Analysis of Mixing in Genetic Algorithms: A Survey

Kumara Sastry

Illinois Genetic Algorithms Laboratory (IlliGAL)
Department of General Engineering
University of Illinois at Urbana-Champaign
104 S. Mathews Ave, Urbana, IL 61801
ksastry@uiuc.edu

Abstract

Ensuring building-block (BB) mixing is critical to the success of genetic and evolutionary algorithms. There has been a growing interest in analyzing and understanding BB mixing and it is necessary to organize and categorize representative literature. This paper presents an exhaustive survey of studies on one or more aspects of mixing. In doing so, a classification of the literature based on the role of recombination operators assumed by those studies is developed. Such a classification not only highlights the significant results and unifies existing work, but also provides a foundation for future research in understanding mixing in genetic algorithms.

1   Introduction

Since the inception of genetic algorithms (GAs), the importance of building blocks have been recognized (Holland, 1975; Goldberg, 1989b). Based on Holland's notion of BBs, Goldberg proposed a design decomposition method for a successful design of GAs (Goldberg, 1991; Goldberg & Liepens, 1991; Goldberg, Deb, & Clark, 1992). This design decomposition currently consists of sevens steps (Goldberg, in press) and can be stated as follows: (1) Know what GAs process—building blocks (BBs), (2) solve problems that are of bounded BB difficulty, (3) ensure an adequate supply of raw BBs, (4) ensure increased market share for superior BBs, (5) know BB takeover and convergence times, (6) ensure that BB decisions are well made, and (7) ensure a good mixing of BBs. Significant progress has been made in developing facetwise models for many of the above decomposition steps and the interested reader should consult Goldberg (in press) and the references therein for further details.

However, researchers have often overlooked the issues of BB identification and mixing/exchange, even though studies on selectorecombinative GAs have indicated that effective *identification* and *exchange* of BBs is critical to innovative success. Furthermore, existing facetwise models such as convergence-time and population-sizing models *assume* tight linkage. That is, alleles of a BB were assumed to be close to one another, and crossover operators are assumed to ensure necessary exchange of BBs with high probability. Even though, the assumption of tight linkage isolates the phenomenon of interest while bracketing the linkage problem, in real-world problems this is not the case, as we don't know which alleles contribute to which BBs *a priori*. This necessitates the incorporation of mixing of BBs into GA dynamics.

It is therefore critical to perform an exhaustive survey of work related to BB mixing in genetic and evolutionary algorithms. Performing such a survey is the objective of this paper. In doing so, existing studies on mixing are classified according to the facets of mixing they model. This

classification not only organizes and highlights significant results, but also unifies representative studies as a whole.

This paper is organized as follows: Section 2 presents a brief overview of genetic algorithms. Section 3 describes the categorization used in surveying the studies on BB mixing. Studies belonging to each of the four categories are discussed in sections 4–7. Finally section 8 presents the summary and key conclusions of the study.

2   A Quick Introduction to Genetic Algorithms

This section presents a brief outline of a simple GA; A more detailed exposition is given elsewhere (Goldberg, 1989b). GAs are search procedures based on the principles of natural selection and genetics. Analogous to genetics, GAs encode the decision variables of the problem at hand into finite-length strings of alphabets of certain cardinality. The letters present in a string are referred to as *genes* or *alleles*. Usually, the parameters are encoded as a fixed length binary string. This string of alleles is referred to as a *chromosome* or an *individual*. In contrast to traditional optimization techniques, GAs work with coding of parameters, rather than the parameters themselves.

Once the problem has been represented, another thing to be determined is the method of deciding between good and bad solutions. This can either be interactively decided by a human or be the result of complex computer simulation. GAs do not differentiate between these different modes of fitness assignments as long as good solutions have higher *fitness* when compared to bad solutions. With the issues of representation and fitness functions decided, we can proceed with the evolution process. To evolve new solutions, an initial population of encoded solutions is created randomly or using some problem-specific knowledge. This population is subjected to genetic operators to create new promising solutions. Different operators exist in GAs, the most popular being (1) *selection*, (2) *crossover*, and (3) *mutation*. This newly created set of individuals replace the old population and the process is continued till some criteria are satisfied.

The selection procedure follows the survival-of-the-fittest principle and allocates more copies to better individuals. Various techniques exist for implementing this idea. They can be broadly classified into two classes, (1) *proportionate* schemes like roulette-wheel selection (Goldberg, 1989b) and stochastic universal selection (Baker, 1985), and (2) *ordinal* schemes like tournament selection (Goldberg, Korb, & Deb, 1989), and truncation selection (Mühlenbein & Schlierkamp-Voosen, 1993).

Recombination or crossover combines parental solutions to form offspring that are likely to be better solutions. Recombination operators are critical in ensuring good mixing of BBs. Various recombination procedures exist in GA literature and a comprehensive description of the same is out of the scope of this study. Instead, a brief description of the crossover techniques that are commonly used is presented. Specifically, one-point crossover, two-point crossover, and uniform crossover methods are discussed. For details on other crossover operators the interested reader should refer elsewhere (Goldberg, 1989b; Booker, Fogel, Whitley, & Angeline, 1997; Spears, 1997) and the references therein.

For the aforementioned three recombination procedures, individuals in the population are paired randomly. Recombination is performed on each pair with a probability equal to the crossover probability, p_c, to obtain two new offspring. In one-point crossover (figure 1), a crossover site is selected at random over the string length, and the alleles on one side of the site are exchanged between the individuals. In two-point crossover (figure 1), two crossover sites are randomly selected. The alleles between the two sites are exchanged between the two randomly paired individuals. Two-point crossover is illustrated in figure 1. An n_c-point crossover can be generalized using a
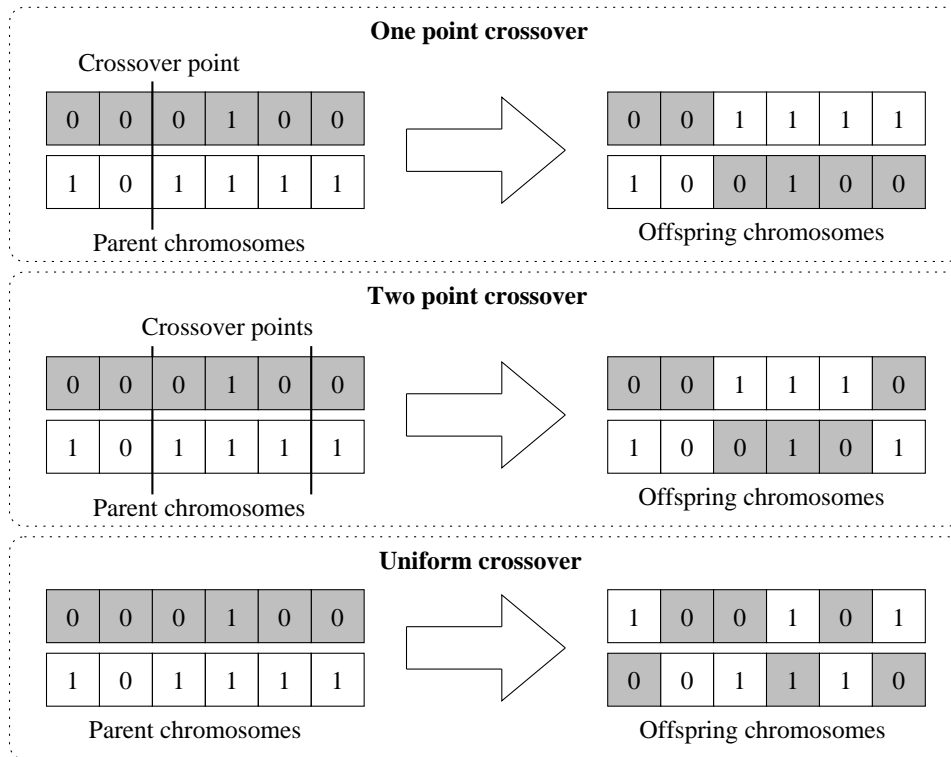
Figure 1: Illustration of one-point, two-point, and uniform crossover methods.

two-point crossover. In uniform crossover (figure 1), every allele is exchanged between the two random individual pairs with a certain probability, p_e, known as the swapping probability. Usually the swapping probability value is taken to be 0.5.

In contrast to recombination which operates on two individuals, mutation operates on a single individual and modifies it slightly. Various mutation procedures exist in GA literature, the most common being the bitwise mutation. In bitwise mutation, each bit in a binary string is changed (a 0 is converted to 1, and vice versa) with a certain probability, p_m, known as the mutation probability. Mutation performs a random-walk in the vicinity of the individual.

Elsewhere, it has been shown that though these operators when analyzed individually are ineffective, however, when combined together they can work well (Goldberg, 1999; Goldberg, in press). This aspect has been explained with the concepts of the *fundamental intuition* and *innovation intuition*. The same study compares a combination of selection and mutation to *continual improvement* (a form of hill climbing), and the combination of selection and recombination to *innovation* (*cross-fertilizing*).

With this brief introduction to GAs, the ground work for introducing BB-mixing has been laid. The following section defines what we mean by the mixing problem, and presents the classification of BB-mixing models developed to understand one or more facets of the mixing problem.

3 Classification of Building-Block-Mixing Models

Holland made two key observations in his monograph (Holland, 1975): (1) BBs with tighter linkage (shorter defining length) have a selective advantage over those with loose linkage (longer defining length), and (2) operators to adapt linkage and to choose allele combinations might be necessary for

GA success. Goldberg, Deb, and Clark (1992) proposed the design decomposition for a successful design of GAs, a core part of which is BB identification and mixing. However, Success stories of GAs with fixed codings and crossover operators have somewhat masked the importance of investigating and designing operators that adapt linkage and efficiently mix BBs (Goldberg, in press).

On a parallel note, it is now known that many operators such as *elitism*, *niching*, *mating restriction*, *inversion*, and *reordering*, proposed to adapt linkage and to enhance GA performance do not achieve their goal. Thierens (1995) showed that elitism and niching either alone or in cohort do not address the mixing issue effectively. He also suggested that any crossover with fixed scheme is highly unlikely to adapt linkage. Bagley (1967) and Frantz (1972) provided empirical evidence to indicate that Holland's (1975) inversion operator does not learn linkage effectively. Goldberg and Bridges (1990) demonstrated that simple reordering operators have limited linkage-learning capability.

Failures with such operators suggest that the critical step in designing better crossover operators is to first analyze fixed crossover operators such as uniform, one-point, and n-point crossovers. Such analysis should tell us if fixed crossover operators are powerful enough to solve problems of bounded difficulty quickly, reliably, and accurately. They should also give us insights as to why such fixed crossover operators fail if they do fail. The question as to how well fixed crossover operators solve GA-easy as well as GA-hard problems is called the mixing question. Many researchers have analyzed different aspects of fixed point crossovers and some of them contain answers to the mixing question. These studies can be broadly classified into four categories based on what they assume the principal role of recombination to be:

1. recombination as a mixer,

2. recombination as a schemata disrupter,

3. recombination as an innovator, and

4. recombination as a mixer, schemata disrupter, and innovator

Each of these roles in briefly explained in the following paragraphs. Next four sections will describe the work done and discuss significant results in each of these categories.

Recombination as a mixer:

Studies belonging to this category model recombination as a tool that mixes solutions in a population (Eshelman, Caruana, & Schaffer, 1989; Booker, 1993; Rabani, Rabinovich, & Sinclair, 1998). They treat recombination as a main source for maintaining diversity in the population. Many of these models are motivated from studies in population or quantitative genetics (Bailey, 1961; Bulmer, 1985; Falconer, 1989; McPeek, 1996). Such models quantify, linkage-disequilibrium, relaxation time, and crossover-induced biases. These models neither consider the effects of schema disruption nor the effects of innovation (mixing of good schemata). Such models are useful in understanding the relative advantages/disadvantages of crossover operators. The results of such models are independent of the search problem, which is both an advantage and a disadvantage. On one hand, the results are universal once the problem coding is defined. On the other hand, their applicability is limited as they incorporate no problem knowledge and linkage is problem specific.

Recombination as a schemata disrupter:

Studies belonging to this category model crossover as disrupter of building blocks (De Jong, 1975; Syswerda, 1989; De Jong & Spears, 1992). Such studies are mainly based on the schema the-

orem (Holland, 1975) and quantify schema survival rates. Elsewhere it has been suggested that schema theorem should be obeyed, and it can be done be easily (Goldberg, Sastry, & Latoza, 2001). Furthermore, satisfying schema theorem does not guarantee BB mixing. It should be noted that schema theorem only focuses on a single BB, but for a GA success (for innovation and successful recombination of best BBs),we must model some aspects about the likelihood of different BBs coming together. Nevertheless, schemata-disruption models are useful in comparing different crossover operators.

Recombination as an innovator:

Studies that visualize recombination as a innovation operator model the likelihood of combining different BBs (Goldberg, Thierens, & Deb, 1993; Thierens & Goldberg, 1993; Thierens, 1996; Thierens, 1999). These facetwise models not only integrate with other models on dimensional grounds, but also yield a *control map* delineating the region of good performance for a GA. Such a control map can be an useful tool in visualizing GA sweet-spots and provide insights in parameter settings (Goldberg, 1999). Such models incorporate problem complexity, in terms of the BB size and number of BBs. However, existing models are valid only for uniform crossover and have to extended to incorporate other crossover operators.

Recombination as a mixer, schemata disrupter, and innovator:

These models combine all the facets of recombination operators (Bridges & Goldberg, 1987; Vose & Liepens, 1991; Nix & Vose, 1992; Stephens & Waelbroeck, 1999; Prügel-Bennett & Shapiro, 1994) using tools such as difference equations, differential equations, Markov chains, and statistical mechanics. They tend to model workings of recombination exactly or more accurately than any of aforementioned three classes of models. That is, they contain all the information about BB mixing that we are interested in. Such information is usually hidden among many other facets of GAs, making these models highly complex. This makes it very hard to extract information on BB mixing (or for that matter, any other single aspect of GAs) out of such models. Unless these models can be simplified, or their asymptotic behavior predicted, it is very difficult to use them to design better GAs or better operators.

4   Recombination As A Mixer

This section discusses representative studies on modeling recombination as a tool for mixing alleles. These models are mainly motivated by analytical framework of population genetics. A detailed description of such a framework is beyond the scope of this study and the interested reader should consult elsewhere (Bailey, 1961; Bulmer, 1985; Falconer, 1989; McPeek, 1996). In these methods crossover is analyzed in isolation, that is, in the absence of mutation and selection. Central to such an approach is *Geiringer's theorem* (Geiringer, 1944).

Geiringer's theorem describes the equilibrium distribution of a population undergoing recombination alone. The theorem states that random mating and recombination without selection lead to chromosome frequencies corresponding to the simple product of initial allele frequencies. A population in such a state is said to be in *linkage equilibrium* or Robbin's equilibrium (Robbins, 1918). The equilibrium distribution can be written as

$$\lim_{t \to \infty} p_t(S) = \prod_{i=1}^{\ell} p_{i,0}(S) \tag{1}$$

5

where, $p_t(S)$ is the proportion of string S at time t and $p_{i,0}(S)$ is the proportion of the ith allele of S in the initial population.

Studies that model recombination as a mixing tool quantify one of the following three properties:

1. *relaxation time*, which measures the rate at which recombination operators converge to linkage equilibrium.

2. coefficient of *linkage disequilibrium*, which measure the deviation of chromosome frequencies at time t from their equilibrium frequencies.

3. Crossover-induced biases such as *positional bias*, *distributional bias*, *recombinative bias* and *schema bias*

Models developed quantifying each of the above three properties is presented in the following two subsections.

4.1 Relaxation-Time and Linkage-Disequilibrium Models

Christiansen (1989) developed analytical framework for the convergence to linkage equilibrium. He suggested that the population dynamics of a population undergoing repeated recombination is governed by *marginal recombination distributions*. He quantified the coefficient of linkage disequilibrium and provided theoretic support that recombination operators that are more disruptive reach linkage equilibrium more quickly. However, since his analysis is pertinent to the field of mathematical genetics, it is not readily applicable to conventional analysis in GAs.

Recently, Rabani, Rabinovich, and Sinclair (1998) modeled crossover as a nonlinear quadratic dynamical system (QDS) and provided tight bounds on relaxation time. They showed that crossover operators can be successfully analyzed as a QDS, even with finite population sizes. Their study overcame earlier claims of Arora, Rabani, and Vazirani (1994) and Pudlák (2001) who suggest that QDS is unlikely to simulate the behavior of crossover. Rabani, Rabinovich, and Sinclair (1998) analyzed uniform, one-point, and Poisson model crossover (Haldane, 1919) and quantified the asymptotic rate of convergence to the stationary equilibrium for each of those operators. The relaxation time, t_m as computed by Rabani, Rabinovich, and Sinclair (1998) for uniform and one-point crossover can be summarized as follows:

- Uniform crossover:
$$\log_2(\ell) - O(1) < t_m \leq 2\log_2(\ell) + \log_2 \epsilon^{-1}. \tag{2}$$

  where ℓ is the string length, and $\epsilon \in (0, 1]$ is the asymptotic distance from linkage equilibrium.

- One-point crossover:
$$\ell \ln \ell \left(\frac{1}{2} - o(1)\right) < t_m \leq \ell \ln \ell + \ell \ln \epsilon^{-1}. \tag{3}$$

In other words, relaxation time for uniform and one-point crossover is $O(\ln \ell)$ and $O(\ell \ln \ell)$ respectively.

Prügel-Bennett (2001) also quantified the mixing rates of uniform, one-point and two-point crossovers. He modeled mixing as a generalized card shuffle problem. In the card shuffle problem, hands refer to strings or chromosomes, cards refer to genes, and suits correspond to different alleles. He defined an order parameter to measure the degree of mixing within the population. The order parameter is defined such that it decays as the strings becomes mixed. His approach is a more detailed, yet intuitive. However, the approach used by Rabani, Rabinovich, and Sinclair (1998) is

more general and is based on a rigorous mathematical framework. Prügel-Bennett (2001) computed that the relaxation time of uniform, one-point and two-point crossover is $O(\ln \ell)$, $O(\ell \ln \ell)$, and $O(\ell \ln \ell)$ respectively. It should be noted that the results of Rabani, Rabinovich, and Sinclair (1998) and Prügel-Bennett (2001) agree, even though they used different approaches.

Spears (2001) used marginal recombination distribution to analyze the transient behavior of crossover operators. He investigated the rate of approaching linkage equilibrium for second order BBs (BB size = 2) with different defining lengths. He analyzed the effects of uniform and one-point crossover. He modeled the rate of approaching the limiting distribution as a set of coupled nonlinear differential equations.

4.2 Crossover-Induced Bias Models

The models discussed in the previous subsection considered the convergence rate of attaining linkage equilibrium. The models presented in this subsection use recombination distributions to quantify bias introduced by crossover operators. Eshelman, Caruana, and Schaffer (1989) attributed two types of bias induced by crossover operators: (1) *distributional bias*, and (2) *positional bias*. Distributional bias refers to the number of loci transmitted during a recombination event and the extent to which some values might be more likely than others. Positional bias refers to the extent of dependence of the probability that a set of loci will be transmitted together depends on the relative positions of those loci on the chromosome. An operator with high positional bias is less likely to disrupt schema which shorter defining length. On the other hand, an operator with low distributional bias disrupts less number of schemata, but will also have less exploratory power.

Eshelman, Caruana, and Schaffer (1989) quantified both distributional and positional biases introduced by different crossover operators including uniform, one-point and n_c-point crossovers. They showed that one-point crossover has high positional bias and low distributional bias, and uniform crossover has low positional bias and high distributional bias. They also showed that n_c-point crossover has lower positional bias when compared to one-point crossover, but has a higher positional bias when compared to uniform crossover. The distributional bias of n_c-point crossover is higher than that of one-point crossover, but is lower than that of uniform crossover.

Booker (1993) revisited Geiringer's work (Geiringer, 1944) and computed the marginal recombination distributions for different crossover operators similar to the approach used in mathematical genetics (Christiansen, 1989). Booker (1993) used these marginal distributions to quantify both the positional and distributional biases induced by uniform, one-point and n_c-point crossover. He obtained results similar to those of Eshelman, Caruana, and Schaffer (1989) and validated the usage of marginal recombination distributions to analyze crossover operators in GAs. Rana (1999) analytical description for the distributional bias induced by uniform, one-point, two-point and HUX (Eshelman, 1991) crossover. Rana quantified distributional bias as the distribution of Hamming distances between parents and offspring. Such a model quantifies the step sizes provided by different crossover operators in the initial generation.

Eshelman and Schaffer (1995) proposed a generalization of distributional bias called *recombinative bias*. Recombinative bias refers to the expected proportion of the differing bits that a crossover operator copies to a child from its furthest parent (in terms of hamming distance). They also proposed a generalization of positional bias called *schema bias*. Schema bias is the extent to which certain schemata are favored over others by a crossover operator. They also provided empirical evidence to suggest that a crossover operator with strong recombinative bias and low schema bias to overcome premature convergence. The also observed strong recombinative bias to be a liability when solving GA-hard problems.

Altenberg (1995) used Geiringer's theorem to predict the evolution of fitness distribution under recombination. He used marginal recombination distribution similar to the approaches used by Booker (1993). Altenberg (1995) incorporated the marginal recombination distributions along with Price's theorem (Price, 1970) for selection into a schema theorem. He suggested that schemata disruption is quantified by linkage disequilibrium and proposed a guidance measure for improving recombination operator. This study, bridges the models belonging to this section and models we will discuss in the next section, in the sense that it links schemata disruption with linkage disequilibrium.

5   Recombination as a Schemata Disrupter

These studies follow along the lines of the schema theorem and model crossover as a disruption operator. They model the schema disruption rates as a function of its defining length for different crossover operators. However, these models still consider only one BB and thus fail to address the mixing question posed in Section 3. Similar to the studies quantifying crossover-induced bias, schema-disruption models provide us with a measure, albeit incomplete, to compare different crossover operators.

It should be noted here that there are many other studies related to schema theorem (Goldberg, 1989b; Goldberg, in press), with the majority of them addressing the issue of BB growth. For example, Goldberg and Sastry (2001) have used schema theorem to develop control maps between selection pressure and crossover probability, but such a control map is the result of BB growth analysis. Furthermore, they showed that obeying schema theorem does not guarantee BB mixing. Therefore, studies belonging to this category are not included in this survey.

Syswerda (1989) analyzed schema survival rates and average number of schemata combination for uniform, one-point and two-point crossover methods. He further provided empirical evidence to suggest that uniform crossover outperformed one-point and two-point crossover methods in most of the cases. Spears and De Jong (1991) analyzed multi-point crossover methods in terms of sampling disruption and compared them to uniform crossover. Their study extended the work of De Jong (1975) by extending the analysis to incorporate k^{th} order schemata. Their study indicated that uniform crossover has a much higher schema disruption rate when compared to one-point and n_c-point crossover. However, they also found that when the defining length is very long ($\sim \ell$), then uniform crossover has a lower schema disruption rate when compared to one-point crossover. Based on their analysis, they suggested that disruption has a positive role to play in balancing the exploration and exploitation during the adaptive search (Spears & De Jong, 1991; De Jong & Spears, 1990). De Jong and Spears (1992) further investigated the effects of crossover operators on the population size. They empirically observed that uniform crossover performed better with small population sizes and two-point crossover performed better with large population sizes. However, they did not give any analytical framework to incorporate the effects of recombination operator on population sizing.

6   Recombination as an Innovator

Models belonging to this category address the BB mixing issue in a more direct manner than those belonging to other categories. These models develop models for predicting *mixing* or *innovation* time. Here, mixing time is defined as the expected number of generations to obtain an instance of the target string. Such models are not only intuitive, but also easy to analyze and compare with other facetwise models using dimensional arguments. Such a comparison can lead to the

construction of a *control map* clearly identifying different competing forces affecting the genetic search.

Goldberg, Thierens, and Deb (1993) addressed the issue of allele-wise mixing building upon Holland's suggestion that limiting probability of success of BB exchange is allele-wise random assortive (Holland, 1975). Goldberg, Thierens, and Deb (1993) calculated the mixing time, t_x, for allele-wise exchange, assuming a population of size n, geometric mean of proportion of target allele p, crossover probability p_c, and a string length ℓ:

$$t_x = \frac{1}{n p_c p^\ell} \tag{4}$$

The compared this model with selection time model (Goldberg & Deb, 1991), drift time (Goldberg & Segrest, 1987) and cross-competition models to construct mixing boundary, drift boundary and cross-competition boundary. They identified the following key points in computing these boundaries:

- When mixing time is less than selection time, then good BB exchange in ensured.

- On the other, hand if mixing time is greater than selection time, then a GA will converge prematurely

- When the selection time is greater than the drift time then, GA drifts and converges to alleles with little or no selection pressure

- When selection pressure is high, BBs in different partitions might compete. Such a competition between non-competing alleles is called *cross competition*.

Goldberg, Thierens, and Deb (1993) used these boundaries to construct a control map on the crossover probability versus selection pressure $(p_c - s)$ plane. They also verified their facetwise models and the control map with empirical evidence. They observed that simple GAs solving easy problems have a large sweet-spot (Goldberg, 1999). As a consequence of a large sweet-spot, a simple GA will successfully solve easy problems with a large range of parameters values to choose from.

The allele-wise mixing model of Goldberg, Thierens, and Deb (1993) was further expanded and used for GA design by Thierens (Thierens, 1995; Thierens, 1996). Specifically, he quantified the interplay between (1) string length and selection pressure, (2) disruption probability and string length, (3) population size and disruption probability, and (4) population size and string length. He showed that for a GA easy problem, the effect of mixing on population size dominates the effect of selection for small string lengths. For larger problems (longer string lengths), the effect of selection on population sizing dominates the effect of mixing.

Thierens and Goldberg (1993) extended the BB mixing model to incorporate BB wise mixing. One of their objectives was the investigate if the sweet-spot remained large enough for boundedly difficult problems. Specifically, they considered *deceptive* trap problems with uniform building-block scale and size (Goldberg, 1987; Goldberg, 1989a; Deb & Goldberg, 1993). They developed a mixing-time model for the case of two BBs of size k and then extended it for the case of multiple BBs of size k. For a problem with m building blocks of size k, the mixing time is given by (Thierens & Goldberg, 1993; Thierens, 1995; Thierens, 1999),

$$t_x = c \frac{2^{\mu k}}{n p_c} \frac{2^m}{m^{\frac{5}{2}}}, \tag{5}$$

where c and μ are constants. Furthermore, comparing this with selection time yields a population-sizing model to satisfy mixing:

$$n \ln n > c \frac{2^{\mu k}}{n p_c} \frac{2^m}{m^{\frac{5}{2}}} \ln s. \tag{6}$$

The above result clearly indicates that the sweet-spot shrinks rapidly as the problem size increases. Their model also showed that for a GA success, the BBs have to be tightly linked in the problem coding structure. One of the main outcome of their analysis is that they provided sufficient evidence to indicate that, to solve boundedly difficult problems in polynomial time, operators that adapt linkage are required. In other words, as the problems become more difficult, population sizes much grow exponentially to ensure that a simple GA using fixed crossover operators converges to good solution (Thierens, 1995; Thierens, 1999). This important result has led to the design of many successful designs of operators that identify and adapt linkage (Goldberg, in press).

7   Recombination As A Mixer, Schemata Disrupter, and Innovator

The studies discussed in the previous three sections investigate one or more aspects of recombination operators in isolation. Such facetwise models are easy to analyze and have been very helpful in understanding and designing GAs. However, in building such facetwise models, we often neglect or assume certain things about other GA aspects. Researchers have developed more complex models that take into account the interplay between the genetic operators. Such models tend to be powerful in simulating GAs and yield exact results. However, usually the property of interest is hidden under the complex formulation, making it difficult to extract the information needed. Therefore the applicability of such complex, yet accurate models in designing competent GAs is limited. Nevertheless, they are still useful and can potentially guide us as to when the facetwise models are applicable or when they are not. They can also validate the facetwise models through some asymptotic analysis.

Bridges and Goldberg (1987) modeled the effects of selection and recombination as a difference equation. Specifically, the considered fitness proportionate selection and one-point crossover. Their analysis of crossover included not only BB disruption, but also BB innovation. They also extended their analytical model to facilitate the computation of schemata propagation. Vose and Liepens (1991) developed difference equations to model the dynamics of a GA similar to those developed by Bridges and Goldberg (1987). Nix and Vose (1992) used the dynamics of recombination operator developed by Vose and Liepens (1991) and proposed a Markov-chain model to simulate GAs. The framework proposed by Nix and Vose (1992) is significant as it proposed a viable method to incorporate finite population into the Markov-chain model. The key idea behind their approach is the identification of a population as a state and the transition probability as the probability of going from one population to another.

Stephens and Waelbroeck (1999) used coarse graining to model schemata evolution. He introduced the notion of effective fitness and suggested that schemata with high effective fitness receive exponentially increasing copies. He also suggested that when schema reconstruction dominates, large schemata are favored and in deceptive problems, short, low order schemata are favored. Furthermore, he derived Geiringer's theorem (Geiringer, 1944) using the coarse-graining approach. Another approach used to model the dynamics of GAs is statistical mechanics modeling proposed by Shapiro and his coworkers (Shapiro, Prügel-Bennet, & Rattray, 1994; Prügel-Bennett & Shapiro, 1994; Rattray, 1996). In statistical-mechanics modeling, the dynamics of cumulants of fitness distribution are quantified. Recombination operators are quantified by both benefits of mixing and BB

disruptions modeled as "interface energy" (Prügel-Bennett, 1997). For further details on statistical-mechanics models the interested should refer to Rattray (1996) and the references therein.

To reiterate, the models discussed in this section are complex, partially due to the fact that the include many facets of GAs, but they yield a good agreement with the dynamics of GAs. However, due the complexity of models, they cannot be easily used in successful GA design.

8 Summary and Conclusions

This study performed an exhaustive survey of models developed to analyze fixed recombination operators. A classification methodology to organize the existing models was proposed based on what aspect of recombination they model. This study suggests that majority of the work has been done in analyzing recombination operators as just a diversity preservation tool. Only few studies have been performed in developing facetwise models to analyze recombination as an innovation operator. Therefore, there is an immediate need for a better facetwise model that addresses the BB mixing problem and analyzes fixed recombination to answer the mixing question. This study also acknowledges the existence of more complex models, however, such models rarely make themselves amenable for designing successful GAs.

References

Altenberg, L. (1995). The schema theorem and Price's theorem. *Foundations of Genetic Algorithms*, *3*, 23–50.

Arora, S., Rabani, Y., & Vazirani, U. V. (1994). Simulating quadratic dynamical systems is PSPACE-complete. *ACM Symposium on Theory of Computing*, 459–467.

Bagley, J. D. (1967). *The behavior of adaptive systems which employ genetic and correlation algorithms.* Doctoral dissertation, University of Michigan, Ann Arbor, MI. (University Microfilms No. 68-7556).

Bailey, N. T. J. (1961). *Introduction to mathematical theory of genetic linkage.* London, UK: Oxford University Press.

Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. *Proceedings of the International Conference on Genetic Algorithms and Their Applications*, 101–111.

Booker, L. B. (1993). Recombination distributions for genetic algorithms. *Foundations of Genetic Algorithms*, *2*, 29–44.

Booker, L. B., Fogel, D. B., Whitley, D., & Angeline, P. J. (1997). Recombination. In Bäck, T., Fogel, D. B., & Michalewicz, Z. (Eds.), *The Handbook of Evolutionary Computation* (Chapter E3.3, pp. C3.3:1–C3.3:27). Philadelphia, PA: IOP Publishing Ltd. and Oxford University Press.

Bridges, C. L., & Goldberg, D. E. (1987). An analysis of reproduction and crossover in binary-coded genetic algorithm. *Proceedings of the Second International Conference on Genetic Algorithms*, 9–13.

Bulmer, M. G. (1985). *The mathematical theory of quantitative genetics.* Oxford: Oxford University Press.

Christiansen, F. B. (1989). The effect of population subdivision on multiple loci without selection. In Feldman, M. W. (Ed.), *Mathematical Evolutionary Theory* (pp. 71–85). Princeton, NJ: Princeton University Press.

De Jong, K. A. (1975). *An analysis of the behavior of a class of genetic adaptive systems.* Doctoral dissertation, University of Michigan, Ann-Arbor, MI. (University Microfilms No. 76-9381).

De Jong, K. A., & Spears, W. M. (1990). An analysis of the interacting roles of population size and crossover in genetic algorithms. *Parallel Problem Solving from Nature*, 38–47. (Also AIC Report No. AIC-90-003).

De Jong, K. A., & Spears, W. M. (1992). A formal analysis of the role of multi-point crossover in genetic algorithms. *Annals of Mathematics and Artificial Intelligence*, *5*(1), 1–26.

Deb, K., & Goldberg, D. E. (1993). Analyzing deception in trap functions. *Foundations of Genetic Algorithms*, *2*, 93–108. (Also IlliGAL Report No. 91009).

Eshelman, L. J. (1991). The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. *Foundations of Genetic Algorithms*, *1*, 265–283.

Eshelman, L. J., Caruana, R. A., & Schaffer, J. D. (1989). Biases in the crossover landscape. *Proceedings of the Third International Conference on Genetic Algorithms*, 10–19.

Eshelman, L. J., & Schaffer, J. D. (1995). Productive recombination and propagating and preserving schemata. *Foundations of Genetic Algorithms*, *3*, 299–313.

Falconer, D. S. (1989). *Introduction to quantitative genetics.* London: Longman Scientific & Technical.

Frantz, D. R. (1972). *Non-linearities in genetic adaptive search.* Doctoral dissertation, University of Michigan, Ann Arbor, MI. (University Microfilms No. 73-11,116).

Geiringer, H. (1944). On the probability theory of linkage in mendelian heridity. *Annals of Mathematical Statistics*, *15*, 25–57.

Goldberg, D. E. (1987). Simple genetic algorithms and the minimal, deceptive problem. In Davis, L. (Ed.), *Genetic algorithms and simulated annealing* (Chapter 6, pp. 74–88). Los Altos, CA: Morgan Kaufmann.

Goldberg, D. E. (1989a). Genetic algorithms and Walsh functions: Part II, deception and its analysis. *Complex Systems*, *3*, 153–171. (Also IllIGAL Report No. 89001).

Goldberg, D. E. (1989b). *Genetic algorithms in search optimization and machine learning.* Reading, MA: Addison-Wesley.

Goldberg, D. E. (1991). *Six steps to GA happiness.* Paper presented at Oregon Graduate Institute, Beaverton, OR.

Goldberg, D. E. (1999). The race, the hurdle, and the sweet spot: Lessons from genetic algorithms for the automation of design innovation and creativity. In Bentley, P. (Ed.), *Evolutionary Design by Computers* (Chapter 4, pp. 105–118). San Mateo, CA: Morgan Kaufmann.

Goldberg, D. E. (in press). *Design of innovation: Lessons from and for competent genetic algorithms.* Boston, MA: Kluwer Acadamic Publishers.

Goldberg, D. E., & Bridges, C. L. (1990). An analysis of a reordering operator on a ga-hard problem. *Biological Cybernetics*, *62*, 397–405. (Also TCGA Report No. 88005).

Goldberg, D. E., & Deb, K. (1991). A comparitive analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms*, *1*, 69–93.

Goldberg, D. E., Deb, K., & Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, *6*, 333–362. (Also IlliGAL Report No. 91010).

Goldberg, D. E., Korb, B., & Deb, K. (1989). Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, *3*(5), 493–530. (Also IlliGAL Report No. 89003).

Goldberg, D. E., & Liepens, G. (1991). Theory tutorial. (Tutorial presented at the 1991 International Conference on Genetic Algorithms, La Jolla, CA).

Goldberg, D. E., & Sastry, K. (2001). A practical schema theorem for genetic algorithm design and tuning. *Proceedings of the Genetic and Evolutionary Computation Conference*, 328–335. (Also IlliGAL Report No. 2001017).

Goldberg, D. E., Sastry, K., & Latoza, T. (2001). On the supply of building blocks. *Proceedings of the Genetic and Evolutionary Computation Conference*, 336–342. (Also IlliGAL Report No. 2001015).

Goldberg, D. E., & Segrest, P. (1987). Finite Markov chain analysis of genetic algorithms. *Proceedings of the Second International Conference on Genetic Algorithms*, 1–8.

Goldberg, D. E., Thierens, D., & Deb, K. (1993). Toward a better understanding of mixing in genetic algorithms. *Journal of the Society of Instrument and Control Engineers*, *32*(1), 10–16. (Also IlliGAL Report No. 92009).

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics*, *8*, 299–309.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

McPeek, M. S. (1996). Introduction to recombination and linkage analysis. In Speed, T., & Waterman, M. S. (Eds.), *Genetic Mapping and DNA Sequencing*, Volume 81 of *IMA Volumes in Mathematics and its Applications* (pp. 1–15). Springer-Verlag.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. continous parameter optimization. *Evolutionary Computation*, *1*(1), 25–49.

Nix, A. E., & Vose, M. D. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, *5*, 79–88.

Price, G. R. (1970). Selection and covariance. *Nature*, *227*, 520–521.

Prügel-Bennett, A. (1997). The dynamics of a genetic algorithm for simple random Ising systems. *Physica D*, *104*, 75–114.

Prügel-Bennett, A. (2001). The mixing rate of different crossover operators. *Foundations of Genetic Algorithms*, *6*, 261–274.

Prügel-Bennett, A., & Shapiro, J. L. (1994). An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, *72*(9), 1305–1309.

Pudlák, P. (2001). Complexity theory and genetics: The computational power of crossing over. *Information and Computation*, *171*, 201–223.

Rabani, Y., Rabinovich, Y., & Sinclair, A. (1998). A computational view of population genetics. *Random Structures & Algorithms*, *12*(4), 313–334.

Rana, S. (1999). The distributional biases of crossover operators. *Proceedings of the Genetic and Evolutionary Computation Conference*, 549–556.

Rattray, L. M. (1996). *Modelling the dynamics of genetic algorithms using statistical mechanics*. Doctoral dissertation, University of Manchester, Manchester, UK.

Robbins, R. (1918). Some applications of mathematics to breeding problems, III. *Genetics*, *3*, 375–389.

Shapiro, J. L., Prügel-Bennet, A., & Rattray, M. (1994). A statistical mechanical formulation of the dynamics of genetic algorithms. In *Lecture Notes in Computer Science*, Volume 865 (pp. 17–27). Berlin: Springer-Verlag.

Spears, W. (1997). Recombination parameters. In Bäck, T., Fogel, D. B., & Michalewicz, Z. (Eds.), *The Handbook of Evolutionary Computation* (Chapter E1.3, pp. E1.3:1–E1.3:13). Philadelphia, PA: IOP Publishing Ltd. and Oxford University Press.

Spears, W. (2001). The equilibrium and transient behavior of mutation and recombination. *Foundations of Genetic Algorithms*, *6*, 241–260.

Spears, W. M., & De Jong, K. A. (1991). An analysis of multi-point crossover. *Foundations of Genetic Algorithms*, *1*, 301–315. (Also AIC Report No. AIC-90-014).

Stephens, C., & Waelbroeck, H. (1999). Schemata evolution and building blocks. *Evolutionary computation*, *7*(2), 109–124.

Syswerda, G. (1989). Uniform crossover in genetic algorithms. *Proceedings of the Third International Conference on Genetic Algorithms*, 2–9.

Thierens, D. (1995). *Analysis and design of genetic algorithms.* Doctoral dissertation, Katholieke Universiteit Leuven, Leuven, Belgium.

Thierens, D. (1996). Dimensional analysis of allele-wise mixing revisited. *Parallel Problem Solving from Nature*, 255–265.

Thierens, D. (1999). Scalability problems of simple genetic algorithms. *Evolutionary Computation*, *7*(4), 331–352.

Thierens, D., & Goldberg, D. E. (1993). Mixing in genetic algorithms. *Proceedings of the Fifth International Conference On Genetic Algorithms*, 38–45.

Vose, M. D., & Liepens, G. E. (1991). Punctuated equilibria in genetic search. *Complex Systems*, *5*(1), 31–44.