygote *fitnesses*, which are ultimately applied to these gamete results, will be designated w_{ij} . A corresponding analysis for haploid is given by Kimura (1965). The difference equations, where x_i designates the proportion of the i^{th} genotype in the next generation (without mutation), are then:

$$x_i' = [x_i \bar{w}_i - rD] / \bar{w}, \quad i = 1..4 \quad (2.1)$$

where for diploid:

$$D = (w_{14}x_1x_4 - w_{23}x_2x_3)$$

$$\bar{w}_i = \sum_{j=1}^4 x_j w_{ij}$$

$$\bar{w} = \sum_{j=1}^4 \sum_{k=1}^4 x_j x_k w_{jk}$$

and for haploid:

$$D = (w_1w_4x_1x_4 - w_2w_3x_2x_3)$$

$$\bar{w}_i = \sum_{j=1}^4 x_j w_j$$

$$\bar{w} = \sum_{j=1}^4 x_j w_j$$

D is referred to as the genetic *disequilibrium*
r is the recombination, or crossover rate.

Similar to Bengtsson (1983), mutation is introduced prior to selection and recombination, by assuming that the rates of forward and backward mutation are identical (as is the case with EP). Designating \mathbf{x}^t as the vector of x_i values from above (these can be viewed as the haploid frequencies prior to recombination) gives the following:

$$\mathbf{x}^t = \begin{bmatrix} (1-u)^2, & u(1-u), & u(1-u), & u^2 \\ u(1-u), & (1-u)^2, & u^2, & u(1-u) \\ u(1-u), & u^2, & (1-u)^2, & u(1-u) \\ u^2, & u(1-u), & u(1-u), & (1-u)^2 \end{bmatrix} \mathbf{x}^{t-1} \quad (2.2)$$

A *deceptive* condition for the 2-locus problem can now be created from the 4 possible (haploid) genotypes. With an appropriate mapping of haploid to diploid fitnesses, as will be defined below, a comparison of EP diploid vs. haploid efficiencies can then be made. In the works previously

cited, this was done by studying conditions under which the AB genotype increases, when initially rare, but without mutation.

**I. Diploid Fitness:
By Haploid Gamete Genotypes from the
Two Parents**

	AB	Ab	aB	ab
AB	w_{11}	w_{12}	w_{13}	w_{14}
Ab	w_{12}	w_{22}	w_{23}	w_{24}
aB	w_{13}	w_{23}	w_{33}	w_{34}
ab	w_{14}	w_{24}	w_{34}	w_{44}

**II. Diploid Fitness:
Genotype States at Both Loci**

	BB	Bb	bb
AA	w_{11}	w_{12}	w_{22}
Aa	w_{13}	w_{23}	w_{24}
aa	w_{33}	w_{34}	w_{44}

**III. Haploid
Fitness**

AB	w_1
Ab	w_2
aB	w_3
ab	w_4

Figure 2.1. Diploid and haploid fitnesses for each genotype, as defined by Bodmer (1967). Diploid subscripts are taken from the 4x4 table that lists all gamete combinations. Table I simplifies to table II as long as the (diploid) fitnesses for Aa= aA and Bb = bB.

Notation for the diploid and haploid fitnesses are defined in figure 2.1. The diploid values are represented with double subscripts. It is assumed here that double heterozygote fitnesses are equal, resulting in $w_{14} = w_{23}$.

The *symmetric* diploid fitness model used is shown in figure 2.2. This model has $w_{11} = w_{14} = 2$ and $w_{44} = 1$, with all other elements set to zero. In order to establish an equivalent haploid fitness landscape, haploid fitnesses can be taken to be $w_1 = w_{14}$ and $w_4 = w_{44}$, with $w_2 = w_3 = 0$. More generally, the haploid fitnesses map to the lower right 2x2 diploid sub matrix. This is because $w_{11} = w_{14}$, and as a consequence, the rate of formation of ABs will be determined by the ability to form heterozygote diploid pairs, which is a function solely of the lower right sub matrix.

In the case of no mutation, at least some A's and B's must exist in the initial population for the optimal AB's to increase in proportion. For this reason, the initial proportions of Ab-s and aB-s (x_2 and x_3) are set to 10^{-10} .

Diploid				Haploid	
	BB	Bb	bb	AB	2
AA	2	2	0	Ab	0
Aa	2	2	0	aB	0
aa	0	0	1	ab	1

Figure 2.2. Diploid fitnesses used for the deceptive, deterministic model. Corresponding haploid fitnesses are taken from the diploid matrix as outlined in bold.

We can now define a simple "deceptive" fitness condition, namely, where $w_{24} = w_{34} < w_{44}$ and $w_{14} > w_{44}$. This corresponds, in the haploid case, to $w_2 = w_3 < w_4$, and $w_1 > w_4$. In either case, and when the global optimum, AB, is initially rare (in generation 0), Bodmer (1967) showed that the recombination rate without mutation must satisfy:

$$r < r_{\max} = \frac{w_{14} - w_{44}}{w_{14}} \quad (2.3)$$

for a diploid population to see an increase in ABs. For the haploid population this corresponds to:

$$r < r_{\max} = \frac{w_1 - w_4}{w_1} \quad (2.4)$$

Conceptually, an excessive crossover rate will cause heterozygotes to be broken up faster than they are created, and thereby completely prevent an increase in AB (optimal) genotypes. Bodmer showed, and iterative calculations shown in the results section confirm, that equation 2.3 is a hard limit when there is no mutation, and is independent of the diploid fitness ratio $w_{11}/w_{14} \gg 1$. A more thorough analysis of similar phenomena affected by crossover and fitness can be found in Felsenstein (1965).

Findings presented in this paper indicate that when mutation *is* present, the r_{\max} limit as defined by equation 2.3, is increased to a larger upper bound that will be designated: $r_{\max}^* \geq r_{\max}$. Experiments with various values of p_m and w_{14} (figures 2.1 and 2.2), in particular, indicate that the size of this inequality increases with the mutation rate as well as w_{14} . An analytic expression has yet to be found for the exact values for r_{\max}^* as a function of p_m and w_{14} .

2.2. Deterministic Analysis of Complete Diploid Dominance

"Complete" dominance, as mentioned, uses the 2 objective values defined by evaluating each diploid homologue. The maximum of these two scalars defines the diploid

individual's fitness. The homologue with highest fitness is then used as the active genotype.

As mentioned, this *scalar* fitness model corresponds exactly to a *single* locus for dominance purposes (homologue string length is arbitrary, and so the term "dominance locus" distinguishes from the usual GA definition of "locus"). As such, analysis of this single gene problem requires a multiple-*allele* model in order to introduce deception. Infinite population size results in exact difference equations that give each possible allele proportion at successive generations.

No recombination will be used here, for simplicity. Otherwise, the diploid population is complicated with two kinds of recombination. The first is the usual one, between individual, real string positions. In this case, crossover occurs *within* the dominance locus. The second applies to chromosomes having multiple genes¹. Such a situation permits recombination to occur *between* dominance loci. Also, the effects of recombination as will be shown, are not entirely obvious. For example, Felsenstein (1965) has shown that recombination can actually slow evolution, depending on the 2nd derivative of the fitness landscape².

2.3. Multiple Allele Deceptive Fitness Function

Mutation and selection are defined for M alleles at a single locus. For the experiments here, M=16. The fitness function used is:

$$f(i) = \begin{cases} 1 + \alpha; & i = 0, \alpha > 0 \\ (i-1)/(M-2); & i \in [1..M-1] \end{cases} \quad (2.5)$$

A value of $\alpha = 0.01$ was found to give reasonable difficulty with short run times, and was therefore used for the single locus experiments described in section 3. Equation 2.5 simulates a difficult portion of a hypothetical fitness landscape, and is similar to the unitary deceptive trap function used by Deb (1993, pg. 95). In order to require that the mutation operator should climb against the entire length of the linear trap, the initial population consists entirely of allele type $i = 15$. The rate of growth of (optimal) type 0's can then be controlled by, among other things, the size of α (larger α makes the problem easier).

A fundamental criterion to be used here, in comparing haploid vs. diploid efficiency, is the rate of increase for the

¹This would apply if the objective function is, itself, a vector of multiple genes.

²More exactly, the 2nd derivative of the log of fitness w/ respect to phenotype in hamming space.

globally optimal genotype³ when it is initially non-existent. The initial population therefore consists *entirely* of the allele that is maximally distant from the global optimum in *operator* space. The mutation "operator" is designed so that small allele changes are encouraged, as described below.

2.4. Multiple Allele (Single Locus) Mutation

Mutation is designed to act like a random walk from the current allele value, according to:

$$x_i^* = (1 - p_m) x_i + p_m \frac{\sum_{k=0, k \neq i}^{M-1} x_k (1 - |i - k| / M)^R}{\sum_{k=0, k \neq i}^{M-1} (1 - |i - k| / M)^R} \quad (2.6),$$

where x_i^* is the proportion of the i^{th} allele value after mutation (but prior to selection). x_i is the corresponding proportion before mutation, p_m is the mutation rate and M is the number of alleles. R controls the mutation *dispersal*, or extent of likely change from the current value, and is analogous to the *step-size* mutation parameter also used in EP (Fogel (1998, pg. 21)). Large R localizes the effect of equation (2.6) in the space defined by the range of ordered allele values. This makes the mutation operator more localized and generally results in slower, but more detailed hill climbing. Such localization is an implicit characteristic of all genetic algorithms, usually due to the relatively large genotype space compared to the space of possible mutations in a single individual per generation. Equation (2.6) is based on the mutation operator utilized by Michalewicz (1992, pg. 88) and uses a double exponential distribution to encourage mutation to allele values that are close to the current generation's value. For simplicity, R is held constant throughout a given run.

2.5. Multiple Allele Selection

The proportion of the i^{th} allele after selection is given by:

$$x_i' = \bar{w}_i / \bar{w}, \quad i = 1..M \quad (2.7)$$

Where: For diploid,

$$\bar{w}_i = \sum_{j=1}^M x_j^* w_{ij}$$

$$\bar{w} = \sum_{j=1}^M \sum_{k=1}^M x_j^* x_k^* w_{jk}$$

and for haploid,

$$\bar{w}_i = x_i^* w_i$$

$$\bar{w} = \sum_{j=1}^M x_j^* w_j$$

also:

$$x_i^* = \text{Proportion of } i^{\text{th}} \text{ allele after mutation}$$

$$w_i = \text{Fitness of the } i^{\text{th}} \text{ haploid genotype}$$

$$w_{ij} = \text{Fitness of the } ij^{\text{th}} \text{ diploid genotype}$$

$$M = \text{Number of alleles.}$$

Equation (2.7) describes the selection process for haploid and diploid populations, and occurs immediately after mutation. As may be seen, all genotype combinations are enumerated. Also, every possible haploid or diploid genotype, has a fitness that is defined by equation 2.5 and the use of complete dominance. x_i^* then gives the exact proportion (excluding roundoff error) for the i^{th} allele type in the next generation.

3. Results and Discussion

In order to facilitate a comparison with normal, finite population EP/GA implementation, the proportion of optimum alleles was used as the criterion for comparing haploid vs. diploid EP efficiency. Efficiency is measured in terms of generations required to reach a specific proportion. "Generations" indicate the number of (parallel) diploid or haploid fitness evaluations required to achieve proportions of 0.001 or 0.005 in the optimum genotype.

It should be noted that, if selection is not elitist (here it is not), the mere appearance of the global optimum does not guarantee that it will stay in a finite population, even to the next generation. Still, the run could be stopped when the global optimum is reached. In any case, such a criterion is nonetheless useful for comparing strategies and provides a low threshold that corresponds roughly to the emergence of a single optimal in populations of nominal size.

³Since the model used here interprets the chromosome as single gene, the terms allele and genotype are used interchangeably.

3.1. Deterministic Analysis of Recombination

Equations (2.1) were iterated using the fitness matrix defined in figure 2.2, to evaluate the effects of crossover (recombination) on optimal allele growth rates. In order to set r_{max} of equation 2.3 to a level consistent with GA usage ($r_{max}=0.5$), a fitness matrix having $w_1=2$ was used as shown in figure 2.2. A family of curves is obtained when run at different mutation rates that illustrates the limits on recombination given by equation (2.4), and the extension of that limit that apparently occurs with increasing mutation rate.

The results of this analysis are shown in figure 3.1. Specifically, the haploid growth rates in particular, show that excessive recombination not only inhibits evolution, but for at least this idealized case can completely stop it. What equations (2.3) and (2.4) do not show, and is indicated by figure 3.1, is that this inhibitory effect is mitigated by mutation.

For the fitness matrix of figure 2.2, growth curves for haploid and diploid at a lower crossover rate are shown in figure 3.3. Both this and figure 3.2 were run with r near r_{max} . Both figures' haploid runs, in particular, have growth rates suppressed by the presence of recombination.

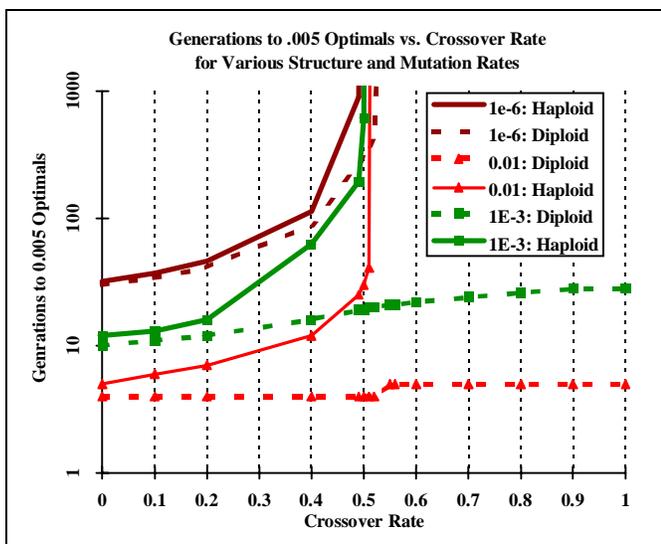


Figure 3.1 Results for fitness structure where $w_1=2$, which gives $r_{max}=0.5$.

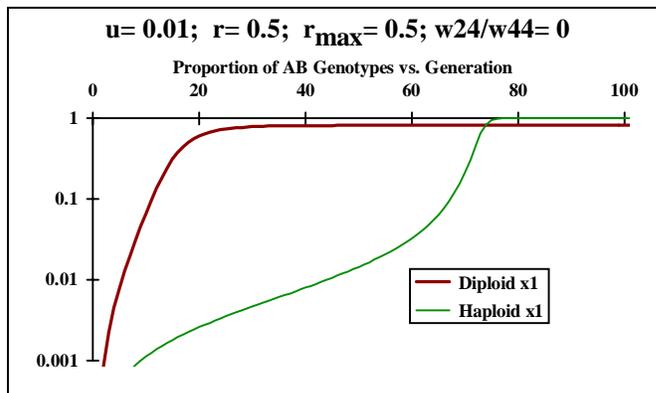


Figure 3.2. $r = r_{max}$ with mutation still slows increase in ABs. $w_{24}/w_{44}=0$ indicates complete deception.

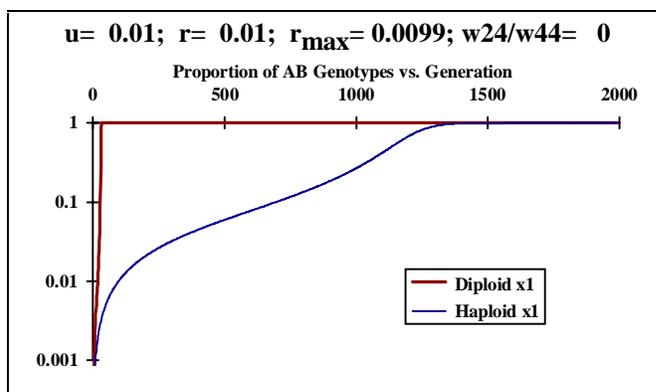


Figure 3.3. Recombination slightly greater than the $r_{max} = (w_{14} \cdot w_{44})/w_{14}$ limit. For this plot only: $w_1=1.01$, causing an order of magnitude slower growth rate.

Note that figure 3.1 suggests that a higher crossover rate *always* reduces evolutionary efficiency. In fact Felsenstein (1965) (pg. 357) shows under what conditions recombination will be beneficial. He shows that a criteria for this is the sign of the fitness 2nd derivative with respect to genotype as measured in Hamming space. In the present case the sign is positive, which indicates that increasing r will always reduce evolutionary efficiency.

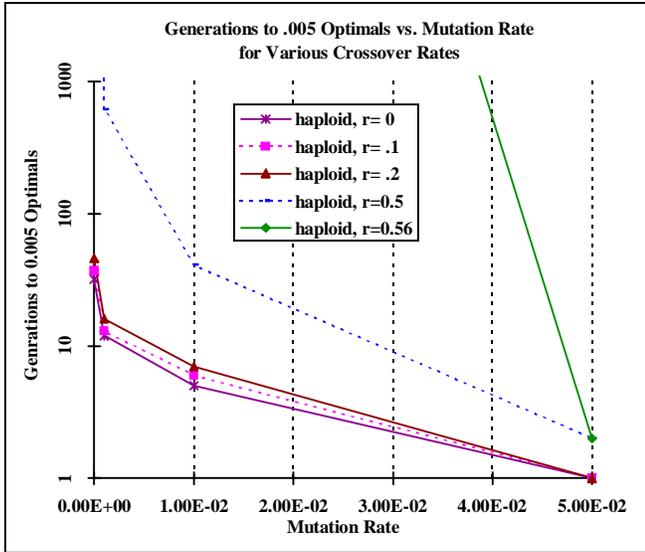


Figure 3.4. Haploid results of figure 1 plotted against mutation ($r_{max} = 0.5$).

Figure 3.4 gives a slightly more detailed look at the effects of haploid efficiency vs. mutation for r ranging from 0 to just past r_{max} . This result can also be more easily compared with the single allele result in section 3.2. Consistent with equation 2.4, as r exceeds r_{max} (in this case about $0.56 > r_{max}$), the production of optimals rapidly comes to a stop.

3.2 Deterministic Analysis of Diploid Dominance (Single Dominance Locus)

The single dominance locus, multiple allele deterministic model was iterated for both haploid and diploid populations, using the fitness function defined by equation 2.5. The experiments below are further analyzed in Greene (1998), and show the proportion of optimal alleles vs. generation for various mutation rates and with a fixed value of $R=6$. This value of R , and range of p_m are similar to that used by Michalewicz (1992) and Greene (1998).

The growth of the optimum allele starts at 0 and asymptotes to an equilibrium level in all cases. Figure 3.5 shows how the rates of growth for haploid and diploid can differ. Both curves will asymptote to different *equilibrium* values. In addition, one can reach a given proportion of optimals much sooner than the other (diploid wins at both the .001 and .005 levels in this case).

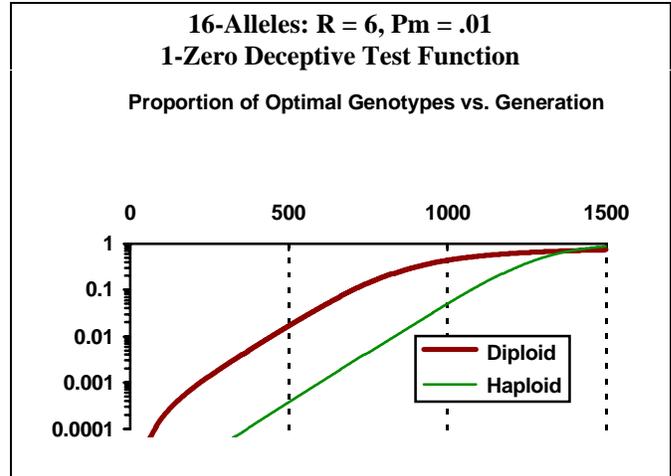


Figure 3.5. Proportion of optimal alleles vs. generation at $p_m = 0.01$.

Figure 3.6 plots the number of diploid and haploid generations that are needed to reach optimal proportions of 0.005 and 0.001. These 2 proportions are fixed for all experiments and are useful for comparing growth rates in the early stages of takeover by the optimal genotype.

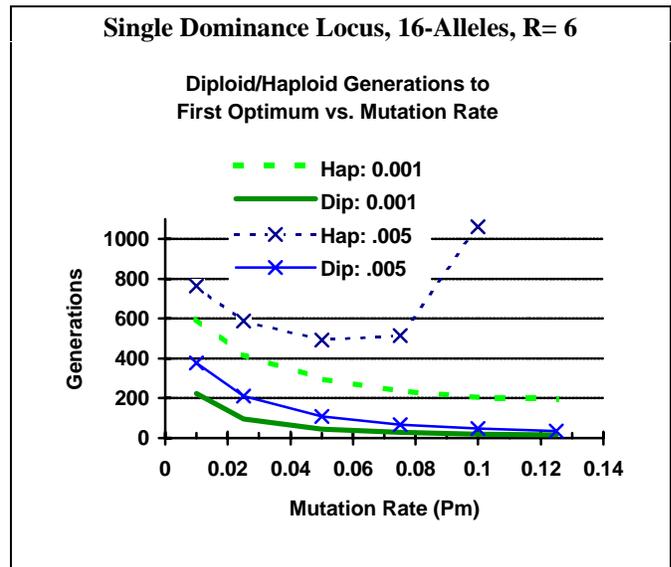


Figure 3.6. Comparison of haploid and diploid efficiencies for various mutation rates. Haploid fails to reach an optimal proportion of .005 for $p_m \geq 0.12$.

The haploid proportion asymptotes to approximately 0.0016 for increasing mutation rate and never reaches 0.005. At such an *equilibrium* point, mutation is destroying optimal alleles as fast as they are being created by selection. In a *finite* EP population of less than 200, with otherwise identical parameters, the optimum allele might remain absent from the population for most initial conditions.

The observed single locus tendencies are not specific to the one level of deceptivity (α). This is demonstrated in figure 8, which plots optimal creation time vs. the fitness deceptivity (α) parameter in the fitness equation. Note that haploid fails to reach an optimal allele proportion of 0.001 for $\alpha \leq 0.001$.

Reducing α increases deceptivity by decreasing the relative fitness of the global optimum, which makes the problem more difficult. This affect is show in figure 3.7 and appears to make the haploid problem difficulty increase much faster than is seen to occur for diploidy. Sufficiently small α (≤ 0.001) precludes the haploid proportion from ever reaching 0.001. In contrast to the recombination results, diploidy can clearly provide a performance boost, at least when fitness is deceptive.

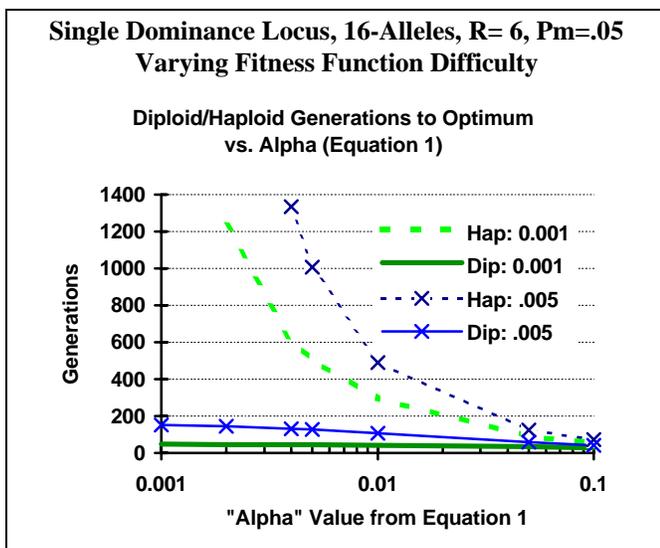


Figure 3.7. Comparison of haploid and diploid efficiencies. Decreasing α (equation 2.5) increases GA difficulty and increases the diploid efficiency advantage over haploid.

4. Conclusions

Recombination is shown to adversely affect evolutionary efficiency in the case of an infinite population with pure deception. This result is in line with theory developed by Felsenstein (1965). As shown, an increasing crossover rate will destroy favorable allele combinations faster than they are generated by selection, and can in the case of no or too little mutation, completely stop evolution. As the mutation rate is decreased, the maximum allowable recombination rate is confirmed to be solely dependent on the w_{44} and w_{14} fitnesses. in equation 2.3. This result is in agreement with the findings of Bodmer (1967).

A method for introducing mutation to previous deterministic approaches is proposed, along with a

straightforward mapping of diploid to haploid fitnesses. This mapping makes it possible to compare 2-locus simulation results with previous findings that utilized a multiple allele, single dominance locus model. The mapping is also reasonable in its behavior, and growth rates under a deceptive condition are restricted according to equation 2.4. In particular, deceptive fitness with no mutation causes evolution to cease at the predicted recombination rate. A finding of this paper is that an increasing mutation rate extends the limit on the allowable recombination rate.

A typical GA fitness landscape is assumed to contain one or more deceptive regions. It should be noted that non-deceptive regions are also common, e.g., in the neighborhood (of operator space) local maxima. Specifically, if fitness asymptotes to a zero slope at a particular local maximum, the fitness 2nd derivative will of necessity be negative. This is one common situation where recombination might be expected to promote evolution. A deceptive trap may still, however, be sufficiently worsened by recombination to effectively stop evolution prior to this. The end result will depend not only on the fitness landscape but in addition, on finite population effects such as initial conditions, that can drastically affect the fitness trajectory.

The observed loss due to recombination, under deceptive conditions, appears to be less severe with diploid than haploid. This particular conclusion implicitly assumes that homologous (diploid) genes are expressed in parallel. Completely parallel gene expression (as found in nature) implies that $w_{14} = w_{23}$, which is the situation here. Even serial gene evaluation may have implicitly parallel behavior, however, if the fitness function allocates partial credit to isolated, but functional genes.

A general diploid speedup *is* fairly conclusive for the single locus, multiple allele results of section 3.2. Moreover, diploidy may provide a speedup in one situation where recombination apparently does not, namely, where the fitness landscape is deceptive.

While deception implies a positive 2nd fitness derivative, with an attending adversity to recombination, the mere lack of deception does not necessarily imply the opposite. For example, Forrest (1992)(pg. 114) notes that "if intermediate stepping stones are too much fitter than the primitive components..." an seemingly crossover friendly problem may still have a positive 2nd derivative and thereby become harder to solve using crossover. The implication is that such a situation may be present with some royal road problems.

Increasing mutation generally speeds up evolution, with some possible exceptions. First, an extremely high mutation

can be detrimental as shown in the 0.001 haploid curve of figure 3.6. This is as expected. Less obvious is that when r is just greater than r_{\max} , slightly insufficient mutation can cause evolution to quickly shut down. This is shown in figure 3.1 where the haploid curves are near $r = 0.5$. Conversely, increasing the mutation rate may have little or no benefit if r is sufficiently greater than r_{\max} .

The results suggest that some insight can be gained into the effects of recombination and diploidy on an idealized GA. Findings are exact, and are made possible due to the use of a deterministic approach. In particular, there is now evidence that too much recombination can under certain circumstances inhibit evolution, particularly when fitness is deceptive. In addition, sufficient mutation can be essential to prevent very destructive recombination. A direct implication is that tests made under idealized conditions of deception may evolve faster without recombination. Finally, while recombination apparently slows evolution in deceptive cases, the evidence suggests that a diploid population may do better than a haploid population in such cases. Subsequent work will likely address finite population effects, more complex fitness landscapes or issues related to royal road experiments.

Bibliography

- Bengtsson, B. (1983). A two-locus mutation-selection model and some of its evolutionary implications. Theoretical Population Biology, 24, 59-77.
- Bodmer, W. F., and Felsenstein, J. (1967). Linkage and selection: Theoretical analysis of the deterministic two locus random mating model. Genetics, 57, 237-265.
- Deb, K., Goldberg, D. (1993). Analyzing Deception in Trap Functions. In D. Whitley (Eds.), FOGA-2 (pp. 93-108). San Mateo: Morgan Kaufmann.
- Eshel, I. a. F. M. (1970). On the evolutionary effects of recombination. Theoretical Population Biology, 1, 88-100.
- Felsenstein, J. (1965). The effect of linkage on directional selection. Genetics, 52(2), 349-363.
- Felsenstein, J. (1974). The evolutionary advantage of recombination. Genetics, 78, 756-737.
- Fogel, D. (Ed.). (1998). Evolutionary Computation: The Fossil Record. New York: IEEE Press.
- Forrest, S., and Mitchell, M. (1992). Relative building-block fitness and the building-block hypothesis. In Whitley (Eds.), Foundations of Genetic Algorithms (pp. 109-126). San Mateo: Morgan-Kaufmann.
- Greene, B. (1998). A deterministic analysis of stationary diploid/dominance. In Proceedings of the 3rd Conference on Genetic Programming, (pp. 770-776). Madison: Morgan Kauffman.
- Greene, F. (1996). A new approach to diploid/dominance and its effects on stationary genetic search. In D. Fogel (Ed.), Proceedings of the Fifth Annual Conference on Evolutionary Programming, . San Diego: MIT Press.
- Kimura, M. (1965). Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. Genetics, 52, 875-890.
- Michalewicz, Z. (1992). Genetic Algorithms + Data Structures = Evolution Programs. Berlin: Springer-Verlag.
- Stansfield, W. D. (1983). Schaum's Outline of Theory and Problems in Genetics (2 ed.). McGraw-Hill.