九州大学大学院理学系研究科

理学博士学位申請論文 (主論文)

Evolution of a Novel Function Fascilitated by Genetic Recombination in Genetic Algorithms

(遺伝的アルゴリズムにおいて組み換えが新しい機能の進化を加速する効果について)

by

鈴木秀明 (Hideaki Suzuki)

Dissertation

submitted in partial fulfillment of the requirements

for the degree of Doctor of Science in Biology

Division of Science, Graduate School of Kyushu University

Contents

Preface

Recently in the engineering field, genetic algorithms (GAs) have attracted a great deal of attention as random search methods for optimization. In GAs, each design of an object is typically coded in a gene-like bit sequence, and in imitation of biological evolution, the optimal (or a close to optimal) design is searched for by operating evolutionary operations (selection, mutation, and genetic recombination) on a population of those sequences. Among these operations, the most characteristic operation of GAs is the genetic recombination, or crossover operation. Since GAs without crossover are nothing more than a parallel hill-climbing method, crossover is a key operation to achieve the optimal design in the shortest number of trials. In most studies using GAs, however, the effectiveness of GAs is questionable. There is a lot of literature on applying GAs to industrial problems; and yet those papers report that the performance of GAs varies diversely depending upon the genetic parameters such as the mutation rate, crossover rate, selection scheme, and fitness landscape on the searching space represented by bit sequences.

Among these various schemes of GAs, I first focus on the fitness landscape. I propose a conspicuously peaked landscape, which stemmed from a study of machine language genetic programming system. When GAs are applied to optimization of a long bit sequence coding a set of machine instructions, GAs must search for an appropriate set of bits which composes some advantageous set of instructions. Mutations of one or few bits have no influence on the final function of a program, and a program can enjoy highly functional advantage only when all the component bits are present in the same program (individual). This fitness model makes a population of programs evolve with an intermittent process wherein drastic adaptive evolution occasionally punctuates long period of stases (neutral evolution), making the evolutionary speed principally determined by the waiting time until creation of an advantageous set of bits. To study this time, I here devise a fitness function for only one advantageous function (which I call a Babel-like fitness landscape in Chapter 2) and study the rate of evolution, especially focusing on the acceleration rate by crossover.

Estimates are made using the following three different methods; the theoretical analysis with mathematical formulas, the numerical method using the vector representation of a population, and the direct simulation method with GAs operated on a population of bit sequences. In Chapter 1, I first develop the second numerical method and simulate a

generation cycle with recurrence formulas. It is shown from the result that at an intermediate mutation rate crossover can greatly reduce the waiting time until an advantageous set of bits dominates the population. A brief mathematical analysis is also given in this chapter to explain the results. Chapter 2 is devoted to the analysis examining how much crossover reduces that waiting time and accelerates evolution. I develop a more detailed analytical methods and estimate the acceleration rate by crossover under different values of genetic parameters. In order to make sure of the theoretical result, experiments using the other two numerical methods are also presented in this chapter. From these results it is concluded that crossover greatly enhances the rate of evolution when genetic parameters are adjusted appropriately.

In the following, I explain the contents of these two chapters in more detail:

Chapter 1. The Optimum Recombination Rate that Realizes the Fastest Evolution of a Novel Functional Combination of Many Genes

The effect of genetic recombination (or crossover) by sexual reproduction is studied on the time until a novel set of genes performing a combined function appears, spreads, and becomes fixed. First, we study a haploid finite population with many binary loci, in which only one sequence (called a functional gene set) is significantly advantageous over the others. The time for evolution of the function (T_{d}) is defined as the mean number of generations until the advantageous sequence dominates in an initially random population. When the sequence diversity is initially stored sufficiently, the evolution time T_{d} is roughly the product of waiting time until the appearance of the advantageous sequence (creation time T_{c}) and the average number of appearances of the advantageous sequence from its absence until its fixation (destruction number N_{d}). Mutation and crossover reduce the former but enlarge the latter. If the mutation rate is low, there is an intermediate optimal rate of crossover that achieves the minimum T_{d}. In contrast, if the mutation rate is sufficiently high, T_{d} is smallest without crossover. Second, the break-down of established functions by recurrent deleterious mutation is examined in an infinite population. The number of functional genes maintained monotonically decreases with the recurrent deleterious mutation rate. Thus in higher organisms having many functional sets of genes in the genome, the mutation rate must be kept very low to preserve them, and hence a high crossover rate made possible by sexual reproduction is important in accelerating the

evolution of novel functional sets of genes. Implication of this long-term advantage of recombination in the maintenance of sexual reproduction in higher organisms is discussed.

Chapter 2. Crossover Accelerates Evolution in GAs with a Babel-like Fitness Landscape: Mathematical Analyses

The effectiveness of crossover in accelerating evolution in genetic algorithms (GAs) is studied with a haploid finite population of bit sequences. A Babel-like fitness landscape is assumed. There is a single bit sequence (schema) that is significantly more advantageous than all the others. We study the time until domination of the advantageous schema (T_{d}). Evolution proceeds with appearance, spread, and domination of the advantageous schema. The most important process determining T_{d} is the appearance (creation) of the advantageous schema, and crossover helps this creation process and enhance the rate of evolution. To study this effect, we first establish an analytical method to estimate T_{d} with or without crossover. Then, we conduct a numerical analysis using the frequency vector representation of the population with the recurrence relations formulated after GA operations. Finally we carry out direct computer simulations with simple GAs operating on a population of binary strings directly prepared in the computer memory to examine the performance of the two analytical methods. It is shown that T_{d} is reduced greatly by crossover with a mildly high rate when the mutation rate is adjusted to a moderate value and that an advantageous schema has a fairly large *order* (the number of bits). From these observations, we can determine implementation criteria for GAs, which are useful when we apply GAs to engineering problems having a conspicuously discontinuous fitness landscape.

CHAPTER 1

The Optimum Recombination Rate that Realizes the Fastest Evolution of a Novel Functional Combination of Many Genes

The paper was published in *Theor. Pop. Biol.* (**51,** 185-200) in 1997.

Introduction

Questions on the evolution of sex and recombination have been among the major unsolved issues in evolutionary biology over the decades. Various hypotheses have been proposed to explain why sexual reproduction is maintained in most higher organisms in spite of a high cost of maintaining it (Williams, 1975; Maynard-Smith, 1978; Lloyd, 1980; Michod and Levin, 1988). Recombination by sexual reproduction accelerates evolution by making advantageous combinations of genes (Crow and Kimura, 1965; Maynard-Smith, 1971; Felsenstein, 1974; Takahata, 1982), or by excluding deleterious genes more effectively (Muller, 1964; Haigh, 1978; Kondrashov, 1988; Redfield, 1994; Kondrashov, 1994). Genetic diversity arising from recombination is beneficial in the continuously changing environment (Maynard-Smith, 1971; Sasaki and Iwasa, 1987) or when subject to attack by pathogens or parasites (Hamilton, 1980; Hamilton *et al.*, 1990).

Above all, the possibility that genetic recombination makes advantageous combinations of genes and accelerates evolution, which was originally suggested by Fisher (1930) and Muller (1932), has been examined extensively since the first quantitative calculation by Crow and Kimura (1965). Various models have been studied, and a number of statements have been made in different ways to describe this long term advantage of recombination (Crow and Kimura, 1965, 1969; Bodmer, 1970; Eshel and Feldman, 1970; Maynard-Smith, 1971; Karlin 1973; Felsenstein, 1974; Takahata, 1982). The evolutionary acceleration mechanism asserted in these papers are summarized as follows; recombination speeds the response to selection by disrupting inter-locus linkage disequilibrium that is continually produced by random genetic drift. This inter-locus linkage disequilibrium was first comprehensively studied by Hill and Robertson (1966) and called “Hill-Robertson effect” by Felsenstein (1974). The latter author also pointed out that the authors using the finite population models had found the advantage of recombination whereas those using the infinite population models had found none.

The studies about advantage or disadvantage of sexual recombination from the individual selection point of view was begun by Nei (1967, 1969). He introduced the recombination-modifying locus (i.e., the special locus that controls the recombination rate between the major selected loci) and studied what kind of allele can increase in this modifying locus. This work had been generalized by Feldman and others using two

major selected loci (Feldman 1972; Feldman *et al.*, 1980; Feldman and Liberman, 1986; Liberman and Feldman, 1986), and was established as “the Reduction Principle” asserting that recombination is reduced on an appropriate initial condition. From the work by Felsenstein and Yokoyama (1976), this approach was extended to the *many* (more than two) major locus model. They assumed a multiplicative selection scheme between twenty major loci and found an increase of recombination under an appropriate condition, which extended the Fisher-Muller theory to the individual selection paradigm. Since this work, a number of authors have studied many locus models with the recombination modifier under different fitness schemes, recombination patterns, and so on, and they have set forth various conditions for advantage or disadvantage of genetic recombination (Maynard-Smith, 1980, 1988; Bergman and Feldman, 1990, 1992; Zhivotovsky *et al.*, 1994). In these studied, however, there are as many models that lead to reduction of recombination as its increase.

Recently in the engineering field, genetic algorithms (GAs) have attracted a great deal of attention as random search methods for optimization (Holland, 1992; Goldberg, 1989; Mitchell *et al.*, 1991; Forrest and Mitchell, 1993; Forrest, 1993; Otto *et al.*, 1994; Vose and Wright, 1995). In GAs, each design of an object is typically coded in a gene-like bit sequence and a population of those sequences is prepared. The optimal (or a close to optimal) design is searched for by evolutionary operations including reproduction and competition between sequences over many generations, which imitates biological evolution, such as natural selection, mutation, genetic recombination, and random drift due to finite population size.

Among these operations, the most characteristic operation of GAs is the genetic recombination, or crossover operation. Since GAs without crossover are nothing more than a parallel hill-climbing method, crossover is a key operation to achieve the optimal design in the shortest number of trials. From this intuitive reasoning, it immediately follows that the theoretical study about the role of crossover in GAs should be closely related to the biological theory asserting that recombination accelerates evolution. Otto *et al.* (1994) studied the advantage (or disadvantage) of sexual recombination from this point of view. They adopted the specific fitness function in the twenty locus model and examined the waiting time until the creation and the domination of the most

advantageous sequence of genes. In spite of such a work, however, there are no general theories on the “optimal” rates of crossover and mutation established in GAs. A general theoretical study on the role of crossover in the rate of adaptive evolution is of large practical importance.

The primary aim of this chapter is to present a biological theory explaining the maintenance of sex from the viewpoint of the long range advantage of recombination. The inspiration was given from the study of GAs. We consider the case in which the novel function is advantageous only if *a large number of* genes are combined. A novel function achieved by a combination of a large number of genes enjoys a very high selective advantage only when all component genes are present in the same individual, and most single mutations are supposed to be *neutral* if they occur separately. This very epistatic fitness scheme was devised by analogy to the fitness landscape of MUNCs, the evolutionary programming system, proposed by Suzuki (1996). In this system, a novel advantageous function is achieved by a subroutine composed of a consecutive bit sequence in the memory. Driven by GAs, MUNCs evolve functional subroutines one by one, and eventually establish a very advantageous function in program memories. Fig. 1.1 symbolically illustrates the evolutionary picture with which MUNCs proceed. In this figure, each individual is a haploid genome that is a sequence of binary (0-1) alleles and is represented in a row of binary matrix. A part of this bit sequence may correspond to a set of genes with a combined function, and through the processes shown in Fig. 1.1(b), those functional sets of genes appear and dominate the population. As a consequence, evolution proceeds as long neutral evolutionary phases (Kimura, 1983) and intermittent short adaptive evolutionary phases, making a resultant stairs-like growth curve of the mean population fitness. See Suzuki (1996) for more detailed explanation of this picture.

Based upon this evolutionary picture, we conduct two computer simulations to examine the two key parameters; the interval time between steps of the fitness growth curve which determines the evolutionary speed and maximum number of steps which determines how many functions can be stored in the genome. In the following, to examine the effect of genetic recombination on the evolutionary speed, I first introduce a many-locus finite-population model and study the time until a novel function comes to predominate. Then some mathematical analysis is given to explain the simulation result and to estimate the maximum crossover rate for the fastest evolution to occur. Although

this first simulation is executed using the vector representation of a population, evolutionary procedures operated on that vector are models of GA operations, so that the derived result showing the crossover's acceleration effect of evolution is immediately true of GAs used for a population of bit sequences. After that, I study the number of functional gene sets maintained against deleterious mutation pressure in an infinite population. Finally, various implications of the model are discussed, including the effectiveness and limitation of GAs.

Evolution of a Novel Function and Optimal Recombination

Here I examine a model for the evolution of a single novel function and study the optimal crossover rate that achieves the fastest evolution. The evolutionary speed is measured by the average number of generations until the advantageous combination of genes dominates the population which is called domination time, T_d. (Here the term 'domination' does not mean that one chromosome reaches 100%. The precise definition of domination will be given later). Note that in the finite population, first several advantageous haplotypes might disappear by chance and the final domination of the advantageous haplotype is generally occasioned not by its first appearance but by the later one. Accordingly, as a very rough estimation, T_d is the product of the average number of generations until the occurrence of a haplotype (combination of genes) having an advantageous combination of genes (creation time, T_c) and the average number of occurrences of the advantageous haplotype from its absence until its fixation in the population (destruction number, N_d);

$$T_d = T_c \cdot N_d. \tag{1.1}$$

Although both T_c and N_d are average values of random variables and T_d (the average value of the product of those random variables) cannot be precisely formulated by the product of the average values, we here considered creation and destruction to be independent processes and neglected the correlation term between them. Moreover, we neglected the time needed for accumulation of the population diversity on the assumption that the initial distribution is completely random, and also neglected the time needed for the advantageous haplotype to spread through the population because its relative advantage over the others is assumed to be significantly large. Since crossover is the randomization process in the nonfunctional region shown in Fig. 1.1, T_c decreases and N_d increases with the crossover rate. This makes T_d minimum at an intermediate crossover rate, which is estimated in the finite population.

Basic assumptions of the model are as follows:

- Individuals are haploid and its genome is represented by a string of a large number of binary (0-1) bits.
- Among all sets of genes, a single sequence, denoted by $[11\cdots1]$, has fitness by far larger than the others, and all the other sequences are the same in fitness.

¶Mating pair is randomly chosen from the sufficiently large population so that the correlation between the bit distributions of the mating pair is negligible.

Owing to this second assumption, creation time T_c is a very important random variable, which eventually determines evolutionary performance.

Let

N = population size;

i = number of loci in which the state is “1”;

x_i = frequency of the population of sets of genes with i of bit 1’s;

I = total number of bits in the set of genes, or
number of genes necessary to realize a novel advantageous function
(we hereafter call this ‘functional order’);

s = selection coefficient of the advantageous set of genes ($i = I$) relative to the others ($i < I$) (s can be large);

u = mutation rate (probability of bit flipping) per locus (bit position) per generation;

c = probability of the crossover point (chiasma) to occur between neighboring binary loci;

r = ratio of the fraction of population which participate in the genetic recombination.

Two parameters, c and r, specify crossover rate. When $r < 1$, the whole population is divided into two parts, the recombining subpopulation and the non-recombining subpopulation. At this time the random mating assumption, which does not literally hold true, is considered to be approximately valid in that participants of the genetic recombination is randomly chosen from the population at each generation. The state of a population is described by frequency vector $(x_0, x_1, \cdots, x_I)$. The next generation vector (x_i) is calculated by the recursion formulas for selection, mutation, crossover, and random drift operation as follows:

$$x_i \overset{\text{sel.}}{\rightarrow} x_i((1 + s\delta_{Ii})/(1 + sx_I)) , \tag{1.2}$$

$$x_i \overset{\text{mut.}}{\rightarrow} \sum_{j=0}^{I} x_j M_{ji}, \tag{1.3}$$

$$x_i \overset{\text{cross.}}{\rightarrow} (1-r)x_i + r\sum_{j=0}^{I}\sum_{k=0}^{I} x_j x_k C_{jki} \ , \tag{1.4}$$

$$x_i \overset{\text{drift}}{\rightarrow} \frac{1}{N}\cdot(\text{number of choices of } i \text{ th state in } N \text{ times roulettes}), \tag{1.5}$$

where M_{ji} is the mutation rate from j to i, and C_{jki} is the transition probability from j to i after recombining with k. (Detailed expressions of M_{ji} and C_{jki} are given in Appendix A and Appendix B respectively.) The probability for i th state to be chosen at roulette in Eq. (1.5) is determined to be proportional to x_i. The initial state is assumed to be a binomial distribution $x_i = \binom{I}{i}\cdot(1/2)^I$, indicating that an allele at each locus is chosen randomly (fifty-fifty) from binary bits in an infinite population. Starting from this initial state, vector (x_i) is calculated recursively with above four formulas until the frequency of the advantageous set of genes x_I exceeds 0.5. (The reason for the choice of the threshold value of 0.5 will be discussed later.) Though x_I takes small positive values after operation with Eqs. (1.3) and (1.4), those values are by far smaller than $1/N$ so that x_I becomes zero after operation with Eq. (1.5) in almost all generations. The term "occurrence" of the advantageous set of genes which was used at defining T_{c} and N_{d} means that x_I jumps up to $1/N$ or more, after operation with Eq. (1.5). Computer simulations are run on SPARC station 10 model 51 (50MHz). For each set of parameters, numerical trials are executed ten times and the mean value of T_{d} is calculated.

Figs. 1.2 and 1.4 show the results about c-dependence at $r = 1$ (all individuals participate in crossover) and Figs. 1.3 and 1.5 show the results about r-dependence at $c = 0.5$ (the occurrence rate of crossover points is high enough to make each locus independent of the others). The population size N is taken to be 10000 through all results. These four figures show that when mutation rate u is low, domination time T_{d} is the shortest at certain intermediate values of c and r, and larger or smaller than this optimal value makes T_{d} larger. This implies that evolution is the fastest for an intermediate recombination rate. When u is high, on the other hand, T_{d} is small enough even at $c = 0$ or $r = 0$ (Figs. 1.2 and 1.3). Therefore optimum value of c or r is minimum if the mutation rate is sufficiently high, implying that the evolution of a novel function is fastest if there is no recombination.

Figs. 1.4 and 1.5 show the results about I-dependence. Note that in both figures domination time T_{d} grows geometrically with functional order I. (It took about one

month on SPARC station 10 to make Figs. 1.4(a) and 1.5(a) for $I = 24$.) The upper limit of c above which T_{d} becomes extremely large decrease with I (Fig 1.4), but the upper limit of r is independent of I (Fig 1.5).

Analysis lumping neutral sets of genes:

The above simulation result can be more clearly understood by the following argument in which neutral sets of genes are lumped together according to the number of 1's. Let x_I be the frequency of the advantageous set of genes. The recursion relations for x_I under selection, mutation, and crossover are:

$$x_I \overset{\mathrm{sel.}}{\rightarrow} ((1+s)x_I)/(1+sx_I) \ , \tag{1.6}$$

$$x_I \overset{\mathrm{mut.}}{\rightarrow} x_I - Iu\left(x_I - \left(\frac{1}{2}\right)^I\right) \ , \tag{1.7}$$

$$x_I \overset{\mathrm{cross.}}{\rightarrow} x_I - r(1-P_{\mathrm{het}}(q))\frac{1-x_I}{1-q^I}(x_I - q^I) \ , \tag{1.8}$$

where q is the frequency of allele 1's at each locus and $P_{\mathrm{het}}(q)$ is the probability that an advantageous set of genes is present after crossover between an advantageous set of genes and a neutral set of genes. Eqs. (1.7) or (1.8) hold true around the stationary distribution for mutation or crossover which we assumed in the derivation of those equations respectively (see Appendix C).

$P_{\mathrm{het}}(q)$ is a function sensitive to q, and q increases abruptly during the spreading process of the advantageous set of genes. We here, however, are interested in the state of the population before or immediately after occurrences of advantageous sets of genes so that we give the expression around the random population denoted by $q = 1/2$;

$$P_{\mathrm{het}}\left(\frac{1}{2}\right) = \left(1 + \frac{3c}{\sqrt{1-2c+9c^2}}\right)\left\{\frac{3}{4}(1-c) + \frac{1}{4}\sqrt{1-2c+9c^2}\right\}^I \ . \tag{1.9}$$

For the derivation of this equation see Appendix D, where we assumed that I is large and that 0-1 distribution is independently determined by $q = 1/2$ at each locus. From this equation, we can obtain two approximate equations;

$$P_{\mathrm{het}}\left(\frac{1}{2}\right) = (1-c)^{I-3} \qquad \text{if } c \text{ is very small } (c \ll 1), \tag{1.10}$$

and

$$P_{\mathrm{het}}\left(\frac{1}{2}\right) = 2\left(\frac{3}{4}\right)^{I} \qquad \text{if } c \text{ is maximum } (c = \frac{1}{2}). \tag{1.11}$$

See Appendix D for the derivation of Eq. (1.10).

According to Eqs. (1.6)~(1.8), selection increases x_I, but mutation and crossover increase or decrease x_I depending upon whether x_I is smaller or larger than threshold values. Threshold values are 0.5^I in Eq. (1.7) and q^I in Eq. (1.8). An increase of x_I below these threshold values decreases creation time T_{c} and makes T_{d} smaller in lower mutation or crossover rates, and a decrease of x_I above these values increases destruction number N_{d} and makes T_{d} larger in higher mutation or crossover rates. I consider these two situations subsequently.

(i) ***Creation time***: From Eqs. (1.7) and (1.8), we can see that around the random distribution mutation increases x_I from zero to $(1/2)^I Iu$, and crossover increases x_I from zero to $(1/2)^I r(1 - P_{\mathrm{het}}(1/2))$. Here $q = 1/2$ was substituted because on the present assumption, bit 0 and bit 1 are evenly distributed in the absence of the advantageous set of genes. The sum of these terms is

$$(1/2)^I\{Iu + r(1 - P_{\mathrm{het}}(1/2))\}\ , \tag{1.12}$$

which is roughly the value of x_I after the transition by Eqs. (1.3) and (1.4). Although this term is small, it is very important as it causes the creation of novel advantageous sets of genes after operation with Eq. (1.5). We can get the theoretical formula of the creation time T_{c}, noting that the expected creation time is the reciprocal of the creation probability that is given by Eq. (1.12) divided by $1/N$;

$$T_{\mathrm{c}} = \left\lceil N\left(\frac{1}{2}\right)^{I}\left\{Iu + r\left(1 - P_{\mathrm{het}}\left(\frac{1}{2}\right)\right)\right\}\right\rceil^{-1} . \tag{1.13}$$

Substituting Eqs. (1.10) or (1.11) into Eq. (1.13), we can get T_{c} at $r = 1$ as a function of low c or T_{c} at $c = 1/2$ as a function of r respectively. These dependences are illustrated in Figs. 1.2~1.5 which show that T_{c} agrees qualitatively well with the dependence of T_{d} for lower c or r. Although T_{c} is proportional to 2^I and causes very large values of T_{d} at large I (Figs. 1.4(a) and 1.5(a)), genetic recombination which randomizes nucleotides quickens the appearance of an advantageous set of genes and reduces T_{d} considerably at low u.

(ii) ***Destruction number***: From Eqs. (1.7) and (1.8), the destruction by mutation and

that by crossover are Iux_I and $r(1-P_{\mathrm{het}}(q))((1-x_I)/(1-q^I))x_I$, respectively. By substituting $x_I = 1$ for x_I in these terms, we can see the difference between mutation and crossover as to destruction; when $x_I = 1$, the former is positive but the latter is zero. (Here we extended Eqs. (1.7) and (1.8) which accurately hold true only around stationary distribution of each process.) This implies that mutation destroys the advantageous set of genes even when it is fixed in the population but crossover does not destroy the advantageous set of genes once it has been fixed. To create a new advantageous set of genes without destroying old functional sets of genes is possible by crossover, but not by mutation.

In the present simulation, the relative fitness advantage of the functional gene set is assumed to be so large that advantageous sequences which could escape from the initial danger of extinction can almost always dominate the population. The created advantageous set of genes is most easily eliminated immediately after its occurrence so that the destruction number N_{d} is crucially dependent upon whether or not selection can overcome the destruction by mutation and crossover at $x_I \sim 1/N$ and $q = 1/2$. This state hardly differs from the stationary state of mutation and crossover and we can use Eqs. (1.7) and (1.8) for analysis. Assigning $q = 1/2$ in Eq. (1.8) and multiplying Eqs. (1.6), (1.7), and (1.8) with approximation $(1/2)^I \ll x_I \ll 1$, we obtain the condition for the spread of the advantageous set of genes;

$$(1+s)(1-Iu)\left|1-r\left(1-P_{\mathrm{het}}\left(\frac{1}{2}\right)\right)\right| > 1 \quad . \tag{1.14}$$

Substituting $P_{\mathrm{het}}(1/2)$ in Eq. (1.14) by Eq. (1.9) gives the condition for the spread, which is illustrated in Fig. 1.6 for different values of s and I. For the two special cases of $r = 1$ or $c = 0.5$, we substitute Eqs. (1.10) or (1.11) for $P_{\mathrm{het}}(1/2)$ in Eq. (1.14) and get the following spread conditions;

$$c < 1-\{(1+s)(1-Iu)\}^{-\frac{1}{I-3}} \quad \text{if } c \ll 1 \text{ and } r = 1, \tag{1.15}$$

and

$$r < 1-\{(1+s)(1-Iu)\}^{-1} \qquad \text{if } c = \frac{1}{2}. \tag{1.16}$$

Approximation $(3/4)^I \ll 1$ was used in derivation of Eq. (1.16). These conditions agree with the values of c or r above which T_{d} becomes extremely large in Figs. 1.2~1.5.

Even if advantageous sets of genes are created by mutation or crossover, if its fitness is not so large as to satisfy above inequalities, the destruction number N_{d} becomes very large, which brings about a very large value of T_{d}.

(iii) ***Upper limit of*** x_I: Because mutation destroys the sets of genes that have already spread through the population, the occupation by the advantageous set of genes cannot be complete, namely, x_I cannot reach 1. With approximation $x_I \sim 1$, we can give the upper limit of x_I as follows. When $x_I \sim 1$ and $q \sim 1$, neutral sets of genes have so many bit 1's that the probability that an advantageous set of genes remains intact after crossover with a neutral set of genes is nearly equal to one (namely, $P_{\mathrm{het}}(q) \sim 1$); therefore x_I is kept unchanged through Eq. (1.8). So, multiplying Eqs. (1.6) and (1.7) and requiring that x_I becomes stationary under the approximation $(1/2)^I \ll 1$, we can get

$$x_I = 1 - \left(1 + \frac{1}{s}\right) Iu \,. \tag{1.17}$$

This says that, for example, when $I = 20$, $u = 0.01$, and $s = 1$, x_I cannot exceed 0.6. Even if the advantageous set of genes is created and spread, its frequency is limited by the value of Eq. (1.17). This is the reason why the threshold value of domination was chosen to be 0.5 in this simulation.

Number of Functional Gene Sets Maintained against Mutation

Here I study the number of functional sets of genes maintained against the pressure of recurrent deleterious mutations. Basic assumptions of the model are as follows:

¶All functional sets of genes in the genome are ordered and numbered (indicated by $j = 1, 2, 3, \cdots$). A functional set of genes (e.g. j) is effective only when all earlier functional sets (numbered $1, 2, \cdots, j-1$) exist in the same genome, and is neutral (nonfunctional) otherwise.

¶Fitness grows in geometric progression with the number of functional sets of genes.

¶Crossover is neglected.

¶Population size is infinitely large.

Crossover was neglected because it no longer breaks down the sets of genes which have already been fixed in the population and hence has no direct effect on the number of stable sets of genes. The maximum number of advantageous sets of genes which can exist stably in the genome is estimated by the simulation based upon standard population genetics.

I use the following notations:

y_j = frequency in the population of individuals with functional sets of genes numbered $1, 2, \cdots, j$ but not functional set of genes numbered $j + 1$;

$w_j = (1 + s)^j$ = fitness of individual which has functional sets of genes numbered $1, 2, \cdots, j$ but not $j + 1$;

f_t = frequency threshold determining domination (namely, some gene set is regarded as dominating the population when its frequency exceeds this value);

p_d = probability of destruction of the functional gene set by mutation per gene set per generation;

p_c = probability of creation of the functional gene set by mutation per generation.

Here we consider $p_c \ll p_d$ assuming that in number of sequences functional sets of genes are far fewer than the neutral ones. Population is described by the frequency vector $(y_0, y_1, y_2, \cdots)$.

Generally speaking, frequency y_j may include individuals which have some of the functional gene sets numbered $j + 2, j + 3, \cdots$, but gene set number $j + 1$ is not

functional. Those cases are eliminated quickly because recurrent mutation pressure makes those later gene sets nonfunctional. I therefore assume that individuals of y_j have none of those later gene sets. Then z_j, which is defined as the frequency in the population of individuals with functional set of genes numbered j, is given by

$$z_j = y_j + y_{j+1} + y_{j+2} + \cdots,$$

and satisfies

$$z_1 \geq z_2 \geq \cdots.$$

I calculate J, defined by

$$z_J \geq f_t \geq z_{J+1},$$

in the stationary vector in the simulation. With this definition, J means the maximum number of functional sets of genes which can be occupied by the fraction of population whose ratio exceeds f_t, or in other words the number of functional sets of genes which can dominate the population.

The transformations of (y_j) through selection and mutation are formulated as follows:

$$y_j \overset{\text{sel.}}{\rightarrow} (y_j w_j) \Big/ \left(\sum_{j'} y_{j'} w_{j'} \right), \tag{1.18}$$

$$y_j \overset{\text{mut.}}{\rightarrow} \sum_{j'=0}^{j} (1-p_d)^{j'} p_c^{j-j'} (1-p_c) y_{j'} + (1-p_d)^j p_d \sum_{j'=j+1}^{\infty} y_{j'}. \tag{1.19}$$

The initial vector is $(1, 0, 0, \cdots)$ representing a population of individuals with no functional set of genes.

Fig. 1.7(a) shows the result of simulations under frequency threshold $f_t = 0.9$. As far as $0 < p_c \ll p_d$ is satisfied, J is insensitive to p_c and changes only with p_d and s. As is evident from Fig. 1.7(a), the lower p_d is and the larger s is, the larger J is. However for larger s, especially when $s > 0.3$, J becomes saturated and insensitive to s. These saturated values of J are plotted with marked points in Fig. 1.7(b), which shows J decreases roughly in inverse proportion to p_d.

This result is explained as follows. According to Eq. (1.17), the parameters s, p_d, and f_t must satisfy the following condition for domination;

$$1 - \left(1 + \frac{1}{s}\right) p_d \geq f_t,$$

where $Iu \sim p_d$ was substituted. Though this inequality is derived for the single gene set model, this can be used for the multiple gene set model by substituting p_d with $J \cdot p_d$. Thus the condition for J functional gene sets to be able to predominate is given by

$$J \cdot p_d \leq (1 - f_t)\frac{s}{1+s}, \tag{1.20}$$

or when $s = \infty$,

$$J \cdot Iu \sim J \cdot p_d \leq 1 - f_t. \tag{1.21}$$

J as a function of p_d given by Eq. (1.21) is also illustrated in Fig. 1.7(b) with a solid line, which agrees well with marked points by simulation.

According to Eqs. (1.20) or (1.21), no matter how advantageous the functional gene set may be, J, the maximum number of those gene sets which can dominate the population decreases in inverse proportion to u (the mutation rate). This relation between J and u is the same as what Eigen *et al.* (1981) mathematically derived with an “error catastrophe” argument about genetic information. It is concluded from this simulation that the lower the mutation rate u is, the more functional sets of genes can be maintained. In other words, u must be low in order for many functional sets of genes to be maintained.

Discussion

From the study of **the evolutionary speed, i.e., the mean number of generations until occurrence and domination of the advantageous set of genes, it was concluded that when the mutation rate is very low, fairly frequent crossover can greatly reduce the time until the domination of a functional set of genes, but when the mutation rate is high, crossover does not enhance the rate of evolution. On the other hand, it was also shown from the second simulation that a high mutation rate causes problems. The mutation rate needs to be low to keep many functional sets of genes stable in the genome. Thus it has become evident that in the evolutionary model wherein the unit of fixation is not one gene but many genes with a combined function, crossover can be a process promoting faster evolution.**

We now discuss some implications of these results.

Evolutionary Maintenance of Sexual Reproduction: As is shown by the second simulation, once a species has stored many functional genes in the genome, the mutation rate must be low to preserve them. In higher organisms this requirement is met by a very low error rate in DNA replication, made possible by molecular machinery for proof-reading. Consequently in those species, the mutation rate is kept very low (typically about 10^{-5} per locus per generation), and evolution without crossover is very slow in creating a novel function. In Figs. 1.2~1.5, T_{d} is very large for low u and c (or r). Hence, in order that such a species might create a novel advantageous function (by accumulating a new advantageous set of genes) and evolve, crossover is indispensible. T_{d} in Figs. 1.2~1.5 comes to decrease quickly as c or r increases. The rate of this acceleration is known from the ratio of T_{c} without crossover to T_{c} with crossover. When $r = 1$, this ratio is written using Eq. (1.10) and (1.13) approximately as follows;

$$\frac{T_{\mathrm{c}}\big|_{\mathrm{without\,crossover}}}{T_{\mathrm{c}}\big|_{\mathrm{with\,crossover}}} = \frac{Iu + (1-(1-c)^{I-3})}{Iu} \sim \frac{u+c}{u}. \tag{1.22}$$

This formula can be very large when the mutation rate is kept very low as is typical for higher organisms. Higher organisms which had stored many functional genes in their genome could have evolved quickly only through this acceleration effect by crossover. This gives a group selection explanation for why sexual reproduction has been

maintained in higher organisms but not in viruses or bacteria. When thinking of the unit of genetic fixation as a set of genes which has a combined function, the destruction process by crossover is overcome by the large fitness of advantageous sets of genes, and crossover can be a beneficial process for fast evolution.

Comparison with Previous Models: The present argument explaining the evolutionary advantage of sexual reproduction is due to the capability of genetic recombination to shorten the waiting time until the occurrence of an individual having an advantageous combination of genes (i.e., the creation time of a 'hopeful monster'). This creation process is enhanced by disruption of the random linkage disequilibrium referred to as the Hill-Robertson effect, so that the theory presented in this chapter might safely be called the Fisher-Muller theory that has been extended to the many-locus model. In this subsection I note the fitness landscape assumed in the many-locus models and argue the difference between the previous models and the present one.

Because the conventional Fisher-Muller theory is essentially based upon the recombination's ability to disrupt linkage between advantageous mutant alleles to create a double mutant chromosome, many authors who had examined the Fisher-Muller effect in the many-locus models had adopted the *multiplicative* fitness scheme, i.e., the scheme in which fitness of a whole chromosome was given by $w = (1 + s)^i$, where s is the selection coefficient of a single advantageous mutant allele and i is the number of advantageous alleles included in the chromosome (Crow and Kimura, 1965; Maynard-Smith, 1971; Felsenstein, 1974; Felsenstein and Yokoyama, 1976). In the individual selection paradigm, many other different schemes have been studied. Maynard-Smith (1980, 1988) introduced the *gaussian* fitness scheme between major loci, and found that directional selection favors recombination whereas stabilizing selection or frequent changes in the direction of selection diminishes recombination. Bergman and Feldman (1990) examined the similar model not only in gaussian but also in *sigmoidal* or *exponential* fitness scheme. Bergman and Feldman (1992) took a more general fitness scheme which is constituted by sine curves of different frequencies. Zhivotovsky *et al.* (1994) calculated using some weak additive-by-additive epistatic selection. Otto *et al.* (1994), whose study is not an individual selection approach, assumed a specific continuous fitness function on twenty major loci and examined the condition under which recombination is an advantageous process.

All these fitness functions are a continuous or quasi-continuous function of the number of advantageous mutant alleles in the chromosome, i. This is because these previous arguments are essentially based upon the notion that evolution proceeded with substitutions by single advantageous (or deleterious) mutant alleles or that a single mutant allele causes a phenotype change and more or less changes the fitness value of the whole chromosome. In the basic evolutionary picture of the present model, however, adaptive evolution is considered to proceed with occurrences and fixations of not single mutant genes but the advantageous combination of many genes. Since the single mutant alleles are assumed to be neutral in fitness, the fitness function of a whole chromosome is a strikingly discontinuous function of i, being flat except its maximum value at $i = I$. The present argument, that asserts the recombination's ability to disrupt linkage to create an advantageous combination of *many* genes, is different from the conventional Fisher-Muller theory that is essentially based upon the recombination's ability to create an advantageous combination of *two* genes.

The present model considers evolution to be a process going forward with accumulation of blocks, i.e., the advantageous (coadapted) sets of genes. For evolutionary genetics, this begs the question of most difficulty, namely, what selective regimes favor formation of such blocks in the first place. It is true the present argument cannot answer all the important problems of evolution, and yet we can think of many examples in living organisms which might correspond to such a block; a set of enzymes constituting a metabolic cycle, a set of structural proteins composing a novel advantageous organ, a set of genes causing some advantageous behavior, morphology, mimicry, and so on. A single component gene constituting such a coadapted gene set cannot be advantageous by itself, but can execute advantageous function when all component genes are present in the same individual.

Evolutionary Discontinuity: According to Figs. 1.2~1.5, recombination shortens or lengthens the time until the domination of an advantageous set of genes depending upon whether the spread condition holds or not. When $r = 1$, this condition is given by Eq. (1.15), which can be transformed approximately as follows;

$$c < 1 - \{(1+s)(1-Iu)\}^{-\frac{1}{I-3}} \sim \frac{s - Iu}{I-3} \sim \frac{s}{I} \quad . \tag{1.23}$$

This equation states that recombination at a constant rate of chiasma probability allows

the spread of a strongly advantageous set of genes only, whose ratio of fitness as compared to the number of component genes (s/I) is large. In contrast recombination eliminates mildly advantageous set of genes whose s/I is small. Therefore, once recombination is programmed at a constant rate within the genetic structure of organisms, it functions to select only the gene combinations performing strongly advantageous function. When considering the evolution of the function achieved by a large number of genes, recombination makes evolution a discontinuous process.

According to the paleontologic studies examining fossils, the morphology of species seems to have changed discontinuously at intervals of millions of years, and between those drastic changes, morphology seems to stay unchanged (Williamson, 1981; Malmgren *et al.*, 1983). The drastic adaptive evolution coming from the creation of very advantageous combinations of genes might make a new advantageous organ like an eye and bring about the dramatic morphological change of species. This might explain the observed evolutionary discontinuity.

Genetic Algorithms: As described before, the basic evolutionary model assumed in this chapter is given from the study of GAs driving MUNCs (Suzuki, 1996). In this subsection, I discuss the performance and the implementation of GAs in the light of the derived results. The basic assumption of GAs, the "building block hypothesis" (Holland, 1992), says that an optimum solution of the problem is created through combinations of good subsolutions by the aid of crossover. Such good subsolutions (disconnected bit sequences) are called "schemas" and are considered to spread with "implicit parallelism" (Goldberg, 1989). In the biological context, on the other hand, the many-locus Fisher-Muller theory assuming the multiplicative fitness scheme is based upon the notion that recombination combines advantageous mutations which arose simultaneously and separately in different individuals into a single individual. Both theories insist on the parallel spread of good (advantageous) sub-solutions, hence make the same point about evolutionary acceleration.

This implicit parallelism was first doubted by Mitchell *et al.* (1991). They assumed the discontinuous fitness function similar to the present one (they called it 'the royal road function'), studied the GA performance, and found that hitchhiking (they called this 'premature convergence'; see for biological citations Maynard-Smith and Haigh, 1974; Kaplan et al., 1989) drops the GA performance. Although this result made them cast

doubt on the implicit parallelism, they had a notion that crossover is useful in combining good schemas so that they only studied the GA performance taking notice of the time necessary for lower-order schemas to combine to form the higher-order schema and did not examined the crossover's influence on the creation time of those schemas.

If the evolutionary acceleration by crossover is due to the mechanism shown in this chapter, however, crossover is useful not because it can spread good schemas parallel at the same time and combine them, but because it can randomize and create a new good schema while preserving old good schemas that are completely held in common within a population. This randomization process is nothing more than that by mutation, hence when only one good schema is searched for in one sequence of bit string, mutation suffices. Although building block hypothesis is right if a schema exactly corresponds to a set of genes with a novel function treated here, the present evolutionary picture says that such component schemas occurs and dominates the population not in parallel but serially (one by one) by the aid of crossover which can selectively randomize the sequences.

These roles and limitations of crossover might be the key point in our understanding of GAs. Though in current researches GAs are applied to various optimization problems with various fitness functions, when we focus the application of GAs on optimization of bit sequences with a discontinuous fitness function as was treated here, we can implement genetic parameters of GAs according to the following prescriptions:

- ¶When the mutation rate u is sufficiently high, crossover cannot accelerate optimization.
- ¶When the mutation rate u is low, the crossover rate (c and r) has an intermediate optimum value which is bounded by Eq. (1.14) or Fig. 1.6.
- ¶When crossover is operated with the "few-points major-participants" (FPAP) ($c \ll 1$ and $r = 1$) mode, fitness coefficient of the optimum solution normalized by the functional order (s/I) needs to be sufficiently larger than the others so that r.h.s. of Eq. (1.23) might be large.
- ¶When crossover is operated with the "many-points minor-participants" (MPIP) ($c = 0.5$ and $r < 1$) mode, fitness of the optimum solution needs to be sufficiently larger than the others so that r.h.s. of Eq. (1.16) might be large.

As was described above, the optimal rate of crossover is simpler for r than for c

because r can be determined regardless of the functional order I. Therefore for such an engineering tool as GAs, the MPIP mode of crossover is easier to use though it is less likely for real organisms. The above criteria make GAs a profitable optimization strategy of a bit string with a discontinuous fitness function as was described in this chapter.

Finally I finish this discussion by pointing out the negative aspect of GAs. GAs are never a dreamy method which can optimize a bit string with large functional order I. According to Figs. 1.4 and 1.5, domination time T_{d} grows geometrically with an increase of functional order I. This comes from the divergence of creation time T_{c} which is proportional to the combinational number 2^I as was shown in Eq. (1.13). It is true crossover considerably quickens the appearance of the advantageous set of genes by the rate given in Eq. (1.22), but the domination time minimized by crossover still grows geometrically with I. GAs, a successful strategy as they are, cannot avoid the problem of combinational explosion occurring in the optimization of a long bit sequence.

Appendix A: CALCULATION OF M_{ji} IN EQ. (1.3)

M_{ji} is the transition probability from a gene set with j of bit 1's to a gene set with i of bit 1's by mutation. When we focus on the situation in which j of bit 1's and i of bit 1's have k pairs of same positional bit 1's, the probability of such a transition's occurring is given by

(the probability that k of bit 1's are preserved and $j-k$ of bit 1's are flipped by mutation)fl(the probability that $I-j-i+k$ of bit 1's are preserved and $i-k$ of bit 1's are flipped by mutation)

$$= \binom{j}{k}(1-u)^k u^{j-k} \times \binom{I-j}{i-k}(1-u)^{I-j-i+k} u^{i-k}$$

$$= \binom{j}{k}\binom{I-j}{i-k}(1-u)^{I-i-j+2k} u^{i+j-2k} .$$

M_{ji} is given by the summation of this term about k.

$$M_{ji} = \sum_{k=\max(0,\, i+j-I)}^{\min(i,\, j)} \binom{j}{k}\binom{I-j}{i-k}(1-u)^{I-i-j+2k} u^{i+j-2k} . \tag{A1}$$

The range of the summation was determined so that $i-k \geq 0$, $j-k \geq 0$, and $I-j-i+k \geq 0$ are satisfied.

Though accurately M_{ji} is given by Eq. (A1), we used simpler formula described below in the simulation. If $u \ll 1$, matrix $M = (M_{ji})$ is approximately given by $M^{(0)}$ defined by

$$M^{(0)}{}_{ji}(u) = \delta_{j+1,\, i} \cdot u(I-j) + \delta_{ji} \cdot (1-Iu) + \delta_{j-1,\, i} \cdot uj . \tag{A2}$$

Using a large integer L that satisfies $u/L \ll 1$, we calculated M for higher u by

$$M = \left[M^{(0)}\left(\frac{u}{L}\right) \right]^L . \tag{A3}$$

Appendix B: CALCULATION OF C_{jki} IN EQ. (1.4)

Tensor (C_{jki}) is calculated with the recursion formula. Crossover between a bit string pair is the operation proceeded with the following three steps:

Step 1: Read the $h+1$-th bits of original strings of j and k and put them in the $h+1$-th cells for the operation.

Step 2: Make chiasma occur between the h-th cells and the $h+1$-th cells with the occurrence probability c.

Step 3: Exchange the bit pair of bits in the $h+1$-th cells if h-th pair had been exchanged and chiasma does not occur or h-th pair had not been exchanged and chiasma occurs.

In Step 1, the probability of the $h+1$-th bit of original string j being 1 is given by

$$p_{jh}(j') = \frac{j - j'}{I - h}, \tag{B1}$$

where j' is the number of bit 1's in h earlier bits in the original string j.

Let $P_h(e', j', k', i')$ be the probability that exchange of the h-th bit pair is in the state of e' ($e' = 1$ means exchanged and $e' = 0$ means unexchanged), the number of bit 1's in h earlier bits in the original string j is j', the number of bit 1's in h earlier bits in the original string k is k', and the number of bit 1's in h earlier bits of the j-side string is i' after crossover.

Since j' is the number of bits in h-length string, $0 \le j' \le h$. And since j' is the number of bit 1's chosen from j bits of 1 in the original string, $j - (I - h) \le j' \le j$. Combining these two conditions and reasoning similarly for k' and i', the intervals of parameters wherein $P_h(e', j', k', i')$ can be positive are specified as follows;

$$\max(0, j + h - I) \le j' \le \min(h, j),$$

$$\max(0, k + h - I) \le k' \le \min(h, k),$$

and

$$\max(0, j + k + h - 2I) \le i' \le \min(h, j + k).$$

$P_h(e', j', k', i')$ satisfies the following recursion formula;

$$P_{h+1}(e', j', k', i') = \sum_{e''}\sum_{j''}\sum_{k''}\sum_{i''} P_h(e'', j'', k'', i'')\,\mathrm{Prob}_{h,h+1}(e'', j'', k'', i''|e', j', k', i'), \tag{B2}$$

where $\sum$ operation of e'' is the summation of $e'' = 0$ and $e'' = 1$, that of j'' is the summation of $j'' = j' - 1$ and $j'' = j'$, that of k'' is the summation of $k'' = k' - 1$ and $k'' = k'$, and that of i'' is the summation of $i'' = i' - 1$ and $i'' = i'$. Considering Step 1~3, transition probability $\mathrm{Prob}_{h,\,h+1}(e'', j'', k'', i''|e', j', k', i')$ is given by

$$\mathrm{Prob}_{h,\,h+1}(e'', j'', k'', i''|e', j', k', i') = \mathrm{Prob}(e''|e') \cdot \mathrm{Prob}_h(j''|j') \times \mathrm{Prob}_h(k''|k') \cdot \mathrm{Prob}_{e',\,j'',\,j',\,k'',\,k'}(i''|i'),$$

where

$$\begin{aligned}
&\mathrm{Prob}(e''|e') = 1 - c && \text{if } e'' = e',\\
&\mathrm{Prob}(e''|e') = c && \text{if } e'' \neq e';\\
&\mathrm{Prob}_h(j''|j') = 1 - p_{jh}(j'') && \text{if } j'' = j',\\
&\mathrm{Prob}_h(j''|j') = p_{jh}(j'') && \text{if } j'' = j' - 1;\\
&\mathrm{Prob}_h(k''|k') = 1 - p_{kh}(k'') && \text{if } k'' = k',\\
&\mathrm{Prob}_h(k''|k') = p_{kh}(k'') && \text{if } k'' = k' - 1;\\
&\mathrm{Prob}_{e',\,j'',\,j',\,k'',\,k'}(i''|i') = 1 && \text{if } j'' = j' \text{ and } e' = 0 \text{ and } i'' = i'\\
&&& \text{or } k'' = k' \text{ and } e' = 1 \text{ and } i'' = i'\\
&&& \text{or } j'' = j' - 1 \text{ and } e' = 0 \text{ and } i'' = i' - 1\\
&&& \text{or } k'' = k' - 1 \text{ and } e' = 1 \text{ and } i'' = i' - 1,\\
&\mathrm{Prob}_{e',\,j'',\,j',\,k'',\,k'}(i''|i') = 0 && \text{otherwise.}
\end{aligned}$$

The initial condition is

$$\begin{aligned}
&P_0(e', j', k', i') = 1 && \text{if } e' = j' = k' = i' = 0,\\
&P_0(e', j', k', i') = 0 && \text{otherwise.}
\end{aligned} \qquad \text{(B3)}$$

We can calculate $P_h(e', j', k', i')$ one by one increasing h, until h comes to be equal to I. C_{jki}, the transition probability from j to i by combining with k, is then given by

$$C_{jki} = P_I(0, j, k, i) + P_I(1, j, k, i).$$

Appendix C: DERIVATION OF EQS. (1.7) AND (1.8)

Let

$p_{\mathrm{d}} = 1-(1-u)^I =$ probability of destruction of the advantageous gene set by mutation per gene set per generation;

$p_{\mathrm{c}} =$ probability of creation of the advantageous gene set from the neutral gene set by mutation per generation.

Though destruction probability can be specified by a constant p_{d} depending upon u, creation probability p_{c} is strictly a variable which changes with the number of bit 1's in the neutral gene set. However, we here lump all neutral gene sets and represent the creation probability from those gene sets by one effective parameter p_{c}. Therefore p_{c} is a parameter that changes with the bit distribution in the population. Using p_{d} and p_{c}, Δx_I through mutation is written as follows;

$$\Delta x_I^{(\mathrm{mut.})} = p_{\mathrm{c}}(1-x_I) - p_{\mathrm{d}} x_I . \tag{C1}$$

We here focus on deriving the expression of $\Delta x_I^{(\mathrm{mut.})}$ around the stationary distribution for mutation. Since mutation is an equal process about exchange between bit 0 and bit 1, the stationary state for mutation is a binomial distribution in which bit 0 and bit 1 are equally distributed at each locus. Requiring $\Delta x_I^{(\mathrm{mut.})} = 0$ at this state (namely, $x_I = (1/2)^I$) and approximating $(1/2)^I \ll 1$, we get the relation $p_{\mathrm{c}} \sim p_{\mathrm{d}}(1/2)^I$. The transition formula through mutation around the stationary state is given with this relation and assumption $u \ll 1$ (which gives the approximation $p_{\mathrm{d}} = 1-(1-u)^I \sim Iu$) as follows;

$$\begin{aligned} x_I &\overset{\mathrm{mut.}}{\to} x_I - p_{\mathrm{d}} x_I + p_{\mathrm{c}}(1-x_I) \\ &\sim x_I - p_{\mathrm{d}}\left(x_I - \left(\frac{1}{2}\right)^I (1-x_I)\right) \\ &\sim x_I - Iu\left(x_I - \left(\frac{1}{2}\right)^I\right) . \end{aligned} \tag{C2}$$

Thus Eq. (1.7) was derived.

For crossover, the following definition is necessary:

$P_{\mathrm{hom}} =$ probability that the advantageous set of genes is created after crossover in the pair of neutral sets of genes.

Like p_c defined above, P_{het} and P_{hom} are effective parameters and varies depending upon the bit distribution in the population. Using P_{het} and P_{hom}, Δx_I through crossover is written as follows:

$$\begin{aligned}\Delta x_I^{(cross.)} &= r[x_I^2 + P_{het}x_I(1-x_I) + P_{hom}(1-x_I)^2] - rx_I \\ &= r(1-x_I)[P_{hom}(1-x_I) - (1-P_{het})x_I] \quad . \end{aligned} \tag{C3}$$

We also focus on deriving the expression of $\Delta x_I^{(cross.)}$ around the stationary distribution for crossover. Since crossover is a process making a bit distribution at each locus independent of the others, the stationary state for crossover is a binomial distribution in which bit 1's are distributed at each locus with frequency q independently. Requiring $\Delta x_I^{(cross.)} = 0$ at this state (namely $x_I = q^I$), we get the relation $P_{hom} = (1-P_{het})(q^I/(1-q^I))$. The transition formula through crossover around the stationary state is given with this relation as follows;

$$\begin{aligned} x_I &\xrightarrow{cross.} x_I + r(1-x_I)[P_{hom}(1-x_I) - (1-P_{het})x_I] \\ &= x_I + r(1-x_I)(1-P_{het})\left[\frac{q^I}{1-q^I}(1-x_I) - x_I\right] \\ &= x_I - r(1-P_{het})\frac{1-x_I}{1-q^I}(x_I - q^I) \end{aligned} \tag{C4}$$

Thus Eq. (1.8) was derived.

Appendix D: DERIVATION OF EQS. (1.9) AND (1.10)

With the comparison between Eq. (1.4) and (C3), the accurate equation defining P_{het} is

$$P_{\mathrm{het}}(1-x_I) = \sum_{k\neq I} x_k C_{IkI} + \sum_{j\neq I} x_j C_{jII}.$$

In this appendix we focus on the state before or immediately after the occurrence of an individual with the advantageous combination of genes, hence this equation can be rewritten with approximation $x_I \ll 1$ as follows;

$$P_{\mathrm{het}} = \sum_{k\neq I} x_k C_{IkI} + \sum_{j\neq I} x_j C_{jII} = P_{\mathrm{pres}} + P_{\mathrm{tran}}, \tag{D1}$$

where $P_{\mathrm{pres}} \equiv \sum_{k\neq I} x_k C_{IkI}$ is the probability that an advantageous gene set is preserved after crossover with a neutral gene set and $P_{\mathrm{tran}} \equiv \sum_{j\neq I} x_j C_{jII}$ is the probability that a neutral gene set is transferred to an advantageous gene set after crossover with an advantageous gene set.

In order to calculate P_{pres} and P_{tran}, we assume that the 0-1 distribution of neutral gene sets is specified only by the parameter q at each locus independently. With this assumption, we can get the recursion formula by considering Step 1~3 as was described in Appendix B. Let $P_h(e')$ be the probability that exchange of the h-th bit pair is in the state of e' ($e' = 1$ means exchanged and $e' = 0$ means unexchanged) and all h earlier bits in the string on the 'advantageous after crossover' side are 1 after crossover. $P_h(e')$ satisfies

$$P_{h+1}(e') = \sum_{e''} P_h(e'') \cdot \mathrm{Prob}(e''|e') \cdot \mathrm{Prob}_{e'}(\mathrm{bit}=1) \ , \tag{D2}$$

where

$$\mathrm{Prob}(e''|e') = 1-c \qquad \text{if } e'' = e',$$

$$\mathrm{Prob}(e''|e') = c \qquad \text{if } e'' \neq e';$$

$$\mathrm{Prob}_{e'}(\mathrm{bit}=1) = q_{\mathrm{a}} \qquad \text{if } e' = 0,$$

$$\mathrm{Prob}_{e'}(\mathrm{bit}=1) = q_{\mathrm{n}} \qquad \text{if } e' = 1.$$

q_{a} (or q_{n}) is the probability that the $h+1$-th bit of original string on the 'advantageous after crossover' side (or the other side) is 1 respectively, and given by

$$q_{\mathrm{a}} = 1 \text{ and } q_{\mathrm{n}} = q \qquad \text{for } P_{\mathrm{pres}},$$

or

$$q_a = q \text{ and } q_n = 1 \qquad \text{for } P_{tran}.$$

This simultaneous recurrence equation can be solved under the initial condition,

$$P_0(0) = 1 \text{ and } P_0(1) = 0\,. \tag{D3}$$

With the solution, we can get the probability that all I bits on the 'advantageous after crossover' side are 1 after crossover as follows;

$$P_I(0) + P_I(1) = \sum_{\pm} \frac{1}{2}\left(1 \pm \frac{(1-c)(q_a - q_n) + 2cq_n}{\sqrt{(1-c)^2(q_a - q_n)^2 + 4c^2 q_a q_n}}\right)\lambda_{\pm}^{I} \quad , \tag{D4}$$

where $\sum$ operation of $\pm$ is the summation of expressions using $+$ or $-$ within formula respectively. $\lambda_{\pm}$ are solutions of the characteristic equation and given by

$$\lambda_{\pm} = \frac{1}{2}(1-c)(q_a + q_n) \pm \frac{1}{2}\sqrt{(1-c)^2(q_a - q_n)^2 + 4c^2 q_a q_n} \quad \text{(respectively).}$$

Therefore, $P_{het}(q)$ is given as follows;

$$\begin{aligned} P_{het}(q) &= P_{pres}(q) + P_{tran}(q) \\ &= (P_I(0) + P_I(1))\big|_{q_a = 1,\, q_n = q} + (P_I(0) + P_I(1))\big|_{q_a = q,\, q_n = 1} \\ &= \sum_{\pm}\left(1 \pm \frac{c(1+q)}{\sqrt{(1-c)^2(1-q)^2 + 4c^2 q}}\right)\lambda_{\pm}^{I} \quad , \end{aligned} \tag{D5}$$

where

$$\lambda_{\pm} = \frac{1}{2}(1-c)(1+q) \pm \frac{1}{2}\sqrt{(1-c)^2(1-q)^2 + 4c^2 q} \quad \text{(respectively).}$$

Now we substitute $q = 1/2$ to study around the random population, then

$$P_{het}\left(\frac{1}{2}\right) = \sum_{\pm}\left(1 \pm \frac{3c}{\sqrt{1 - 2c + 9c^2}}\right)\lambda_{\pm}^{I} \quad , \tag{D6}$$

where

$$\lambda_{\pm} = \frac{3}{4}(1-c) \pm \frac{1}{4}\sqrt{1 - 2c + 9c^2} \quad \text{(respectively).}$$

In the domain of $0 \le c \le 0.5$, λ_{+} and λ_{-} move within $0.75 \le \lambda_{+} \le 1$ and $0 \le \lambda_{-} \le 0.5$ respectively; hence, if I is so large that $(0.5)^I \ll 1$, λ_{-}^{I} can be neglected and Eq. (1.9) is derived.

Eq. (1.9) can be transformed with approximation $c \ll 1$ as follows;

$$P_{\text{het}}\left(\frac{1}{2}\right) = \left(1 + \frac{3c}{\sqrt{1-2c+9c^2}}\right)\left\{\frac{3}{4}(1-c) + \frac{1}{4}\sqrt{1-2c+9c^2}\right\}^I$$

$$= (1+3c(1+c))\left(\frac{3}{4}(1-c) + \frac{1}{4}(1-c)\right)^I$$

$$= (1+3c)(1-c)^I$$

$$= (1-c)^{I-3},$$

which gives Eq. (1.10).

Figure 1.1

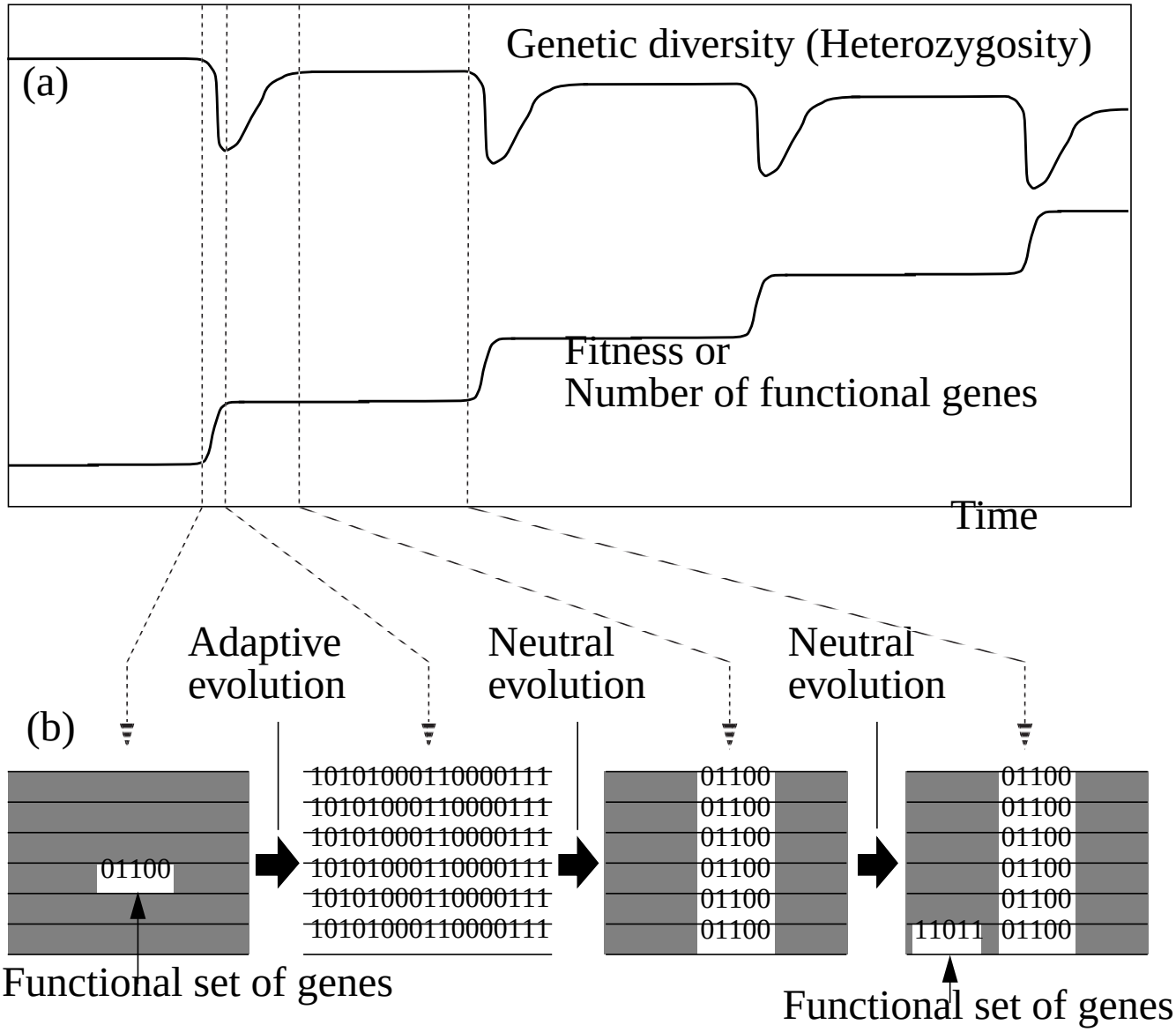

(a)
Genetic diversity (Heterozygosity)
Fitness or
Number of functional genes
Time
(b)
Adaptive
evolution
Neutral
evolution
Neutral
evolution
01100
10101000110000111
01100
11011
Functional set of genes
Functional set of genes

Figure 1.2

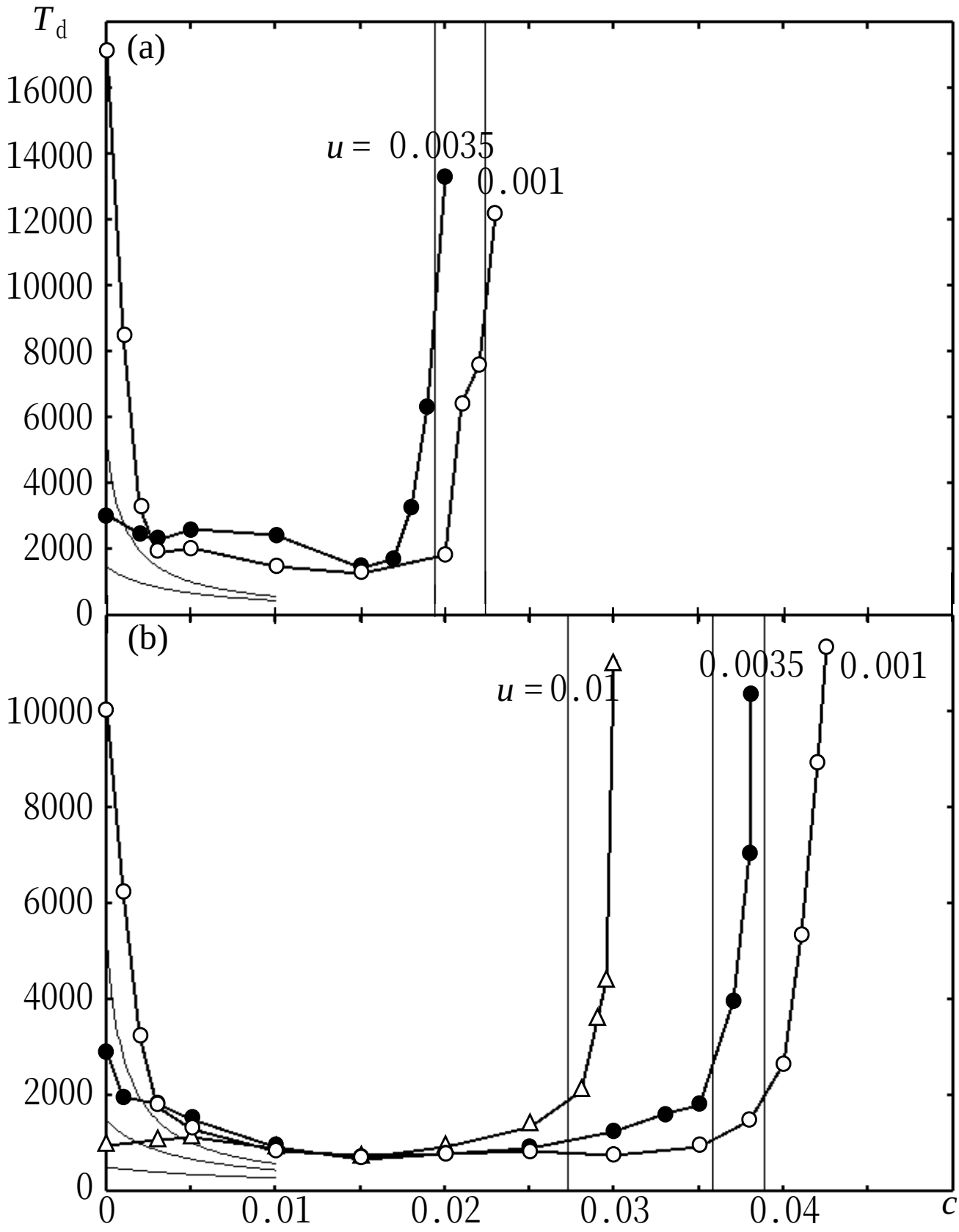

T_d
(a)
16000
14000
12000
10000
8000
6000
4000
2000
0
u = 0.0035
0.001
(b)
10000
8000
6000
4000
2000
0
u = 0.01
0.0035
0.001
0
0.01
0.02
0.03
0.04
c

Figure 1.3

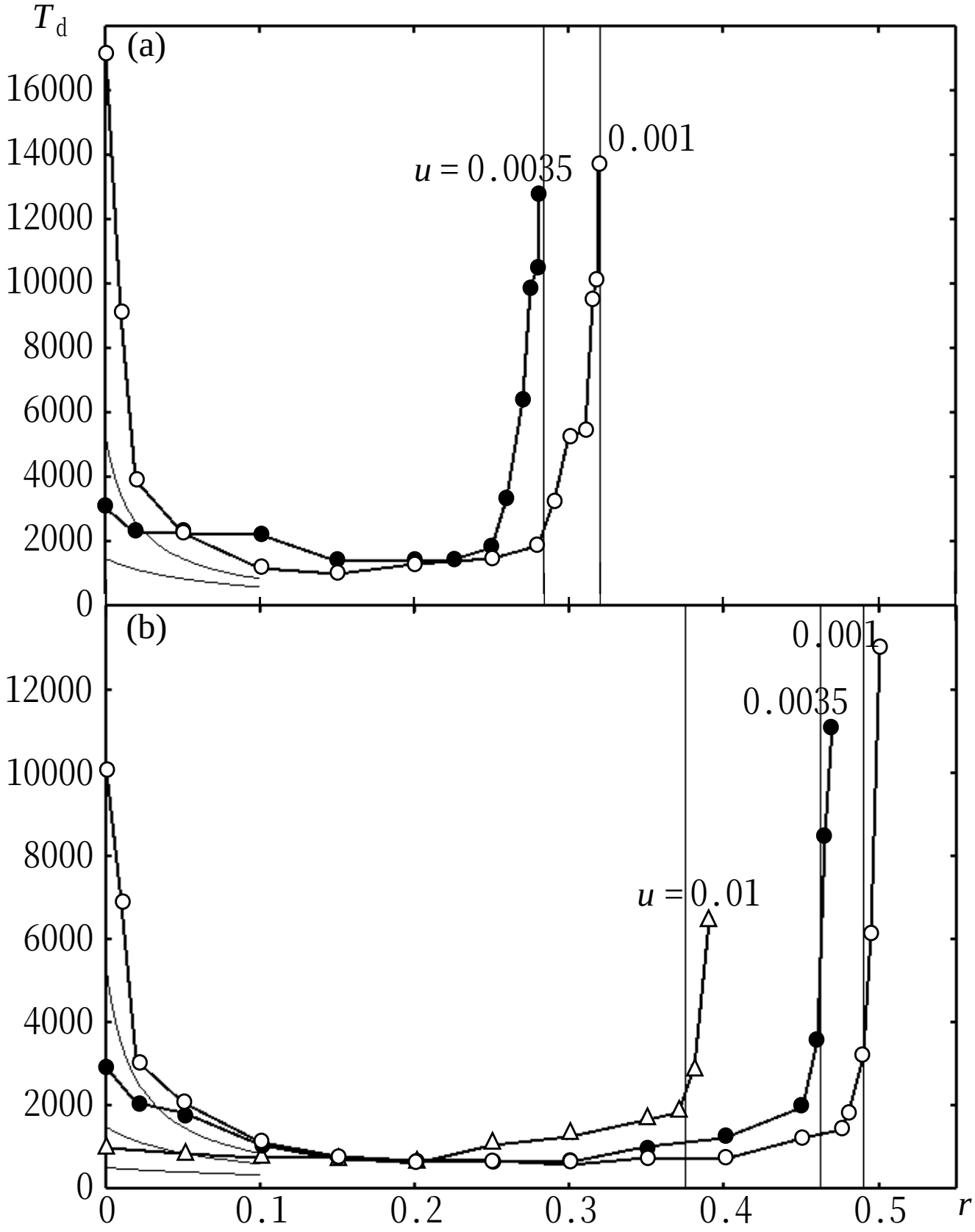

T_d
(a)
16000
14000
12000
10000
8000
6000
4000
2000
0
$u = 0.0035$
0.001
(b)
12000
10000
8000
6000
4000
2000
0
0.001
0.0035
$u = 0.01$
0
0.1
0.2
0.3
0.4
0.5
r

Figure 1.4

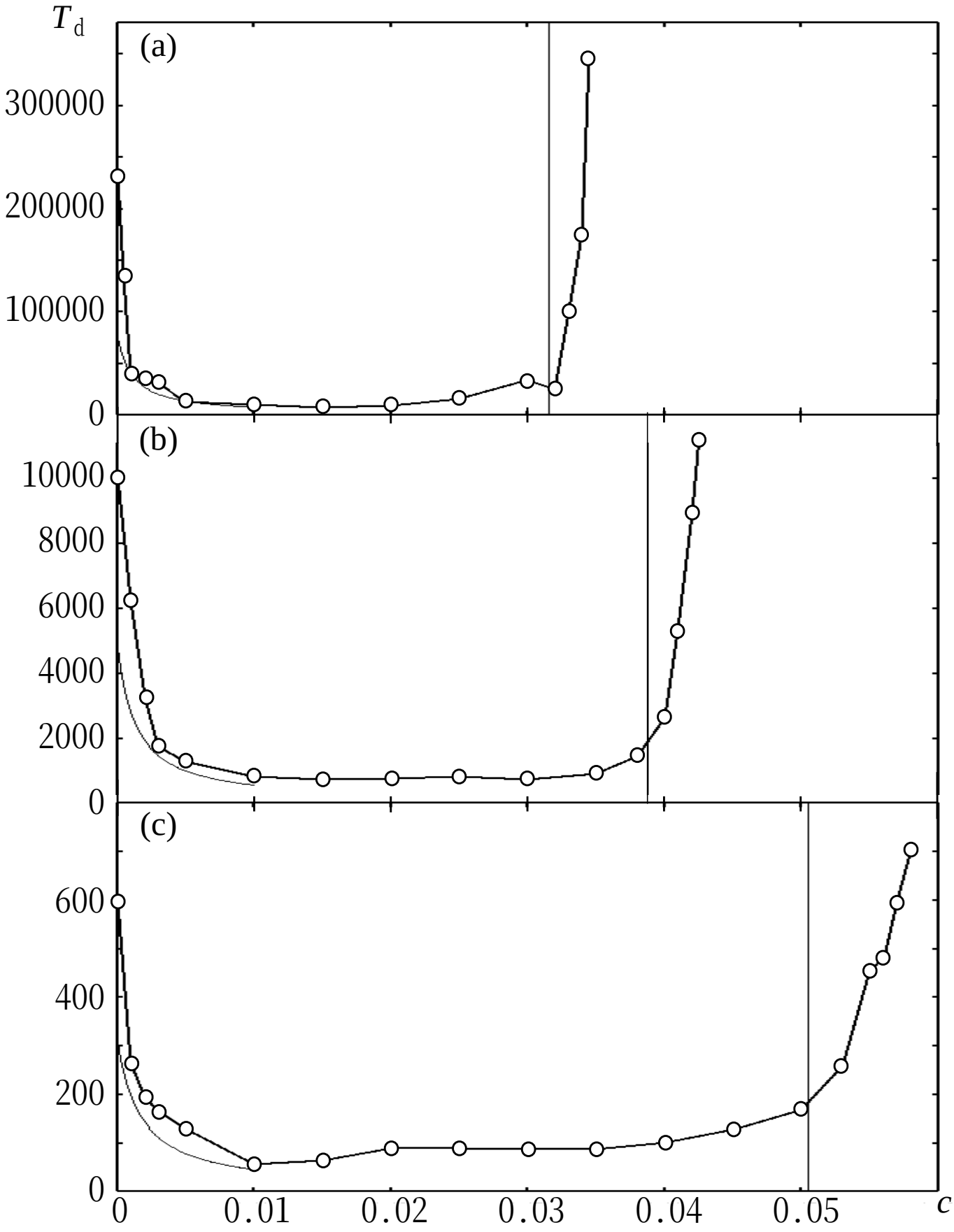

T_d
(a)
300000
200000
100000
0
(b)
10000
8000
6000
4000
2000
0
(c)
600
400
200
0
0
0.01
0.02
0.03
0.04
0.05
c

Figure 1.5

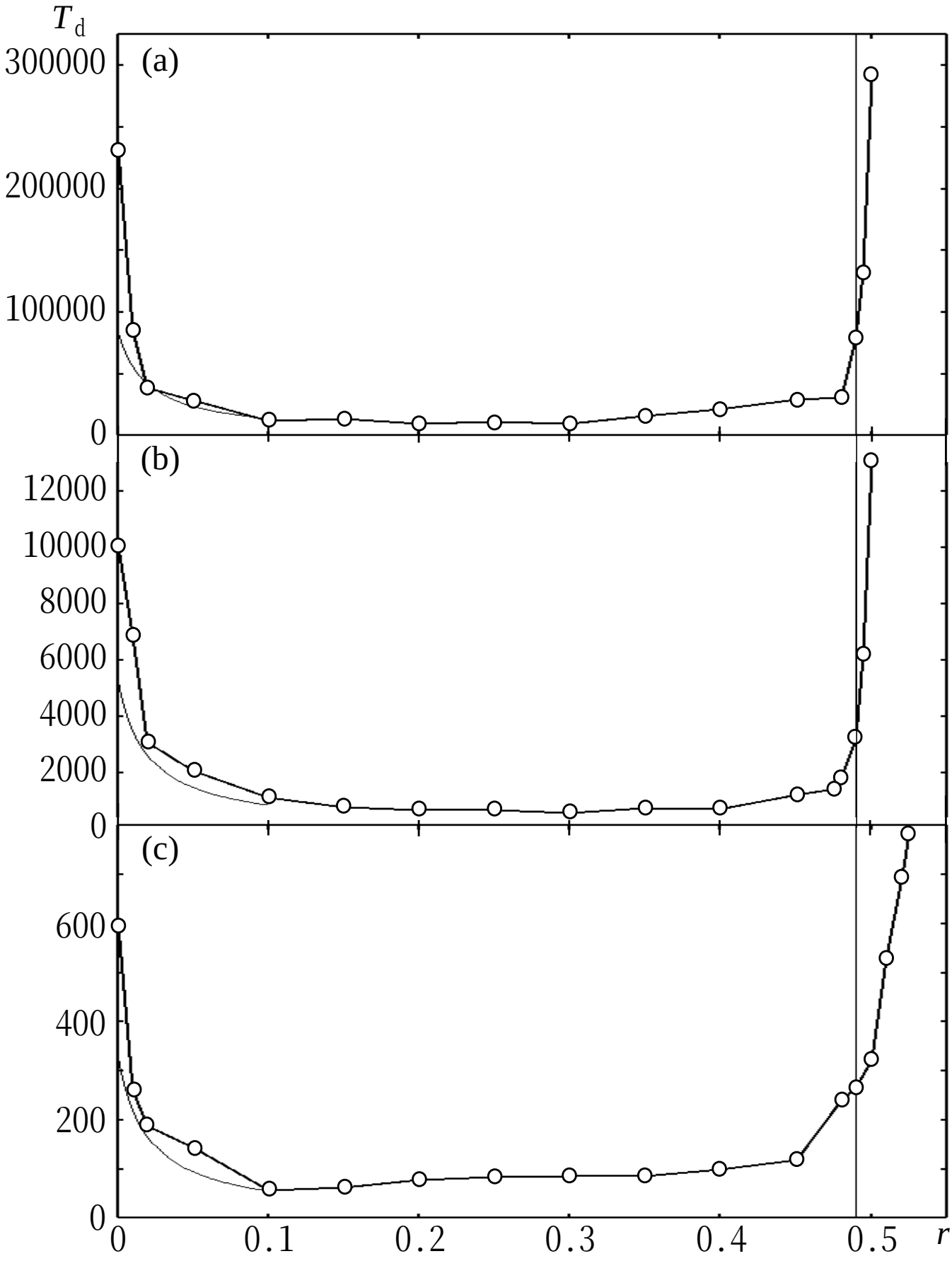

T_d
(a)
300000
200000
100000
0
(b)
12000
10000
8000
6000
4000
2000
0
(c)
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
r

Figure 1.6

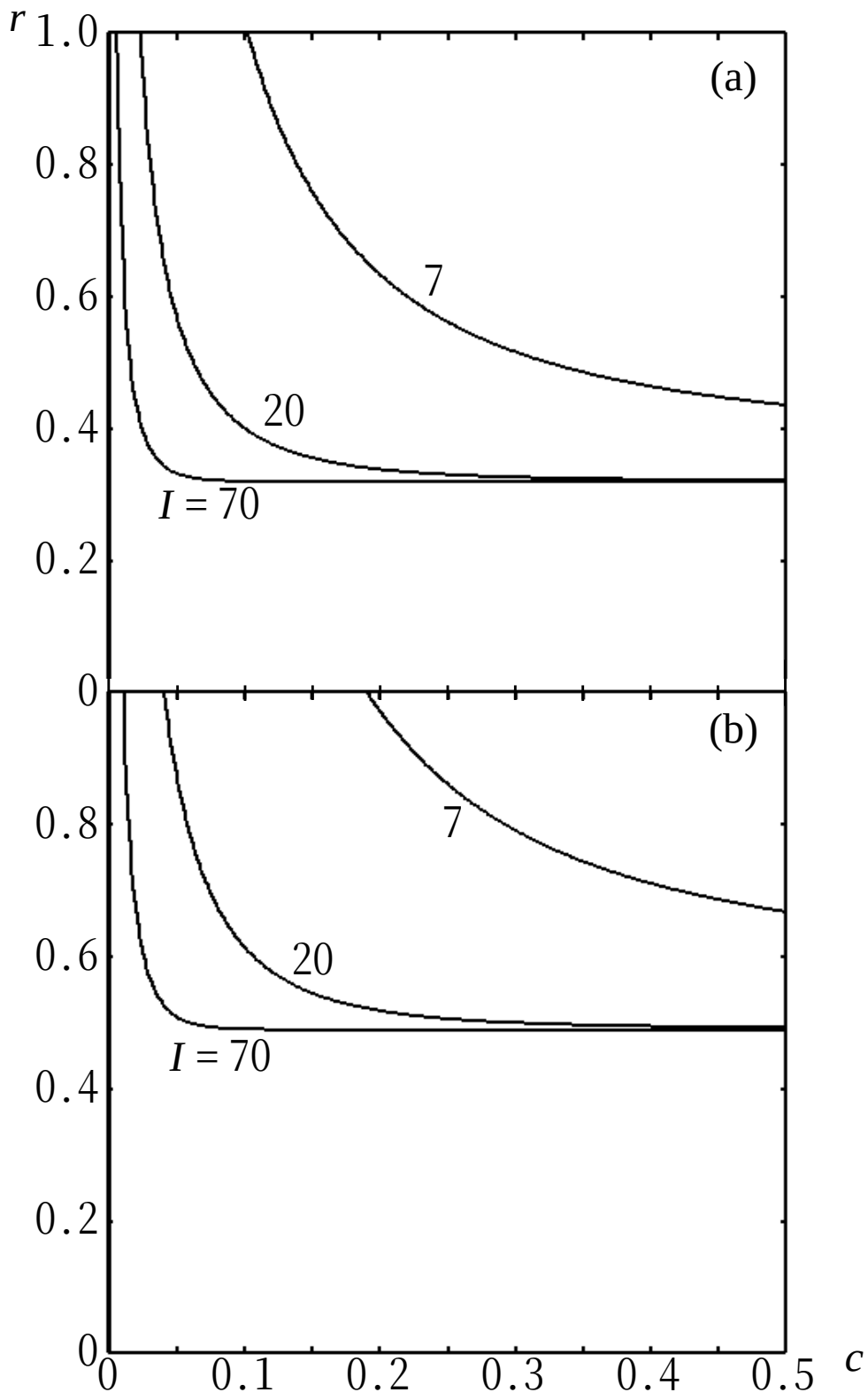
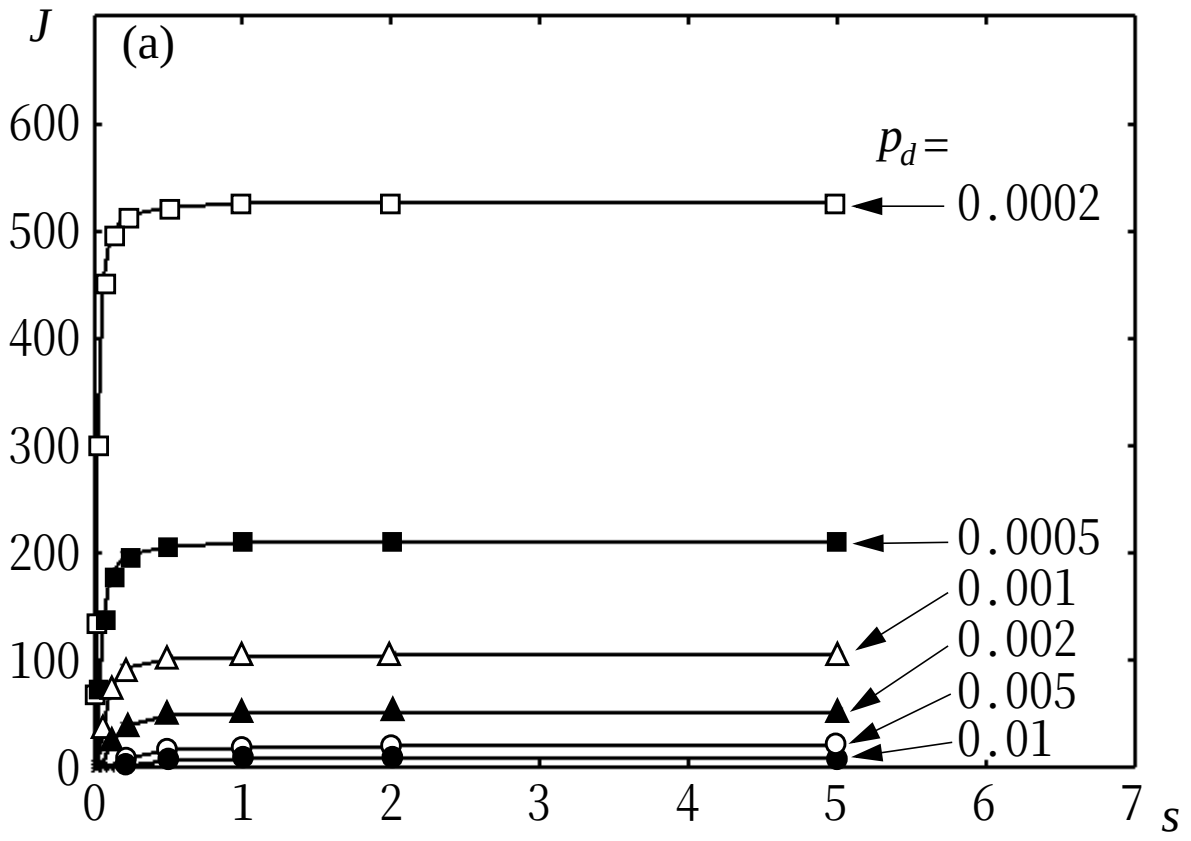


Figure 1.7


J
(a)
600
500
400
300
200
100
0
p_d =
0.0002
0.0005
0.001
0.002
0.005
0.01
0
1
2
3
4
5
6
7
s

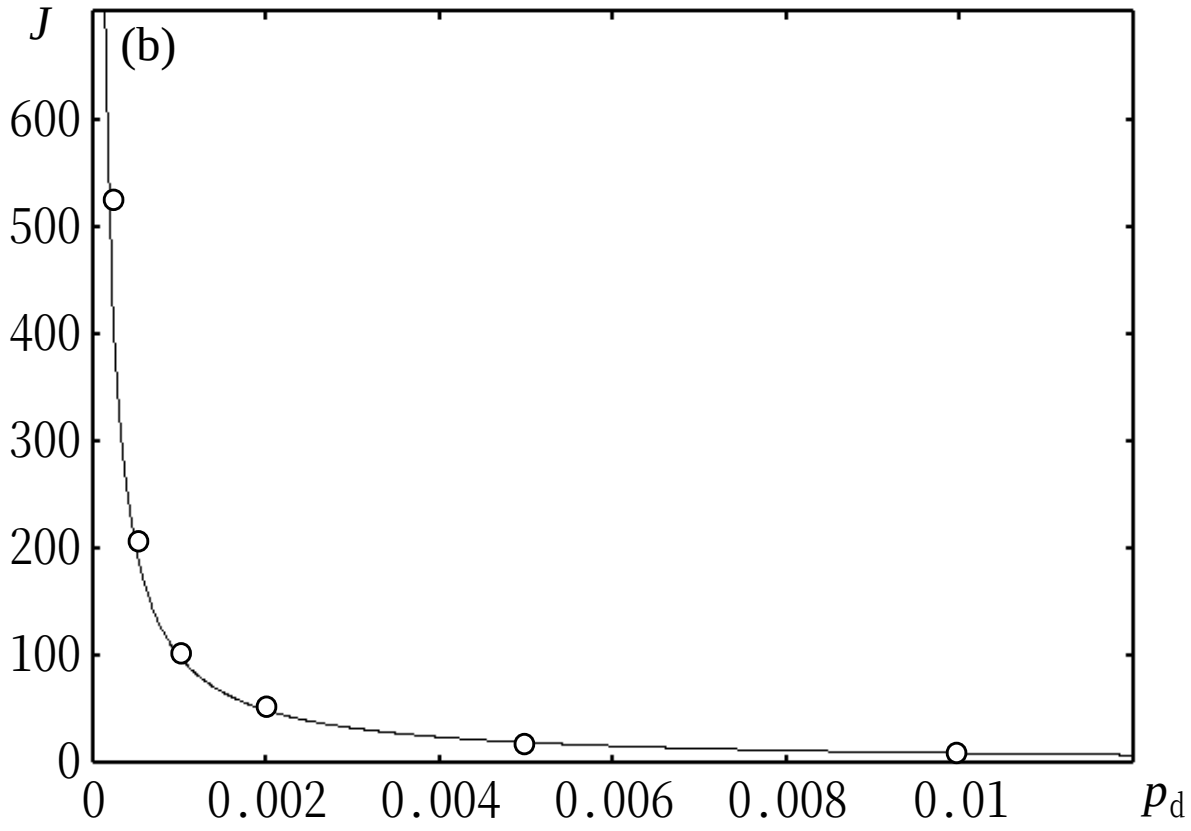

J
(b)
600
500
400
300
200
100
0
0
0.002
0.004
0.006
0.008
0.01
p_d

Figure Legend

Fig. 1.1. General evolutionary picture. (a) The genetic diversity, the number of functional genes, and mean population fitness as a function of time. (b) Genome sketch of the population at each stage of evolution. Shading represents the region wherein genes (or nucleotides) are diverse among individuals.

Fig. 1.2. Mean number of generations required for the advantageous set of genes to dominate the population (domination time T_d) as a function of the occurrence probability of crossover points c under the condition of all individuals participating in crossover ($r = 1$). The values of the other parameters are $N = 10000$, $I = 20$, (a)$s = 0.5$ or (b)$s = 1$, and $u = 0.001$(○), $u = 0.0035$(●), or $u = 0.01$(△). Here and in the subsequent figures in Chapter 1 (Figs. 1.3~1.5), marked points and line segments connecting between them are the results of simulation, curved lines in lower c or r are T_c given by Eqs. (1.10)~(1.13), and vertical lines in higher c or r are maximum values given by Eq. (1.15) or (1.16).

Fig. 1.3. Domination time T_d as a function of the population ratio of crossover r under the condition of high occurrence probability of crossover points ($c = 0.5$). The values of the other parameters are $N = 10000$, $I = 20$, (a)$s = 0.5$ or (b)$s = 1$, and $u = 0.001$(○), $u = 0.0035$(●), or $u = 0.01$(△).

Fig. 1.4. Domination time T_d as a function of c under the condition of $r = 1$. The values of the other parameters are $N = 10000$, $s = 1$, $u = 0.001$, and (a)$I = 24$, (b)$I = 20$, or (c)$I = 16$.

Fig. 1.5. Domination time T_d as a function of r under the condition of $c = 0.5$. The values of the other parameters are $N = 10000$, $s = 1$, $u = 0.001$, and (a)$I = 24$,

(b)$I = 20$, or (c)$I = 16$.

Fig. 1.6. Upper limitations of general c and r given by Eqs. (1.9) and (1.14), with $u = 0.001$, (a) $s = 0.5$ or (b) $s = 1$, and different I.

FIG. 1.7. (a) The number of dominant functional sets of genes J as a function of fitness coefficient s under different destruction probabilities p_d. The plotted values are the results of simulation assuming $f_t = 0.9$. (b) The maximum value of J at large s, as a function of p_d. Marked points are the simulation results and the solid line are the hyperbolic curve given by Eq. (1.21).

CHAPTER 2

Crossover Accelerates Evolution in GAs with a Babel-like Fitness Landscape: Mathematical Analyses

This chapter was done in collaboration with Prof. Yoh Iwasa, Kyushu University.

The paper is to be published in *Evolutionary Computation Journal* after minor revision.

Introduction

Recently in the engineering field, genetic algorithms (GAs) have attracted a great deal of attention as optimization methods (Holland, 1992; Goldberg, 1989; Mitchell, Forrest, & Holland, 1991; Forrest, 1993; Forrest & Mitchell, 1993, 1994; Otto, Feldman, & Christiansen, 1994; Vose & Wright, 1995; Mitchell, 1996). In GAs, each design of an object is typically coded in a gene-like bit sequence and a population of those sequences is prepared in the computer memory. The optimal (or a close to optimal) design is searched for by evolutionary operations including reproduction, natural selection, mutation, and genetic recombination. Since GAs without genetic recombination, or crossover, are nothing more than a parallel hill-climbing method, crossover is a key operation to achieve the optimal design in the shortest number of trials. In previous studies, however, the effectiveness of crossover in GAs has met with varying degrees of success, depending upon the problem domain and the shape of fitness landscape. We are still far from achieving a complete understanding of the role of crossover in GAs, which is important not only for an engineering purpose but also from a biological point of view.

Recently, Suzuki (1996, 1997) proposed an evolutionary programming methodology, Multiple von Neumann computers with machine language architecture (MUNCs). MUNCs evolve and create a program that solves a problem prepared in an environmental database. Suzuki used GAs for the evolution of MUNCs and found that crossover contributed to the acceleration. A key point causing this result was the fitness function of bit sequences optimized by GAs. In MUNCs, an advantageous function is achieved by a particular combination of machine instructions, and without such sequence, no functional advantage is conferred. A particular combination of instructions corresponds to a schema with fairly large *order*; hence in MUNCs, bit sequences optimized by GAs have a highly discontinuous fitness function.

Up to the present, a discontinuous fitness function has studied by many GA researchers, exemplified by the royal road function (Mitchell *et al.*, 1992; Jones, 1994; Wu & Lindsay, 1996; Nimwegen, Crutchfield, & Mitchell, 1996, 1997). This function takes discrete fitness values of the bit sequence, depending upon the number of ‘blocks’ (schemata) that is included in a haplotype. Recent work by Nimwegen *et al.* (1996, 1997) investigated the dynamics of GAs in this function theoretically and established a

method to analyze GAs in a multiple block fitness function.

However, Nimwegen *et al.*, inferred the role of crossover based only on the numerical experiment, but did not make any quantitative estimate of the effectiveness of crossover. This was done in Chapter 1 for the case with a much simpler fitness landscape. In order to study the rate of evolution of a single advantageous function, Suzuki assumed a fitness function with a single advantageous schema and examined the time until domination of the advantageous schema by the simulation using the frequency vector representation of the population.

We extend this study by assuming the same fitness function (we called this a Babel-like fitness function) and examined the GA performance in both experimental and theoretical ways. A preliminary result was reported in Suzuki & Iwasa (1997). In the simulation, we used simple GAs (Goldberg, 1989) directly operating on the population of bit sequences. Besides the theoretical study, we also estimated the time until domination as a function of several genetic parameters. These results showed that crossover with a mildly high rate can greatly enhance the evolution if genetic parameters are adjusted to appropriate values.

Although the theoretical analysis by Suzuki & Iwasa (1997) were more accurate than the analysis in Chapter 1, their analyses were still unsatisfactory in the following points. First, their theoretical estimation of the time until creation of the advantageous schema was based upon the simplifying assumption that the diversity of a population is so large that the bit distribution does not depend upon the history of previous distribution of bit sequences. Second, since they evaluated the final formula by the aid of the Monte Carlo method, they could not estimate accurate values of the acceleration rate when the order of an advantageous schema is large.

The aim of this chapter is to remedy these drawbacks and to present a more exact estimation of the evolutionary rate by GAs. As in previous works, we focus on a single advantageous function and assume a Babel-like fitness landscape (among a large number of possible bit configurations, a single sequence is much more advantageous than the others). The estimation of the evolutionary rate is made in three different ways. First, we develop a mathematical estimation method. The second one, which was developed in Chapter 1, is an analysis using recurrence formulas for a frequency vector expressing the composition of the population. The third one is a straightforward computer simulation

using simple GAs.

The organization of the chapter is as follows. In the next section, a basic fitness model is presented and the evolutionary process under this function is described. Then we explain the three methods one by one; first we present the analytical formulas for two extreme cases, next we explain the simulation method using vector representation, and then we describe the computer simulation method with GAs. Finally we compare results of all three methods. We also discuss the mechanism of GAs and roles of crossover, together with the implementation criteria for several genetic parameters.

The Basic Fitness Model and the Evolutionary Picture

Here we present a basic fitness model for the evolution of a single advantageous function. The assumptions of the model are as follows.

- ¶Individuals are haploid. The chromosome of each individual is represented by a string of a number of binary (0-1) loci.
- ¶Among all sequences, a single sequence, denoted by $[11\cdots1]$, has fitness by far larger than the others, and all the other sequences have the same fitness.
- ¶In the initial population, all individuals have the same haplotype which was chosen randomly.
- ¶Mating pair is randomly chosen in the population which is sufficiently large.

Evolution under this fitness function proceeds with the following steps.

Step (i) [Diversification]: Mutation stores diversity until the population reaches an equilibrium state determined as the balance between mutation and genetic drift.

Step (ii) [Creation]: Mutation and crossover search for and create an advantageous sequence. Due to the finiteness of the population size, a newly created advantageous sequence may be lost by chance, and the cycles of creation and extinction of the advantageous sequence are repeated several times until the advantageous sequence begins to spread through the population.

Step (iii) [Spread]: The advantageous sequence spreads and dominates the population.

Here the term ‘domination’ does not mean that one chromosome completely occupies the whole population but it implies that the frequency exceeds 50U of the whole population.

We designate the average number of generations from the initial setting until some domination criterion is satisfied the domination time T_{d}. Based on the above evolutionary process, T_{d} is the sum of three parts;

$$T_{\mathrm{d}} = T_{\mathrm{v}} + T_{\mathrm{c}}N_{\mathrm{c}} + T_{\mathrm{s}}, \tag{2.1}$$

where

T_{v}: the diversification time, defined as the average number of generations taken for the population to store sufficient diversity starting from the initial homogeneous distribution;

T_c: the creation time, defined as the average number of generations between an appearance of the advantageous sequence followed by the extinction of its descendants and the next appearance in a sufficiently diverse population;

N_c: the creation number, defined as the average number of appearances of the advantageous sequence from its absence until it begins to spread;

T_s: the spread time, defined as the average number of generations for a novel advantageous sequence to spread through the population.

Note that N_c is the same as the destruction number N_d defined in Chapter 1. For the second term, we neglected the correlation between T_c and N_c, or in other words, we assumed that one appearance of the advantageous sequence does not affect next creation of the advantageous sequence.

As pointed out in Chapter 1, crossover randomizes diverse regions of the genome and helps to create novel sequences, and hence T_c decreases and N_c increases with the crossover rate, which realizes T_d minimum at an intermediate crossover rate. This dependence was estimated in Chapter 1 with some crude approximations. In Chapter 2, we estimate T_d more accurately by the use of Eq. (2.1). In addition, we here pay a particular attention to the evolutionary acceleration effect by crossover. In order to estimate this effect quantitatively, we define the acceleration rate of crossover as

$$A_{\mathrm{cross}} = \frac{T_d\big|_{\mathrm{without\ crossover}}}{T_d\big|_{\mathrm{with\ crossover}}} \tag{2.2}$$

and evaluate this under various values of genetic parameters. The genetic parameters used in this chapter are as follows.

I: total number of bits in the sequence, or the number of bits necessary to realize a novel advantageous function (we hereafter call this “epistatic number”);

N: population size;

N_e: effective population size;

s: selection coefficient of the advantageous sequence relative to the others;

u: mutation rate (probability of bit flipping) per locus (bit position) per generation;

c: probability of a crossover point (chiasma) occurring per pair of neighboring binary loci;

r: fraction of the sequences which participates in crossover.

Throughout this chapter, the value of s is taken to be 1. The crossover rate is specified by two parameters, c and r. In the following, we consider two modes of crossover designated as FPAP-mode (few-points major-participants mode, namely $c < 0.5$ and $r = 1$) and MPIP-mode (many-points minor-participants mode, namely $c = 0.5$ and $r < 1$). For MPIP-mode, the population is divided into two parts, the recombining subpopulation and the non-recombining subpopulation.

Mathematical Analyses

Here we describe the mathematical analyses and present the estimation method of the domination time T_{d} and the acceleration rate A_{cross}. Let q_i be the frequency of bit 1s at the *i*-th locus. We first give the following simplifying assumptions.

¶The 0-1 distribution at the *i*-th locus is determined only by the frequency parameter q_i and is independent of the other loci.

¶The spread time T_{s} is neglected.

Strictly speaking in the simulation model of finite population size, some linkage disequilibrium (non-zero correlation) automatically develops between loci due to less than complete recombination. However, as an approximation, we here neglect correlation between loci and express the state of the population using the I-dimensional frequency vector $\{q_i\} = \{q_1, \cdots, q_I\}$. The second approximation is acceptable because the selection coefficient $s = 1$ is large enough to make the advantageous sequence, once escaped from the initial danger of extinction, spread rapidly until its final domination.

In the following, we give the formulas for T_{v}, T_{c}, and N_{c} one by one.

Diversification Time: Diversification is a process promoted by mutation. Although crossover facilitates creation of novel sequences, the creation process is based upon currently stored diversity, and crossover itself cannot increase the diversity at each locus. Therefore, $H^{(t)}$, that is, heterozygosity at each locus at generation *t*, is determined only by mutation and random genetic drift, and is given by the following equation.

$$H^{(t)} = H^{(\infty)}\left[1 - \exp\left(-\left(\frac{1}{N_{\mathrm{e}}} + 4u\right)t\right)\right]$$

$$= \frac{2N_{\mathrm{e}}u}{1 + 4N_{\mathrm{e}}u}\left[1 - \exp\left(-\left(\frac{1}{N_{\mathrm{e}}} + 4u\right)t\right)\right] . \qquad (2.3)$$

In deriving of this equation, approximations $u \ll 1$ and $N_{\mathrm{e}} \gg 1$ are used (see Kimura (1983) chap. 8). According to this equation, the value of $H^{(t)}$, which is initially 0, gradually increases and approaches the value of $H^{(\infty)}$ with relaxation time $(1/N_{\mathrm{e}} + 4u)^{-1}$. Hence, the diversification time T_{v} is roughly estimated as

$$T_{\mathrm{v}} \approx \left(\frac{1}{N_{\mathrm{e}}} + 4u\right)^{-1} . \qquad (2.4)$$

General Strategies for calculating Creation Time T_{c} ***and Creation Number*** N_{c}: We

present in this subsection a general strategy to evaluate T_c and N_c. According to population genetics theory (Crow & Kimura, 1970; Ewens, 1979, p.155), after the lapse of T_v (namely after the population have stored sufficient diversity at a given parameter condition) and without selection, distribution at each locus reaches an equilibrium state in which q_is obey the beta distribution described by

$$h(q) \equiv \frac{\Gamma(4\beta)}{\{\Gamma(2\beta)\}^2} q^{2\beta-1}(1-q)^{2\beta-1} \ , \ \beta \equiv N_e u. \tag{2.5}$$

This formula states that when $\beta \gg 1$, q_i is likely to have some intermediate value around 0.5, but when $\beta \ll 1$, allele frequency q_i tends to either near 0 or near 1 even after T_v has elapsed.

Here we introduce a parameter l defined as the total number of bit 1's for a haplotype (a bit sequence). l represents the distance of a haplotype to the advantageous one, and its population average value, denoted by $Q \equiv E(l)$, is rewritten as

$$Q = E(l) = E(\sum_{i=1}^{I} b_i) = \sum_{i=1}^{I} E(b_i) = \sum_{i=1}^{I} q_i,$$

where b_i is an allele value parameter at the i-th locus. Because for a finite population the frequency vector $\{q_i\}$ fluctuates every generation, the value of Q also fluctuates satisfying some probability distribution. We express this distribution by $P(Q)$ and calculate it using Eq. (2.5) as

$$P(Q) = \int_0^1 \cdots \int_0^1 \delta(Q - \sum q_i) \prod_i h(q_i) dq_i. \tag{2.6}$$

This equilibrium distribution is illustrated in Fig. 2.1 for different values of β. According to this figure, when β is large, $P(Q)$ has a continuous distribution with a single peak, whereas when β is much smaller than one, the value of Q is discretized and is most likely to have a value around an integer.

Hence in the following, we formulate T_c and N_c in two extreme cases, large β ($\beta \gtrsim 1$) and small β ($\beta \ll 1$). When $\beta \gtrsim 1$, the vector $\{q_i\}$ fluctuates so fast that a population reaches the quasi-equilibrium state (the state of the population before creation of the advantageous sequence) in a time much shorter than the creation time T_c. We can calculate in this case the creation rate of the advantageous sequence by using the weighted average for the quasi-equilibrium state. When $\beta \ll 1$, on the other hand, the

value of $\{q_i\}$ strongly depends upon previous values and the waiting time until a population reaches the quasi-equilibrium state is much larger than T_c. In this case, we need another approximation method to evaluate T_c and N_c.

Creation Time for large β: Since creation of the advantageous sequence results from randomization processes caused by mutation and crossover, the creation time T_c is calculated as

$$T_c = (\text{the probability per generation for the advantageous sequence to be created from absence in the population})^{-1}$$

$$= [N \cdot (\langle p_c^{(\text{mut.})} \rangle_{BC} + \langle p_c^{(\text{cross.})} \rangle_{BC})]^{-1} , \tag{2.7}$$

where $p_c^{(\text{mut.})}$ and $p_c^{(\text{cross.})}$ are the probabilities of an arbitrary sequence becoming advantageous by mutation or by crossover respectively. Symbol $\langle A \rangle_{BC}$ denotes the expected value of a quantity A 'Before Creation' of the advantageous sequence. See Appendix E for detailed formulas of $p_c^{(\text{mut.})}$ and $p_c^{(\text{cross.})}$.

The averaging operation $\langle\ \rangle_{BC}$ is calculated as follows. When β is large, the vector $\{q_i\}$ is continuously distributed (Fig. 2.1a). Under this circumstances, the vector $\{q_i\}$ fluctuates so fast that we can calculate $\langle A \rangle_{BC}$ with a weighted average formulated as

$$\langle A \rangle_{BC} = (\text{normalization factor}) \times \int_0^1 \cdots \int_0^1 (W_{BC} \cdot A) \prod_i dq_i . \tag{2.8}$$

W_{BC} is the weight for the quasi-equilibrium state before creation of the advantageous sequence and the normalization factor is a constant determined by the requirement $\langle 1 \rangle_{BC} = 1$. If the value of $\{q_i\}$ does not depend upon the history, W_{BC} is the product of two factors: the probability in the equilibrium and the probability of the absence of the advantageous sequence. The probability distribution in the equilibrium is proportional to the product given by $\prod h(q_i)$ where Eq. (2.5) is used. The second factor is calculated using Poisson distribution. Since the probability of a randomly chosen sequence being advantageous is very small, the number of the advantageous sequence in the population (denoted by n_I) obeys the Poisson distribution with average $N\prod q_i$,

$$p(n_I) = \frac{(N\prod q_i)^{n_I}}{n_I!} \exp(-N\prod q_i) . \tag{2.9}$$

Then the second factor is $p(0) = \exp(-N\prod q_i)$. Thus W_{BC} is written as

$$W_{\mathrm{BC}} = \prod h(q_i) \cdot \exp(-N \prod q_i) \,. \tag{2.10}$$

We can calculate $\langle p_{\mathrm{c}}^{(\mathrm{mut.})} \rangle_{\mathrm{BC}}$ and $\langle p_{\mathrm{c}}^{(\mathrm{cross.})} \rangle_{\mathrm{BC}}$ by using Eqs. (2.8) and (2.10). See Appendix E for the detailed method for the numerical evaluation.

Creation Number for large β: Although both mutation and crossover help creating the advantageous sequence, once the creation has achieved, both operate to destroy the newly created sequence. The creation number N_{c} is determined from the balance between these destructive forces and the selective advantage of the created sequence. Let z be the extinction probability of a newly created advantageous sequence becoming extinct rather than spreading. The number of times of the appearances of the advantageous sequence from absence until the final domination obeys a geometric distribution $z^{k-1}(1-z)$ $(k = 1, 2, 3, \cdots)$. N_{c} is the average of this distribution:

$$N_{\mathrm{c}} = \sum_{k=1}^{\infty} k \cdot z^{k-1}(1-z) = \frac{1}{1-z} \,. \tag{2.11}$$

In order to calculate z, we use the technique of branching processes. When the sequence is created, there is only one copy in the population. Assume that the number of offspring of the advantageous sequence has a Poisson distribution with a mean value μ. Then, the standard population genetic calculation (see Appendix G) gives the relation

$$z = \exp(-\mu(1-z)) \,. \tag{2.12}$$

μ is the expected value of the ratio of the frequency of the advantageous sequence at the next generation compared to the current frequency, and can be approximately formulated as the product of the frequency ratios by selection, mutation, and crossover;

$$\mu \approx (1+s)(1 - \langle p_{\mathrm{d}}^{(\mathrm{mut.})} \rangle_{\mathrm{AC}})(1 - \langle p_{\mathrm{d}}^{(\mathrm{cross.})} \rangle_{\mathrm{AC}}) \quad , \tag{2.13}$$

where $p_{\mathrm{d}}^{(\mathrm{mut.})}$ and $p_{\mathrm{d}}^{(\mathrm{cross.})}$ are the probabilities of a newly created advantageous sequence being destroyed by mutation and crossover respectively. $\langle A \rangle_{\mathrm{AC}}$ denotes the expected value of a quantity A immediately 'After Creation' of the advantageous sequence. See Appendix E for the formulas for $p_{\mathrm{d}}^{(\mathrm{mut.})}$ and $p_{\mathrm{d}}^{(\mathrm{cross.})}$.

$\langle\ \rangle_{\mathrm{AC}}$ is calculated in the similar way to that for $\langle\ \rangle_{\mathrm{BC}}$:

$$\langle A \rangle_{\mathrm{AC}} = (\mathrm{normalization\ factor}) \times \int_0^1 \cdots \int_0^1 (W_{\mathrm{AC}} \cdot A) \prod_i dq_i \,. \tag{2.14}$$

W_{AC} is the weight for the quasi-equilibrium state immediately after creation of the

advantageous sequence and is formulated as the product of $\prod h(q_i)$ and the probability of there existing only one advantageous sequence in the population. The latter factor is given by $p(1) = N(\prod q_i)\exp(-N\prod q_i)$ (see Eq. (2.9)), so that omitting a constant coefficient, we get

$$W_{\mathrm{AC}} = \prod q_i h(q_i) \cdot \exp(-N\prod q_i)\,. \tag{2.15}$$

$\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{\mathrm{AC}}$ and $\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{\mathrm{AC}}$ are calculated from the combination of Eqs. (2.14) and (2.10). See also Appendix E for the evaluation method.

Creation Time and Creation Number for small β: As was shown in Figs. 2.1b and 1c, the value of q_i tends to be near 0 or 1 and the value of $Q = \sum q_i$ is discretized when β is much smaller than one. For a small β, the state transition of the population cannot be approximated by a random sampling for all the probabilities. Rather the population makes Markovian jumps between "partial distributions" which are centered around integers. Within a partial distribution the vector $\{q_i\}$ is chosen anew every generation. We express the state of the population at generation t using the $(I+1)$-dimensional probability vector $\{\hat{x}_i^{(t)}\} = (\hat{x}_0^{(t)}, \cdots, \hat{x}_I^{(t)})$ whose i-th entry $\hat{x}_i^{(t)}$ denotes the probability that the population at generation t is in the i-th partial distribution $(Q \approx i)$ and that no advantageous sequence has ever been created. On this vector representation, we formulate the generation cycle with the recursion formulas for mutation and creation. As we focus on the history before the first creation of the advantageous sequence, we need not formulate selection. Crossover is considered only in the creation step because crossover itself does not change the value of $\{q_i\}$.

Transition formulas for mutation and creation are

$$\hat{x}_i \overset{\mathrm{mut.}}{\to} \sum_{j=0}^{I} \hat{x}_j M_{ji}, \tag{2.16a}$$

$$\hat{x}_i \overset{\mathrm{cre.}}{\to} \hat{x}_i O_i, \tag{2.16b}$$

where M_{ji} is the probability of the population in the j-th partial distribution shifting to the i-th partial distribution by mutation. To formulate this matrix, we consider the probability of the allele distribution concentrating $q \approx 0$ $(q \approx 1)$ shifting to the distribution concentrating $q \approx 1$ $(q \approx 0)$. Since this probability is the product of the probability of a mutation occurring in the whole population and the probability of its fixation, it is $\{1-(1-u)^N\} \times (1/N) \approx Nu \cdot (1/N) = u$. This is the same as the

mutation rate itself so that M_{ji} is the same as the probability of transition from a sequence with j 1 bits to a sequence with i 1 bits by mutation. See Appendix A for the detailed formula of M_{ji}. To calculate O_i, we use a similar argument as that used in the derivation of Eq. (2.7). Since an advantageous sequence is created through mutation and crossover, O_i is formulated as

$$O_i = 1 - [N \cdot (\langle p_{\mathrm{c}}^{(\mathrm{mut.})} \rangle_{Q \approx i} + \langle p_{\mathrm{c}}^{(\mathrm{cross.})} \rangle_{Q \approx i})] \text{ if } i < I, \tag{2.17a}$$

$$O_I = 0, \tag{2.17b}$$

where $p_{\mathrm{c}}^{(\mathrm{mut.})}$ and $p_{\mathrm{c}}^{(\mathrm{cross.})}$ are the creation probabilities by mutation and crossover given by Eqs. (E2) and (E8) respectively. $\langle A \rangle_{Q \approx i}$ means the expected value of quantity A under the i-th partial distribution. See Appendix E for the evaluation of $\langle p_{\mathrm{c}}^{(\mathrm{mut.})} \rangle_{Q \approx i}$ and $\langle p_{\mathrm{c}}^{(\mathrm{cross.})} \rangle_{Q \approx i}$.

Starting from the initial binomial distribution $\hat{x}_i^{(0)} = \binom{I}{i}\left(\frac{1}{2}\right)^i$ $(i = 0, 1, \cdots, I)$, we simulate the generation cycle by the use of recurrence formulas (2.16a) and (2.16b). The value of $\sum_i \hat{x}_i^{(t)}$ is initially equal to one, decreases, and gradually approaches zero. Through the recursion, we calculate two quantities, the creation time T_{c} and the creation probability vector $\{e_i\}$, using

$$T_{\mathrm{c}} = \sum_{t=1}^{\infty} t \times \left(\sum_i \hat{x}_i^{(t-1)} - \sum_i \hat{x}_i^{(t)} \right) = \sum_{t=0}^{\infty} \sum_i \hat{x}_i^{(t)} ,$$

$$e_i = \sum_{t=0}^{\infty} (1 - O_i) \hat{x}_i^{(t)} .$$

The i-th entry e_i of the vector $\{e_i\}$ is the probability of creation of the advantageous sequence occurring from the i-th partial distribution, and from this vector, we can calculate the creation number N_{c} in the following way. N_{c} is calculated from the extinction probability z (see Eq. (2.11)). Since the value of z depends upon the partial distribution from which the advantageous sequence is created, we estimate $\langle z \rangle_{Q \approx i}$ for different i and calculate the weighted average

$$z = \frac{\sum_i e_i \langle z \rangle_{Q \approx i}}{\sum_i e_i}. \tag{2.18}$$

Like Eqs. (2.12) and (2.13), $\langle z \rangle_{Q \approx i}$'s are calculated from

$$\langle z\rangle_{Q\approx i} = \exp(-\langle\mu\rangle_{Q\approx i}\cdot(1-\langle z\rangle_{Q\approx i}))\,, \tag{2.19}$$

$$\langle\mu\rangle_{Q\approx i} \approx (1+s)(1-\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{Q\approx i})(1-\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{Q\approx i})\ . \tag{2.20}$$

The formulas for $p_{\mathrm{d}}^{(\mathrm{mut.})}$ and $p_{\mathrm{d}}^{(\mathrm{cross.})}$ are given in Appendix E (Eqs. (H1a) and (H1b)). See Appendix E for $\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{Q\approx i}$ and $\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{Q\approx i}$.

Simulation using a Vector Representation of the Population

The second method we adopted is a simulation using a vector representation of the population. This method is almost the same as that developed in Chapter 1, and here we again adopt this method to show the validity of this method if some modifications are made as to the initial condition and population diversity, and also to compare its performance with a new method described in the next section.

The experimentation is as follows. We represented a population at generation t by the frequency vector $(x_0^{(t)}, \cdots, x_I^{(t)})$ whose i-th entry $x_i^{(t)}$ represents the frequency of sequences with i of bit 1's at generation t. We started with initial frequency vector $(x_i^{(0)}) = (\delta_{ik}) = (0, \cdots, 1, \cdots, 0)$ (where k is determined from the number of bit 1's in a bit sequence that is chosen randomly), and simulated the generation cycle by operating recursion relations formulated for selection, mutation, crossover, and random genetic drift. The recurrence formulas were the same as those given in Chapter 1, except for a slight modification of the crossover tensor (C_{jki}). Since crossover has no effect in a perfectly homogeneous (monomorphic) population, the effectiveness of crossover changes with heterozygosity H. To incorporate this dependence, we changed the entries of (C_{jki}) according to the theoretical value of $H^{(t)}$ calculated from Eq. (2.3). See Appendix K for the modified method to calculate (C_{jki}). The generation cycle with the recurrence formulas was repeated until the frequency of the advantageous sequence (x_I) exceeded 0.5, and then the final generation number was recorded as the domination time T_{d}. For each parameter set, 30 replicates of numerical trials were conducted and the average value of T_{d} was calculated.

Simulation with GAs using a Direct Representation of the Population

The third method we adopted to estimate the evolutionary rate is the simple GA that operated on the population of bit strings directly prepared in the computer memory. The simulation procedure was as follows:

Step 0: [Initial setting] A binary number of I bits is chosen by the random number generator. All the N sequences of I bits in the population have the same binary number (the population is homogeneous).

Step 1: [Selection] In proportion to fitness values ($1 + s = 2$ for the advantageous sequence and 1 for all the other sequences), select N sequences randomly from the whole population and create their copies, which compose a population of sequences for next generation (the standard roulette selection).

Step 2: [Mutation] Choose a bit randomly out of all sequences and flip it. This modification is repeated so that the number of binary loci subject to modifications might be approximately equal to NIu.

Step 3: [Crossover] Choose Nr sequences out of the population and pair them randomly. Between each pair, recombine binary strings at the rate that the probability for cross-point to occur per interbit gap equal to c. Then, return to Step 1.

Note that $N_e = N$ in this experiment owing to the standard roulette selection in Step 1. The generation cycle (recursion of Steps 1~3) was repeated until the mean population fitness exceeded 1.5, i.e., until the frequency of the advantageous sequence exceeds 0.5. (This is the definition of ‘domination’ in this experiment. See Chapter 1 for the background of the choice of this threshold frequency.) The program for the simulation was written in the ANSI-C language and run on a desk-top MIMD computer, Parsytec XPLORER with sixteen PowerPC’s (80MHz). For each set of parameters, numerical trials were conducted fifteen times and the mean value was calculated.

Results

Figure 2.2a shows the results for T_d and $T_v + T_c$ as a function of c in FPAP mode, and Fig. 2.2b shows those as a function of r in MPIP mode. The other parameters were $I = 20$, $N = N_e = 4096$, and $u = 0.002$. The lines with dots are the results from the simulation by GAs, and the curved solid lines are the results from the theoretical estimation using the formula for large β (note that $\beta = 4096 \times 0.002 \approx 8$ is much larger than one). The results obtained from the two different methods agree well with each other. They are qualitatively the same as results of the simulation with the vector representation, developed in Chapter 1. These results suggest that T_d is the minimum at an intermediate optimum crossover rate. This can be explained as follows. Since both mutation and crossover are randomization processes to create novel sequences, when the mutation rate is not as high as to completely randomize the population, crossover helps to create novel sequences of bits and reduces T_c. In Fig. 2.2 both T_d and $T_v + T_c$ decrease with c (or r) in the region of low c (or r). If c (or r) is too high, on the other hand, the created advantageous sequences are destroyed by crossover and cannot spread in the population. This increases the creation number N_d and brings about an conspicuous increase in T_d. In the region of higher c (or r) in Fig. 2.2, $T_v + T_c$ does not increase with c (or r), whereas T_d increases markedly.

Figures 2.3~2.9 show the acceleration rate by crossover A_{cross} defined by Eq. (2.2) as a function of the mutation rate u. The crossover rate were $c = 0.015$ and $r = 1$ for FPAP-mode (Figs. 2.3, 2.4, 2.5, 2.6a, 2.7, 2.8a, and 2.9a), and $c = 0.5$ and $r = 0.2$ for MPIP-mode (Figs. 2.6b, 2.8b, and 2.9b). In all these figures, the thick solid lines are the results from the mathematical formulas for large β, the fine solid lines are those from the mathematical formulas for small β, the lines with black dots are those from the GA simulation, and the lines with white dots are those from the simulation using the vector representation of the population.

In Figs. 2.3~2.5, the epistatic number was $I = 12$ and the (effective) population size was $N = N_e \approx 500$ (Fig. 2.3), $N = N_e \approx 2000$ (Fig. 2.4), or $N = N_e = 10000$ (Fig. 2.5), respectively. In each figure, the results obtained from different methods are qualitatively similar. The results from the simulation (shown in lines with dots in the figures) are given only for larger u, because for very low u, the computational load of

the simulation was so large that the experiments cannot be completed within a practical simulation time. In Fig. 2.6, theoretical results are shown for different values of the (effective) population sizes.

According to Figs. 2.3~2.6, we can conclude that A_{cross} is the maximum at an intermediate mutation rate u. This result is explained as follows. When u is very low, the genetic diversity is so small that crossover which can create novel sequences by stored diversity cannot be effective. When u is sufficiently high, on the other hand, the population is fully randomized by mutation, and then crossover cannot enhance creation of novel sequences further. This also decreases A_{cross} in the region of higher u. Hence there is an intermediate value of mutation u that makes crossover the most effective.

According to Figs. 2.3, 2.4, 2.5, and 2.6, the analytical estimation assuming small β suggests that A_{cross} does not necessarily depend monotonically on u for low u and has a second peak in the region of lower u. Although we could not confirm this result with the simulation (because of the limitation of computational time), this might be a real phenomena by the following reason. The reduction of A_{cross} for a low u is caused by the reduction in randomization by crossover due to the reduction in the population diversity. However, a low u also reduces the randomization by direct effect of mutation itself. Since A_{cross} is determined by the balance between randomization by mutation and crossover, if the creation of the advantageous sequence by mutation is suppressed more strongly than by crossover, this increases A_{cross} and can make the second peak of A_{cross}.

According to Fig. 2.6, when the population size N is not very large, the theoretical results obtained from the two different analyses (large-β analysis and small-β analysis) are similar, but they differ considerably for larger N ($N = 10000$ or 50000). We can explain this in the following way. As discussed above, the decrease in A_{cross} with u in the region of large u comes from the reduced of H, or in other words, because q_is tend to near 0 or 1. In the analysis for large β, however, continuous distribution of vector $\{q_i\}$ is assumed and this effect is not fully considered. This causes the underestimation of A_{cross}. Actually, the simulation results shown in Fig. 2.5 quantitatively agree with the results from the small-β analysis. When we use the value of A_{cross} estimated from the large-β analysis, we need to note this underestimation for low β.

Figures 2.7~2.9 show the results for larger epistatic numbers I ($I = 20$ for Figs. 2.7 and 2.8 and $I = 40$ for Fig. 2.9). According to these figures, we can conclude that the

value of A_{cross} *geometrically* increases with I. Crossover greatly enhances the creation of advantageous schemata when their orders are fairly large.

Finally we point out that two modes of crossover, FPAP-mode and MPIP-mode did not differ much. According to the results shown in Figs. 2.6, 2.8, and 2.9, the two modes of crossover did not make any significant difference in the acceleration effect and the evolutionary outcome.

Discussion

We studied the performance of GAs and the acceleration effect by crossover under a conspicuously discontinuous (Babel-like) fitness function in which only one sequence is much more advantageous than the others. The estimation was made with three different methods, and the results obtained from those methods demonstrated that crossover clearly accelerates evolution when its rate is not very high and the mutation rate u has an intermediate optimum value. The key point responsible for this result is the fitness landscape assumed in the model. In this landscape, evolutionary speed is primarily determined by the creation rate of an advantageous schema. Both mutation and crossover can enhance this creation process; and, especially when the mutation rate is not very high, crossover can significantly promote the speed of evolution.

Traditionally the theory of GAs has been centered upon the 'building block hypothesis' (BBH) (Holland, 1992) which states that the final solution is achieved through the combination of good component schemata. Many good schemata which are candidates for components of the final solution were considered to spread in 'implicit parallelism' (Goldberg, 1989), and the crossover operation in GAs was regarded to help to combine those component schemata into a single individual to create a more advantageous sequence (Fig. 2.10a). However, when the fitness landscape is discontinuous as assumed in the present chapter, evolution proceeds according to the picture illustrated in Fig. 2.10b, rather than the one shown in Fig. 2.10a. The BBH is valid, if it only asserts that the final solution is made of component good schemata. However, in this picture, creation and domination of advantageous component schemata take place not in parallel but serially (one by one), and the major roles of crossover are not combining advantageous schemata but the following two:

- ¶[Creation]: randomizing sequences and helping to create a novel advantageous schema,
- ¶[Preservation]: maintaining advantageous schemata which has already been fixed in the population.

The effect of crossover on the creation of schemata is essential to the acceleration effect. However this has often been overpassed in the GA studies. For example, a discontinuous fitness landscape has been studied using the royal road function. Mitchell

et al. (1991) studied the GA performance focusing on the time necessary for smaller-order schemata to combine to form the larger-order schema, but they did not examine the influence of crossover on the creation time of each schema. Although hitchhiking (which they called ‘premature convergence’) made them doubt the implicit parallelism, they sticked to the conventional notion about GAs (shown in Fig. 2.10a), without examining the waiting time until creation of advantageous schemata. Recently, however, Nimwegen *et al.* (1996, 1997) threw a new light on the royal road GA studies. Inspired by the study by Shapiro *et al.* (Prugel-Bennett & Shapiro, 1994, 1996; Rattray & Shapiro, 1996), Nimwegen *et al.* expressed a population using a fitness distribution and analyzed the dynamics of the royal road GA. Their mathematical analysis was restricted to GAs without crossover. However, by conducting a simulation incorporating crossover, they conjectured the same mechanism of GAs as shown in Fig. 2.10b. The Babel-like function is a special case of the royal road function, a case in which the number of block is limited to one. The present study, which has quantitatively estimated the advantage of crossover, can be considered an extension of Nimwegen *et al.* to the case where GAs include crossover.

Crossover has a preservation function. Acceleration of creation process is achieved by mutation as well as by crossover. The main difference between them comes from preservation by crossover. Mutation, which is a blind substitution of bits, destroys the established schemata, whereas crossover, which has no effect on the homogenous regions of the sequence, can preserve advantageous schemata once they are fixed in the population. Because of this blindness of mutation, the mutation rate must be kept very low if a large number of good schemata are to be maintained in the population. Destructive effect of mutation is called ‘error catastrophe’ problem (see Chapter 1 or Eigen, Gardiner, Schuster, & Winkler, 1981), and it is concluded that the mutation rate must be kept low, inversely proportional to the total order of good schemata. Crossover can be a beneficial process in such a case. A population which has once accumulated many functional schemata can evolve with creation by crossover which selectively randomizes only nonfunctional regions of bit sequences without destroying functional schemata.

These roles of crossover are the key points in application of GAs to the problem domain with a discontinuous fitness landscape. The implementations for genetic

parameters in these problems are summarized as follows:

¶The crossover rate should not be too high nor too low for fast evolution.

¶The mutation rate must be adjusted to a moderate value to enhance evolutionary acceleration by crossover.

¶To achieve a large acceleration effect by crossover, the order of the advantageous schemata to be created needs to be sufficiently large.

Although these rules are not applicable to all the engineering problems for which GAs are used today, there exists an important class of problems which has a discontinuous fitness function as assumed here. Genetic programming (GP) using machine language architecture is one such problem. In machine language GP, a program is represented by a very long bit sequence, and a functional set of machine codes corresponds to an advantageous schema. As a consequence, evolution typically proceeds with the discontinuous picture, and the use of GAs can be a powerful strategy for developing the optimal program in such systems. In Suzuki (1996, 1997), GAs were used to create an advantageous function in the program memory and crossover evidently contributed to the acceleration of creation process. Figure 7 in Suzuki (1997) qualitatively agrees well with Figs. 2.3~2.8.

We finish this discussion by pointing out the biological meaning of the above results. In this chapter, we examined the advantage of crossover focusing on the waiting time until creation and domination of a highly advantageous schema. In biology, this is interpreted as the effect of genetic recombination which induces creation of an advantageous combination of genes. In the living world, we can think of many examples of coadaptive combinations of genes; an advantageous function such a metabolic cycle, an advantageous anatomical structure such as eyes, and an advantageous behavior which requires the improvements of many elements. Such functions enjoy highly selective advantage when all component genes are present in the same individual, and the fitness landscape is considered to be conspicuously discontinuous in the DNA regions coding these functions. Recombination can promote creation of such a very advantageous combination of genes which causes drastic adaptive evolution.

However, above argument is not directly applicable to the explanation of the widespread occurrence of sexuality in real species. The shown advantage of

recombination is a long term one and is easily overwhelmed by the two-fold cost of sex when a mutation favoring parthenogenesis occurs in a population. In a biological context, this problem has been an enigma over the decades and many evolutionary biologists have made various arguments searching for a short term advantage of recombination (Crow & Kimura, 1965; Williams, 1975; Maynard Smith, 1971, 1978; Lloyd, 1980; Michod & Levin, 1988; Feldman, Otto, & Christiansen, 1997). Modern arguments on this problem are made using modifier models which have a special locus controlling the recombination rate between the major selected loci (Nei, 1967, 1969; Feldman, 1972; Felsenstein & Yokoyama, 1976; Feldman, Christiansen, & Brooks, 1980; Zhivotovsky, Feldman, & Christiansen, 1994; Bergman, Otto, & Feldman, 1995a, 1995b). Two different theories have recently attracted a great deal of attention; the deleterious mutation theory (Muller, 1964; Haigh, 1978; Kondrashov, 1988, 1994; Redfield, 1994; Zeyl & Bell, 1997) and the parasite theory (Jayakar, 1970; Jaenike, 1978; Hamilton, 1980; Bremermann, 1980; Tooby, 1982; Bell & Maynard Smith, 1987; Hamilton, Axelrod, & Tanese, 1990). Among them, the parasite theory has close relation to the present study. Based upon the Red Queen hypothesis, this theory states that sex is favored when a species is subject to attack by pathogens or parasites and needs to evolve under the fluctuating environment due to the host-parasite competition. This theory was tested by using the modifier models with cyclic selection (Charlesworth, 1993; Andreasen & Christiansen, 1995), and it was shown that advantage of recombination given by the host-parasite cycle is not decisive but delicate. This theory remains controversial in the biological community. The creation of an advantageous combination of genes which stands up to attack by parasites is essentially the same as the creation of an advantageous schema causing drastic adaptive evolution. The present quantitative result about the creation process by recombination may be of use for examining the parasite theory.

Appendix E: FORMULAS FOR $p_{\mathrm{C}}^{(\mathrm{mut.})}$ AND $p_{\mathrm{C}}^{(\mathrm{cross.})}$

Let x_I be the frequency of the advantageous sequence. To calculate $p_{\mathrm{C}}^{(\mathrm{mut.})}$, we first estimate the expected value of $\Delta x_I^{(\mathrm{mut.})}$, namely the differential of x_I by mutation per generation. In Chapter 1 Δx_I was calculated assuming that all q_is have the same value q around the stationary distribution. Here we derive more general expressions with the hope that the analyses holds for the situations wherein the heterozygosity H is smaller than 0.5.

We consider the elementary process of mutation to be divided into the following steps: first we randomly choose a sequence from the population and substitute it for the site A. Next, we flip bits in the sequence with the probability u, and then we put the sequence back into the population. Let A_{bef} be the event that a sequence chosen for site A happens to be the unique advantageous sequence before mutation, and A_{aft} be the event that the sequence at site A is advantageous after mutation. Hereafter we use the following notations for probability; $P(X)$ is the probability for event X to occur, and $P(X|Y)$ or $P_Y(X)$ are the conditional probability for event X to occur on the condition of event Y. Then, $\Delta x_I^{(\mathrm{mut.})}$ is written as

$$\Delta x_I^{(\mathrm{mut.})} = [P(A_{\mathrm{aft}}|A_{\mathrm{bef}})x_I + P(A_{\mathrm{aft}}|\overline{A_{\mathrm{bef}}})(1-x_I)] - x_I$$

$$= \frac{P(\overline{A_{\mathrm{bef}}} \wedge A_{\mathrm{aft}})}{P(\overline{A_{\mathrm{bef}}})}(1-x_I) - \{1 - P(A_{\mathrm{aft}}|A_{\mathrm{bef}})\}x_I$$

$$= \frac{P(A_{\mathrm{aft}}) - P(A_{\mathrm{bef}} \wedge A_{\mathrm{aft}})}{P(\overline{A_{\mathrm{bef}}})}(1-x_I) - \{1 - P(A_{\mathrm{aft}}|A_{\mathrm{bef}})\}x_I$$

$$= \frac{P(A_{\mathrm{aft}}) - P(A_{\mathrm{bef}})P(A_{\mathrm{aft}}|A_{\mathrm{bef}})}{1 - P(A_{\mathrm{bef}})}(1-x_I) - \{1 - P(A_{\mathrm{aft}}|A_{\mathrm{bef}})\}x_I$$

By substituting $P(A_{\mathrm{aft}}) = \prod\{(1-u)q_i + u(1-q_i)\}$, $P(A_{\mathrm{bef}}) = \prod q_i$ (these are valid due to the assumption of independent distribution at each locus), and $P(A_{\mathrm{aft}}|A_{\mathrm{bef}}) = (1-u)^I$, we obtain the formula for $\Delta x_I^{(\mathrm{mut.})}$ as

$$\Delta x_I^{(\mathrm{mut.})} = \frac{\prod\{(1-u)q_i + u(1-q_i)\} - (\prod q_i)(1-u)^I}{1 - \prod q_i}(1-x_I) - \{1-(1-u)^I\}x_I$$

$$= \frac{(\prod\{u+(1-2u)q_i\} - \prod q_i)(1-x_I) + \{1-(1-u)^I\}(\prod q_i - x_I)}{1-\prod q_i} \quad . \tag{E1}$$

Then $p_c^{(\mathrm{mut.})}$, the creation probabilities by mutation, is formulated as $\Delta x_I^{(\mathrm{mut.})}$ substituted with $x_I = 0$;

$$p_c^{(\mathrm{mut.})} \approx \prod\{u+(1-2u)q_i\} - (1-u)^I \prod q_i \quad , \tag{E2}$$

where the denominator factor $1-\prod q_i$ was neglected because $1-\prod q_i \approx 1$ holds true before creation of the advantageous sequence.

The basic strategy to calculate $p_c^{(\mathrm{cross.})}$ is the same as the one for $p_c^{(\mathrm{mut.})}$. We first formulate the expected value of $\Delta x_I^{(\mathrm{cross.})}$ (i.e., the differential of x_I by crossover per generation). Let the elementary process of crossover be divided into the three steps: first we randomly choose two sequences from the population and substitute them for sites A and B, next we recombine these two sequences, and finally we put them back into the population. Let A_{bef} and B_{bef} be the events that original sequences chosen for site A and B are advantageous before crossover respectively, and A_{aft} and B_{aft} denote the events that the sequences on site A and B are advantageous after crossover respectively. With these symbols, we define the following probabilities:

$$P_{\mathrm{an}} \equiv P(A_{\mathrm{aft}} \wedge \overline{B_{\mathrm{aft}}} \,|\, A_{\mathrm{bef}} \wedge \overline{B_{\mathrm{bef}}}),$$

$$P_{\mathrm{na}} \equiv P(A_{\mathrm{aft}} \wedge \overline{B_{\mathrm{aft}}} \,|\, \overline{A_{\mathrm{bef}}} \wedge B_{\mathrm{bef}}),$$

$$P_{\mathrm{nn}} \equiv P(A_{\mathrm{aft}} \wedge \overline{B_{\mathrm{aft}}} \,|\, \overline{A_{\mathrm{bef}}} \wedge \overline{B_{\mathrm{bef}}}).$$

Then, $\Delta x_I^{(\mathrm{cross.})}$ is written as

$$\begin{aligned} \Delta x_I^{(\mathrm{cross.})} &= r[x_I^2 + (P_{\mathrm{an}} + P_{\mathrm{na}})x_I(1-x_I) + P_{\mathrm{nn}}(1-x_I)^2] - rx_I \\ &= r(1-x_I)[P_{\mathrm{nn}}(1-x_I) - (1-P_{\mathrm{an}} - P_{\mathrm{na}})x_I] \quad . \end{aligned} \tag{E3}$$

First we transform P_{nn}.

$$P_{\mathrm{nn}} = P(A_{\mathrm{aft}} \wedge \overline{B_{\mathrm{aft}}} \,|\, \overline{A_{\mathrm{bef}}} \wedge \overline{B_{\mathrm{bef}}})$$

$$= \frac{P(A_{\mathrm{aft}} \wedge \overline{A_{\mathrm{bef}}} \wedge \overline{B_{\mathrm{bef}}})}{P(\overline{A_{\mathrm{bef}}} \wedge \overline{B_{\mathrm{bef}}})}$$

$$= \frac{P(A_{\text{aft}} \wedge \overline{A_{\text{bef}}}) - P(A_{\text{aft}} \wedge \overline{A_{\text{bef}}} \wedge B_{\text{bef}})}{P(\overline{A_{\text{bef}}})P(\overline{B_{\text{bef}}})}$$

$$= \frac{P(A_{\text{aft}}) - P(A_{\text{aft}} \wedge A_{\text{bef}}) - P(A_{\text{aft}} \wedge B_{\text{bef}}) + P(A_{\text{aft}} \wedge A_{\text{bef}} \wedge B_{\text{be}}}{(1 - P(A_{\text{bef}}))(1 - P(B_{\text{bef}}))}$$

$$= \frac{P(A_{\text{aft}}) - P(A_{\text{aft}} \wedge A_{\text{bef}}) - P(B_{\text{aft}} \wedge A_{\text{bef}}) + P(A_{\text{bef}} \wedge B_{\text{bef}}}{(1 - P(A_{\text{bef}}))(1 - P(B_{\text{bef}}))}$$

$$= \frac{P(A_{\text{aft}}) - P(A_{\text{bef}})P(A_{\text{aft}}|A_{\text{bef}}) - P(A_{\text{bef}})P(B_{\text{aft}}|A_{\text{bef}}) + P(A_{\text{bef}})P(B_{\text{bef}}|A_{\text{be}}}{(1 - P(A_{\text{bef}}))(1 - P(B_{\text{bef}}))}.$$

Since crossover does not change the values of q_is which determines the probability of a randomly chosen sequence being an advantageous one, $P(A_{\text{aft}}) = P(A_{\text{bef}}) = \prod_{i=1}^{I} q_i$. Therefore,

$$P_{\text{nn}} = \frac{P(A_{\text{bef}})[1 - P(A_{\text{aft}}|A_{\text{bef}}) - P(B_{\text{aft}}|A_{\text{bef}}) + P(A_{\text{aft}} \wedge B_{\text{aft}}|A_{\text{bef}})}{(1 - P(A_{\text{bef}}))(1 - P(B_{\text{bef}}))}$$

$$= \frac{P(A_{\text{bef}})[1 - P(A_{\text{aft}} \vee B_{\text{aft}}|A_{\text{bef}})]}{(1 - P(A_{\text{bef}}))(1 - P(B_{\text{bef}}))}$$

$$= \frac{P(A_{\text{bef}})P(\overline{A_{\text{aft}}} \wedge \overline{B_{\text{aft}}}|A_{\text{bef}})}{(1 - P(A_{\text{bef}}))(1 - P(B_{\text{bef}}))}$$

$$= \frac{(\prod q_i)P_{\text{des}}}{(1 - \prod q_i)^2}, \tag{E4}$$

where we defined

$$P_{\text{des}} \equiv P(\overline{A_{\text{aft}}} \wedge \overline{B_{\text{aft}}}|A_{\text{bef}}). \tag{E5}$$

P_{des} is the probability of an advantageous sequence being destroyed by crossover with a randomly chosen sequence. Similarly, $1 - P_{\text{an}} - P_{\text{na}}$ is transformed as

$$1 - P_{\text{an}} - P_{\text{na}} = 1 - P(A_{\text{aft}} \wedge \overline{B_{\text{aft}}}|A_{\text{bef}} \wedge \overline{B_{\text{bef}}}) - P(A_{\text{aft}} \wedge \overline{B_{\text{aft}}}|\overline{A_{\text{bef}}} \wedge B_{\text{bef}}$$

$$= 1 - P(A_{\text{aft}} \wedge \overline{B_{\text{aft}}}|A_{\text{bef}} \wedge \overline{B_{\text{bef}}}) - P(\overline{A_{\text{aft}}} \wedge B_{\text{aft}}|A_{\text{bef}} \wedge \overline{B_{\text{be}}}$$

$$= P(\overline{A_{\text{aft}}} \wedge \overline{B_{\text{aft}}}|A_{\text{bef}} \wedge \overline{B_{\text{bef}}}) + P(A_{\text{aft}} \wedge B_{\text{aft}}|A_{\text{bef}} \wedge \overline{B_{\text{be}}}$$

$$= \frac{P(\overline{A_{\text{aft}}} \wedge \overline{B_{\text{aft}}} \wedge A_{\text{bef}} \wedge \overline{B_{\text{bef}}})}{P(A_{\text{bef}} \wedge \overline{B_{\text{bef}}})}$$

$$= \frac{P(\overline{A_{\mathrm{aft}}} \wedge \overline{B_{\mathrm{aft}}} \wedge A_{\mathrm{bef}}) - P(\overline{A_{\mathrm{aft}}} \wedge \overline{B_{\mathrm{aft}}} \wedge A_{\mathrm{bef}} \wedge B_{\mathrm{bef}})}{P(A_{\mathrm{bef}}) - P(A_{\mathrm{bef}} \wedge B_{\mathrm{bef}})}$$

$$= \frac{P(\overline{A_{\mathrm{aft}}} \wedge \overline{B_{\mathrm{aft}}} \wedge A_{\mathrm{bef}})}{P(A_{\mathrm{bef}})(1 - P(B_{\mathrm{bef}}))}$$

$$= \frac{P(\overline{A_{\mathrm{aft}}} \wedge \overline{B_{\mathrm{aft}}} \,|\, A_{\mathrm{bef}})}{1 - P(B_{\mathrm{bef}})}$$

$$= \frac{P_{\mathrm{des}}}{1 - \prod q_i}. \tag{E6}$$

Substituting Eqs. (E4) and (E6) for Eq. (E3), the formula for $\Delta x_I^{(\mathrm{cross.})}$ is derived as

$$\Delta x_I^{(\mathrm{cross.})} = \frac{r \cdot P_{\mathrm{des}}(1 - x_I)(\prod q_i - x_I)}{(1 - \prod q_i)^2}. \tag{E7}$$

Then $p_{\mathrm{c}}^{(\mathrm{cross.})}$, the creation probabilities by crossover, is formulated as $\Delta x_I^{(\mathrm{cross.})}$ substituted with $x_I = 0$;

$$p_{\mathrm{c}}^{(\mathrm{cross.})} \approx r \cdot P_{\mathrm{des}} \prod q_i, \tag{E8}$$

where the denominator factor $1 - \prod q_i$ was neglected as in the case of mutation.

The detailed formula for P_{des} is derived as follows. From Eq. (E5),

$$P_{\mathrm{des}} = P_{A_{\mathrm{bef}}}(\overline{A_{\mathrm{aft}}} \wedge \overline{B_{\mathrm{aft}}})$$

$$= 1 - P_{A_{\mathrm{bef}}}(A_{\mathrm{aft}} \vee B_{\mathrm{aft}})$$

$$= 1 + P_{A_{\mathrm{bef}}}(A_{\mathrm{aft}} \wedge B_{\mathrm{aft}}) - P_{A_{\mathrm{bef}}}(A_{\mathrm{aft}}) - P_{A_{\mathrm{bef}}}(B_{\mathrm{aft}})$$

$$= 1 + P_{A_{\mathrm{bef}}}(B_{\mathrm{bef}}) - P_{A_{\mathrm{bef}}}(A_{\mathrm{aft}}) - P_{A_{\mathrm{bef}}}(B_{\mathrm{aft}})$$

Expressions for $P_{A_{\mathrm{bef}}}(A_{\mathrm{aft}})$ and $P_{A_{\mathrm{bef}}}(B_{\mathrm{aft}})$ can be derived from the recursion method established in Appendix D. Let U_i and E_i denote the events that i-th bit pair is unexchanged and exchanged by crossover respectively, and A_i and B_i denote the events that leftmost (earlier) i bits on the site A and B are all 1 after crossover respectively. (Note that A_{aft} and B_{aft} coincide with A_I and B_I respectively.) On the assumption of independent allele distribution at each locus, we can set up a recursion formula for the vector $\left[P_{A_{\mathrm{bef}}}(U_i \wedge A_i)\ P_{A_{\mathrm{bef}}}(E_i \wedge A_i)\right]$ as

$$\left[P_{A_{\text{bef}}}(U_i \wedge A_i)\ P_{A_{\text{bef}}}(E_i \wedge A_i)\right] = \left[P_{A_{\text{bef}}}(U_{i-1} \wedge A_{i-1})\ P_{A_{\text{bef}}}(E_{i-1} \wedge A_{i-1})\right] \begin{bmatrix} 1-c & cq_i \\ c & (1-c)q_i \end{bmatrix} \tag{E9}$$

Recursively using Eq. (E9) from the initial vector $\left[P_{A_{\text{bef}}}(U_0 \wedge A_0)\ P_{A_{\text{bef}}}(E_0 \wedge A_0)\right] = \left[1\ 0\right]$, we can write down a straightforward expression for $P_{A_{\text{bef}}}(U_I \wedge A_I)$ and $P_{A_{\text{bef}}}(E_I \wedge A_I)$ as

$$\left[P_{A_{\text{bef}}}(U_I \wedge A_I)\ P_{A_{\text{bef}}}(E_I \wedge A_I)\right] = \left[1\ 0\right] \begin{bmatrix} 1-c & cq_1 \\ c & (1-c)q_1 \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} 1-c & cq_I \\ c & (1-c)q_I \end{bmatrix}.$$

Similarly, the formula for the site B is derived as

$$\left[P_{A_{\text{bef}}}(U_I \wedge B_I)\ P_{A_{\text{bef}}}(E_I \wedge B_I)\right] = \left[1\ 0\right] \begin{bmatrix} (1-c)q_1 & c \\ cq_1 & 1-c \end{bmatrix} \cdot \ldots \cdot \begin{bmatrix} (1-c)q_I & c \\ cq_I & 1-c \end{bmatrix}.$$

By using these formulas, a detailed expression for P_{des} is derived as

$$\begin{aligned} P_{\text{des}} &= 1 + P(B_{\text{bef}}) - P_{A_{\text{bef}}}(A_I) - P_{A_{\text{bef}}}(B_I) \\ &= 1 + \prod q_i - (P_{A_{\text{bef}}}(U_I \wedge A_I) + P_{A_{\text{bef}}}(E_I \wedge A_I) + P_{A_{\text{bef}}}(U_I \wedge B_I) + P_{A_{\text{bef}}}(E_I \wedge B_I)) \\ &= 1 + \prod q_i - \left[1\ 0\right] \begin{bmatrix} 1-c & cq_1 \\ c & (1-c)q_1 \end{bmatrix} \cdots \begin{bmatrix} 1-c & cq_I \\ c & (1-c)q_I \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &\qquad - \left[1\ 0\right] \begin{bmatrix} (1-c)q_1 & c \\ cq_1 & 1-c \end{bmatrix} \cdots \begin{bmatrix} (1-c)q_I & c \\ cq_I & 1-c \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned} \tag{E10}$$

Appendix F: NUMERICAL EVALUATION OF $\langle p_{\rm c}^{(\rm mut.)}\rangle_{\rm BC}$ AND $\langle p_{\rm c}^{(\rm cross.)}\rangle_{\rm BC}$

In order to formulate $\langle p_{\rm c}^{(\rm cross.)}\rangle_{\rm BC}$ in an appropriate form for evaluation, we first derive a polynomial expression for $P_{\rm des}$. Hereafter, we express all q_is by a parameter q. Under this notation, we consider q^k to be the product of k different q_is because every term in the right-hand side of Eq. (E10) includes q_i only once (namely the degree of q_i is not larger than one). Although a polynomial expression for $P_{\rm des}$ can be derived from the solution of the recurrence equation (E9) as well (Chapter 1, Appendix D), here we express $P_{A_{\rm bef}}(U_I \wedge A_I)$, $P_{A_{\rm bef}}(E_I \wedge A_I)$, $P_{A_{\rm bef}}(U_I \wedge B_I)$, and $P_{A_{\rm bef}}(E_I \wedge B_I)$ by polynomial formulas with a set of coefficient parameters and establish recurrence equations for those parameters. This method enables us to express $P_{\rm des}$ by a positive coefficient formula that is numerically evaluated without occurrence of cancellation.

Since in all terms of polynomial expressions for $P_{A_{\rm bef}}(U_I \wedge A_I)$ and $P_{A_{\rm bef}}(E_I \wedge A_I)$ the sum of degrees of $1-c$ and c is always I, we can write these probabilities as

$$P_{A_{\rm bef}}(U_I \wedge A_I) = \sum_{i=0}^{I}\sum_{j=0}^{I} g_{I,i,j}^{(UA)}(1-c)^j c^{I-j} q^i, \tag{F1a}$$

$$P_{A_{\rm bef}}(E_I \wedge A_I) = \sum_{i=0}^{I}\sum_{j=0}^{I} g_{I,i,j}^{(EA)}(1-c)^j c^{I-j} q^i. \tag{F1b}$$

Substituting these formulas for Eq. (E9),

$$\left[\sum_{i=0}^{I}\sum_{j=0}^{I} g_{I,i,j}^{(UA)}(1-c)^j c^{I-j} q^i,\ \sum_{i=0}^{I}\sum_{j=0}^{I} g_{I,i,j}^{(EA)}(1-c)^j c^{I-j} q^i\right]$$

$$= \left[\sum_{i=0}^{I-1}\sum_{j=0}^{I-1} g_{I-1,i,j}^{(UA)}(1-c)^j c^{I-1-j} q^i,\ \sum_{i=0}^{I-1}\sum_{j=0}^{I-1} g_{I-1,i,j}^{(EA)}(1-c)^j c^{I-1-j} q^i\right]\begin{bmatrix}1-c & cq\\ c & (1-c)q\end{bmatrix}$$

$$= \left[\sum_{i=0}^{I}\sum_{j=0}^{I} (g_{I-1,i,j-1}^{(UA)} + g_{I-1,i,j}^{(EA)})(1-c)^j c^{I-j} q^i,\right.$$

$$\left.\sum_{i=0}^{I}\sum_{j=0}^{I} (g_{I-1,i-1,j}^{(UA)} + g_{I-1,i-1,j-1}^{(EA)})(1-c)^j c^{I-j} q^i\right].$$

Comparing terms of the same degrees of $1-c$, c, and q, we can get the following recurrence equations for $g_{I,i,j}^{(UA)}$ and $g_{I,i,j}^{(EA)}$;

$$g^{(UA)}_{I,i,j} = g^{(UA)}_{I-1,i,j-1} + g^{(EA)}_{I-1,i,j}, \tag{F2a}$$

$$g^{(EA)}_{I,i,j} = g^{(UA)}_{I-1,i-1,j} + g^{(EA)}_{I-1,i-1,j-1}. \tag{F2b}$$

From $P_{A_{\mathrm{bef}}}(U_0 \wedge A_0) = 1$ and $P_{A_{\mathrm{bef}}}(E_0 \wedge A_0) = 0$, the initial values of $g^{(UA)}_{0,i,j}$ and $g^{(EA)}_{0,i,j}$ are given by

$$g^{(UA)}_{0,i,j} = 1 \text{ for } i = j = 0,\ g^{(UA)}_{0,i,j} = 0 \text{ otherwise, and} \tag{F3a}$$

$$g^{(EA)}_{0,i,j} = 0 \text{ for all } i\text{'s and } j\text{'s}. \tag{F3b}$$

To solve these simultaneous recurrence equations, we introduce dummy parameters k_{a}, k_{b}, k_{c}, and k_{d}, and write Eqs. (F2a) and (F2b) as

$$g^{(UA)}_{I,i,j} = g^{(UA)}_{I-1,i,j-1}k_{\mathrm{a}} + g^{(EA)}_{I-1,i,j}k_{\mathrm{b}} \tag{F4a}$$

$$g^{(EA)}_{I,i,j} = g^{(UA)}_{I-1,i-1,j}k_{\mathrm{c}} + g^{(EA)}_{I-1,i-1,j-1}k_{\mathrm{d}}. \tag{F4b}$$

Recursively substituting Eqs. (F4a) and (F4b) and finally substituting Eqs. (F3a) and (F3b), we can get

$$\begin{aligned} g^{(UA)}_{I,i,j} &= g^{(UA)}_{I-1,i,j-1}k_{\mathrm{a}} + g^{(EA)}_{I-1,i,j}k_{\mathrm{b}} \\ &= (g^{(UA)}_{I-2,i,j-2}k_{\mathrm{a}} + g^{(EA)}_{I-2,i,j-1}k_{\mathrm{b}})k_{\mathrm{a}} + (g^{(UA)}_{I-2,i-1,j}k_{\mathrm{c}} + g^{(EA)}_{I-2,i-1,j-1}k_{\mathrm{d}})k_{\mathrm{b}} \\ &= \cdots \\ &= f(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}})g^{(UA)}_{0,0,0}k_{\mathrm{a}}^{n_{\mathrm{a}}}k_{\mathrm{b}}^{n_{\mathrm{b}}}k_{\mathrm{c}}^{n_{\mathrm{c}}}k_{\mathrm{d}}^{n_{\mathrm{d}}}, \end{aligned} \tag{F5}$$

where $f(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}})$ is an integer coefficient denoting the number of terms with the same degrees of k's. First we focus on the relation between (i, j) and $(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}})$. $n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}$, and n_{d} satisfying Eq. (F5) are not free parameters but should satisfy relations

$$n_{\mathrm{a}} + n_{\mathrm{b}} + n_{\mathrm{c}} + n_{\mathrm{d}} = I, \tag{F6a}$$

$$n_{\mathrm{c}} + n_{\mathrm{d}} = i, \tag{F6b}$$

$$n_{\mathrm{a}} + n_{\mathrm{d}} = j, \text{ and} \tag{F6c}$$

$$n_{\mathrm{b}} = n_{\mathrm{c}}. \tag{F6d}$$

The Eq. (F6a) is obvious. Eq. (F6b)/(F6c) is due to the fact that when using Eqs. (F4a) or (F4b), i/j is decreased by the term including k_{c} or $k_{\mathrm{d}}/k_{\mathrm{a}}$ or k_{d}. To derive Eq. (F6d) we consider the order of dummy parameters k's. Although in Eq. (F5) we assumed the

commutative law about the product of k's, if we had not allowed that law, the term $k_{\mathrm{a}}^{n_{\mathrm{a}}}k_{\mathrm{b}}^{n_{\mathrm{b}}}k_{\mathrm{c}}^{n_{\mathrm{c}}}k_{\mathrm{d}}^{n_{\mathrm{d}}}$ should be substituted with the term

$$k_{\mathrm{a}}{}^{*}k_{\mathrm{c}}k_{\mathrm{d}}{}^{*}k_{\mathrm{b}}k_{\mathrm{a}}{}^{*}\cdots k_{\mathrm{a}}{}^{*}k_{\mathrm{c}}k_{\mathrm{d}}{}^{*}k_{\mathrm{b}}k_{\mathrm{a}}{}^{*} = k_{\mathrm{a}}{}^{*}(k_{\mathrm{c}}k_{\mathrm{d}}{}^{*}k_{\mathrm{b}}k_{\mathrm{a}}{}^{*})^{n_{\mathrm{b}}}\ , \tag{F7}$$

where the symbol '*' denotes any nonnegative integer. Eq. (F7) includes the same number of k_{b}'s and k_{c}'s so that Eq. (F6d) must be satisfied. From Eqs. (F6a)~(F6d), we can represent $n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}$, and n_{d} by I, i and j;

$$n_{\mathrm{a}} = \frac{I+j}{2} - i,\ n_{\mathrm{b}} = n_{\mathrm{c}} = \frac{I-j}{2},\ n_{\mathrm{d}} = \frac{-I+j}{2} + i.$$

Next we formulate the coefficient $f(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}})$. Since $f(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}})$ is the number of the disposition of k_{a}'s and k_{d}'s in Eq. (F7), if $n_{\mathrm{b}} > 0$ (i.e., $j < I$), we can formulate that by the product of the repeated combinations;

$$\begin{aligned} f(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}}) &= {}_{n_{\mathrm{b}}+1}H_{n_{\mathrm{a}}} \cdot {}_{n_{\mathrm{b}}}H_{n_{\mathrm{d}}} \\ &= \binom{n_{\mathrm{b}}+n_{\mathrm{a}}}{n_{\mathrm{a}}}\binom{n_{\mathrm{b}}+n_{\mathrm{d}}-1}{n_{\mathrm{d}}} \\ &= \binom{I-i}{\frac{I+j}{2}-i}\binom{i-1}{\frac{-I+j}{2}+i} \\ &= \binom{I-i}{\frac{I-j}{2}}\binom{i-1}{\frac{I-j}{2}-1}. \end{aligned}$$

If $n_{\mathrm{b}} = 0$ (i.e., $j = I$), on the other hand, the number of the disposition of k_{a}'s and k_{d}'s in Eq. (F7) is 1 if $n_{\mathrm{d}} = i = 0$ and 0 if $n_{\mathrm{d}} = i > 0$. Hence, combining them, $f(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}})$ is written as

$$f(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{c}}, n_{\mathrm{d}}) = \binom{I-i}{\frac{I-j}{2}}\binom{i-1}{\frac{I-j}{2}-1} + \delta_{i,0}\delta_{j,I}\,. \tag{F8}$$

The formula for $g_{I,i,j}^{(UA)}$ is given by Eqs. (F5) substituted with (F3a), (F8), and $k_{\mathrm{a}} = k_{\mathrm{b}} = k_{\mathrm{c}} = k_{\mathrm{d}} = 1$;

$$g_{I,i,j}^{(UA)} = \binom{I-i}{\frac{I-j}{2}}\binom{i-1}{\frac{I-j}{2}-1} + \delta_{i,0}\delta_{j,I}\,. \tag{F9a}$$

Similarly, the formula for $g_{I,i,j}^{(EA)}$ is derived as

$$g^{(EA)}_{I,i,j} = \binom{I-i}{\frac{I-j-1}{2}}\binom{i-1}{\frac{I-j-1}{2}} . \tag{F9b}$$

Now that solutions of the recurrence equations were obtained, we can write polynomial formulas for $P_{A_{\text{bef}}}(U_I \wedge A_I)$ and $P_{A_{\text{bef}}}(E_I \wedge A_I)$. Substituting Eqs. (F9a) and (F9b) for Eqs. (F1a) and (F1b), we get

$$P_{A_{\text{bef}}}(U_I \wedge A_I) = \sum_{i=0}^{I}\sum_{j=0}^{I}\binom{I-i}{\frac{I-j}{2}}\binom{i-1}{\frac{I-j}{2}-1}(1-c)^j c^{I-j} q^i + (1-c)^I , \tag{F10a}$$

$$P_{A_{\text{bef}}}(E_I \wedge A_I) = \sum_{i=0}^{I}\sum_{j=0}^{I}\binom{I-i}{\frac{I-j-1}{2}}\binom{i-1}{\frac{I-j-1}{2}}(1-c)^j c^{I-j} q^i . \tag{F10b}$$

For derivation of polynomial expressions for $P_{A_{\text{bef}}}(U_I \wedge B_I)$ and $P_{A_{\text{bef}}}(E_I \wedge B_I)$, we can make a similar argument. The results are

$$P_{A_{\text{bef}}}(U_I \wedge B_I) = \sum_{i=0}^{I}\sum_{j=0}^{I}\binom{i}{\frac{I-j}{2}}\binom{I-i-1}{\frac{I-j}{2}-1}(1-c)^j c^{I-j} q^i + (1-c)^I q^I , \tag{F10c}$$

$$P_{A_{\text{bef}}}(E_I \wedge B_I) = \sum_{i=0}^{I}\sum_{j=0}^{I}\binom{i}{\frac{I-j-1}{2}}\binom{I-i-1}{\frac{I-j-1}{2}}(1-c)^j c^{I-j} q^i . \tag{F10d}$$

Then the final polynomial formula for P_{des} is derived from Eq. (E10) substituted with Eqs. (F10a)~(F10d);

$$\begin{aligned} P_{\text{des}} &= \{1-(1-c)^I\}(1+q^I) \\ &\quad -\sum_{i=0}^{I}\sum_{j=0}^{I}\left[\binom{I-i}{\frac{I-j}{2}}\binom{i-1}{\frac{I-j}{2}-1}+\binom{I-i}{\frac{I-j-1}{2}}\binom{i-1}{\frac{I-j-1}{2}}\right. \\ &\qquad \left.+\binom{i}{\frac{I-j}{2}}\binom{I-i-1}{\frac{I-j}{2}-1}+\binom{i}{\frac{I-j-1}{2}}\binom{I-i-1}{\frac{I-j-1}{2}}\right](1-c)^j c^{I-j} q^i \\ &= \{1-(1-c)^I\}(1+q^I) \\ &\quad -\sum_{i=0}^{I}\sum_{j=0}^{I}\left[\binom{I-i}{\frac{I-j}{2}}\binom{i-1}{\frac{I-j}{2}-1}+\binom{I-i}{\frac{I-j-1}{2}}\binom{i-1}{\frac{I-j-1}{2}}\right](1-c)^j c^{I-j}(q^i+q^{I-i}) \\ &= \{1-(1-c)^I\}(1+q^I) - \sum_{j=0}^{[I/2]}\sum_{i=j}^{I-j}\Omega_{I,i,j}(c)(q^i+q^{I-i}) , \end{aligned} \tag{F11a}$$

where

$$\Omega_{I,i,j}(c) \equiv \left[\binom{i-1}{j-1} + \binom{i-1}{j}\frac{c}{1-c}\right]\binom{I-i}{j}(1-c)^{I-2j}c^{2j} \quad . \tag{F11b}$$

To derive the last formula, parameter conversions $j \to I-2j$ (for the former term) or $j \to I-2j-1$ (for the latter term) were used. [ ] in $[I/2]$ means the Gauss's notation.

With this formula and Eqs. (E2) and (E8), $\langle p_c^{(\mathrm{mut.})}\rangle_{\mathrm{BC}}$ and $\langle p_c^{(\mathrm{cross.})}\rangle_{\mathrm{BC}}$ are formulated as

$$\begin{aligned}\langle p_c^{(\mathrm{mut.})}\rangle_{\mathrm{BC}} &= \langle \prod\{u+(1-2u)q_i\} - (1-u)^I \prod q_i\rangle_{\mathrm{BC}} \\ &= \langle \{u+(1-2u)q\}^I - (1-u)^I q^I\rangle_{\mathrm{BC}} \\ &= \sum_{i=0}^{I}\binom{I}{i}u^{I-i}(1-2u)^i\langle q^i\rangle_{\mathrm{BC}} - (1-u)^I\langle q^I\rangle_{\mathrm{BC}} \quad ,\end{aligned} \tag{F12a}$$

$$\begin{aligned}\langle p_c^{(\mathrm{cross.})}\rangle_{\mathrm{BC}} &= \langle r\cdot P_{\mathrm{des}}\prod q_i\rangle_{\mathrm{BC}} \\ &= \langle r\cdot\left[\{1-(1-c)^I\}(1+q^I) - \sum_{j=0}^{[I/2]}\sum_{i=j}^{I-j}\Omega_{I,i,j}(c)(q^i+q^{I-i})\right]q^I\rangle_{\mathrm{BC}} \\ &= r\{1-(1-c)^I\}(\langle q^I\rangle_{\mathrm{BC}} + \langle q^Iq^I\rangle_{\mathrm{BC}}) - r\sum_{j=0}^{[I/2]}\sum_{i=j}^{I-j}\Omega_{I,i,j}(c)(\langle q^iq^I\rangle_{\mathrm{BC}} + \langle q^{I-i}q^I\rangle_{\mathrm{BC}}) \\ &\quad .\end{aligned} \tag{F12b}$$

q^i and q^I in Eqs. (F12a) are also considered to be the product of i and I different q_is respectively because in the expansion of $\prod\{u+(1-2u)q_i\}$ there is no term which includes q_i with degrees larger than 1.

$\langle q^i\rangle_{\mathrm{BC}}$ or $\langle q^iq^I\rangle_{\mathrm{BC}}$ in Eqs. (F12a) and (F12b) are calculated as follows. From Eqs. (2.8) and (2.10), the formula for $\langle q^i\rangle_{\mathrm{BC}}$ and $\langle q^iq^I\rangle_{\mathrm{BC}}$ is written as

$$\langle q^i(q^I)^\nu\rangle_{\mathrm{BC}} = \frac{\int_0^1\cdots\int_0^1 \exp(-N\prod q_i)q^i(q^I)^\nu\prod_i h(q_i)dq_i}{\int_0^1\cdots\int_0^1 \exp(-N\prod q_i)\prod_i h(q_i)dq_i} \quad , \ \nu = 0 \text{ or } 1. \tag{F13}$$

Hence, to evaluate $\langle q^i(q^I)^\nu\rangle_{\mathrm{BC}}$, we have to calculate

$$J(\nu, i) \equiv \int_0^1 \cdots \int_0^1 \exp(-N \prod q_i) q^i (q^I)^\nu \prod_i h(q_i) dq_i \ , \tag{F14}$$

where i is an integer parameter whose value is $i = 0, 1, \cdots, I$. $J(\nu, i)$ is transformed as

$$\begin{aligned} J(\nu, i) &= \sum_{k=0}^{\infty} \frac{(-N)^k}{k!} \int_0^1 \cdots \int_0^1 q^i \prod_{i'} q_{i'}^{k+\nu} h(q_{i'}) dq_{i'} \\ &= \sum_{k=0}^{\infty} \frac{(-N)^k}{k!} \left\{ \int_0^1 q^{k+\nu} h(q) dq \right\}^{I-i} \left\{ \int_0^1 q^{k+\nu+1} h(q) dq \right\}^i \\ &= \sum_{k=0}^{\infty} \frac{(-N)^k}{k!} \left(\frac{(2\beta)_{k+\nu}}{(4\beta)_{k+\nu}} \right)^{I-i} \left(\frac{(2\beta)_{k+\nu+1}}{(4\beta)_{k+\nu+1}} \right)^i \ . \end{aligned} \tag{F15}$$

$(a)_n$ is the Pochhammer symbol defined by

$$(a)_n \equiv a(a+1)\cdots(a+n-1) = \Gamma(a+n)/\Gamma(a) \ .$$

Since the direct summation of the series in Eq. (F15) gives rise to cancellation, we express Eq. (F15) with the modified generalized Hypergeometric Function and calculate it in the complex plane. Using the complex parameter sets (a_j) and (b_j), ***non-negative integer parameter sets*** (m_j) ***and*** (n_j), and a complex variable z, the modified generalized Hypergeometric Function is defined by

$${}_p\hat{F}_q\left(\frac{(a,m)_p}{(b,n)_q} \middle| z \right) \equiv \sum_{k=0}^{\infty} \frac{\prod_{j=1}^{p} (a_j)_{k+m_j}}{\prod_{j=1}^{q} (b_j)_{k+n_j}} \frac{z^k}{k!} \tag{F16}$$

where $(a,m)_p$ and $(b,n)_q$ are contracted notations interpreted as $((a_1, m_1), (a_2, m_2), \cdots, (a_p, m_p))$ and $((b_1, n_1), (b_2, n_2), \cdots, (b_q, n_q))$ respectively. Using this definition, $J(\nu, i)$ is formulated as

$$J(\nu, i) = {}_I\hat{F}_I\left(\frac{(2\beta, \nu)_{I-i} (2\beta, \nu+1)_i}{(4\beta, \nu)_{I-i} (4\beta, \nu+1)_i} \middle| (-N) \right) \ . \tag{F17}$$

To evaluate $\hat{F}$ for a large negative variable, we transform Eq. (F16) into the Mellin-Barnes type integral in the complex plane. Like standard theory for the generalized Hypergeometric Function (Luke, 1969; Mathai & Saxena, 1973; Prudnikov &

Marichev 1986), Eq. (F16) is transformed as

$$ {}_p\hat{F}_q\left(\frac{(a,m)_p}{(b,n)_q} \middle| z \right) = \frac{\prod_{j=1}^{q} \Gamma(b_j)}{\prod_{j=1}^{p} \Gamma(a_j)} \frac{1}{2\pi i} \int_{(\alpha - i\infty)}^{(\alpha + i\infty)} \frac{\prod_{j=1}^{p} \Gamma(a_j + m_j - w)}{\prod_{j=1}^{q} \Gamma(b_j + n_j - w)} \Gamma(w)(-z)^{-w} dw \quad . $$

The path in the right-hand side of this formula is taken in such a manner that the poles of $\Gamma(w)$ are separated from those of $\Gamma(a_j + m_j - w)$ for $j = 1, \cdots, p$. ***We can calculate*** $J(\nu, i)$ ***with this formula*** by numerically integrate the complex integral.

Appendix G: DERIVATION OF EQ. (2.12)

According to the standard population genetic method of branching process (Fisher, 1930; Ewens, 1979, p.22), the extinction probability z satisfies

$$z = p_0 \cdot 1 + p_1 \cdot z + \cdots = \sum_{k=0}^{\infty} p_k \cdot z^k, \tag{G1}$$

where p_k is the probability that the advantageous sequence bears k surviving offspring in the next generation. If k obeys the Poisson distribution given by

$$p_k = \frac{\mu^k}{k!} e^{-\mu}, \tag{G2}$$

Eq. (G1) is transformed as

$$z = \sum_{k=0}^{\infty} \left(\frac{\mu^k}{k!} e^{-\mu} \right) \cdot z^k = \exp(-\mu(1-z)) \ . \tag{G3}$$

Appendix H: EVALUATION OF $\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{\mathrm{AC}}$ AND $\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{\mathrm{AC}}$

We first formulate the detailed expression for $p_{\mathrm{d}}^{(\mathrm{mut.})}$ and $p_{\mathrm{d}}^{(\mathrm{cross.})}$. By using x_I (the frequency of the advantageous sequence), the destruction probability $p_{\mathrm{d}}^{(\mathrm{mut.})}$ ($p_{\mathrm{d}}^{(\mathrm{cross.})}$) is formulated as $(-\Delta x_I^{(\mathrm{mut.})})/x_I$ ($(-\Delta x_I^{(\mathrm{cross.})})/x_I$) substituted with $x_I = 1/N$. Hence, from Eqs. (E1) and (E7),

$$p_{\mathrm{d}}^{(\mathrm{mut.})} = -\frac{\Delta x_I^{(\mathrm{mut.})}}{x_I}$$

$$= (\prod\{u+(1-2u)q_i\} - \prod q_i)(1-N) + \{1-(1-u)^I\}(1-N\prod q_i)$$

, (H1a)

$$p_{\mathrm{d}}^{(\mathrm{cross.})} = -\frac{\Delta x_I^{(\mathrm{cross.})}}{x_I} = r\cdot P_{\mathrm{des}}\left(1-\frac{1}{N}\right)(1-N\prod q_i), \quad \text{(H1b)}$$

where the denominator factor $1-\prod q_i$ was omitted because $1-\prod q_i \approx 1$ holds true immediately after creation of the advantageous sequence.

By use of Eqs. (H1a), (H1b) and (F11a), we can derive the polynomial formulas for $\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{\mathrm{AC}}$ and $\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{\mathrm{AC}}$ as

$$\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{\mathrm{AC}}$$

$$= \langle(\prod\{u+(1-2u)q_i\} - \prod q_i)(1-N) + \{1-(1-u)^I\}(1-N\prod q_i)\rangle_{\mathrm{AC}}$$

$$= (1-N)\sum_{i=0}^{I}\binom{I}{i}u^{I-i}(1-2u)^i\langle q^i\rangle_{\mathrm{AC}} - \{1-N(1-u)^I\}\langle q^I\rangle_{\mathrm{AC}} + \{1-(1-u)^I\}$$

, (H2a)

$$\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{\mathrm{AC}}$$

$$\approx r\cdot\langle P_{\mathrm{des}}(1-N\prod q_i)\rangle_{\mathrm{AC}}$$

$$= r\cdot\left\langle\left[\{1-(1-c)^I\}(1+q^I) - \sum_{j=0}^{[I/2]}\sum_{i=j}^{I-j}\Omega_{I,i,j}(c)(q^i+q^{I-i})\right](1-Nq^I)\right\rangle_{\mathrm{AC}}$$

$$= r\{1-(1-c)^I\}\{1+(1-N)\langle q^I\rangle_{\mathrm{AC}} - N\langle q^I q^I\rangle_{\mathrm{AC}}\}$$

$$-r\sum_{j=0}^{[I/2]}\sum_{i=j}^{I-j}\Omega_{I,i,j}(c)\{\langle q^i\rangle_{\mathrm{AC}}+\langle q^{I-i}\rangle_{\mathrm{AC}}-N(\langle q^iq^I\rangle_{\mathrm{AC}}+\langle q^{I-i}q^I\rangle_{\mathrm{AC}})\}\quad , \qquad \text{(H2b)}$$

As in Appendix E, q^k in Eqs. (H2a) and (H2b) represents the product of k different q_is. The evaluation method for $\langle q^i\rangle_{\mathrm{AC}}$ and $\langle q^iq^I\rangle_{\mathrm{AC}}$ is quite similar to that for $\langle q^i\rangle_{\mathrm{BC}}$ and $\langle q^iq^I\rangle_{\mathrm{BC}}$ respectively. Combining Eqs. (2.14) and (2.15), $\langle q^i(q^I)^\nu\rangle_{\mathrm{AC}}$ ($\nu = 0$ or 1) is formulated as

$$\langle q^i(q^I)^\nu\rangle_{\mathrm{AC}} = \frac{\int_0^1\cdots\int_0^1\exp(-N\prod q_i)q^i(q^I)^{\nu+1}\prod_i h(q_i)dq_i}{\int_0^1\cdots\int_0^1\exp(-N\prod q_i)q^I\prod_i h(q_i)dq_i} . \qquad \text{(H3)}$$

The numerator and denominator factors are evaluated by calculating $J(\nu, i)$ (defined in Eq. (F14)) for $\nu = 0$, 1 and 2. Note that the numerical evaluation method for $J(\nu, i)$ (Appendix E) also holds true when $\nu = 2$.

Appendix I: EVALUATION OF $\langle p_{\rm c}^{(\rm mut.)}\rangle_{Q\approx i}$ AND $\langle p_{\rm c}^{(\rm cross.)}\rangle_{Q\approx i}$

Substituting Eqs. (E2), (E8), and (E10), $\langle p_{\rm c}^{(\rm mut.)}\rangle_{Q\approx i}$ and $\langle p_{\rm c}^{(\rm cross.)}\rangle_{Q\approx i}$ are transformed as

$$\begin{aligned}\langle p_{\rm c}^{(\rm mut.)}\rangle_{Q\approx i} &= \langle\prod\{u+(1-2u)q_i\}-(1-u)^I\prod q_i\rangle_{Q\approx i}\\ &= \{u+(1-2u)\langle q\rangle_0\}^{I-i}\{u+(1-2u)\langle q\rangle_1\}^i-(1-u)^I\langle q\rangle_0^{I-i}\langle q\rangle_1^i \quad ,\end{aligned} \tag{I1a}$$

$$\begin{aligned}\langle p_{\rm c}^{(\rm cross.)}\rangle_{Q\approx i} &= r\cdot\langle P_{\rm des}\prod q_i\rangle_{Q\approx}\\ &= r\langle\prod q_i\rangle_{Q\approx i}+r\langle\prod q_i^2\rangle_{Q\approx i}\\ &\quad -r\left\langle\begin{bmatrix}1&0\end{bmatrix}\begin{bmatrix}1-c & cq_1\\ c & (1-c)q_1\end{bmatrix}\cdots\begin{bmatrix}1-c & cq_I\\ c & (1-c)q_I\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\prod q_i\right\rangle_{Q\approx i}\\ &\quad -r\left\langle\begin{bmatrix}1&0\end{bmatrix}\begin{bmatrix}(1-c)q_1 & c\\ cq_1 & 1-c\end{bmatrix}\cdots\begin{bmatrix}(1-c)q_I & c\\ cq_I & 1-c\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\prod q_i\right\rangle_{Q\approx i}\\ &= r\langle q\rangle_0^{I-i}\langle q\rangle_1^i+r\langle q^2\rangle_0^{I-i}\langle q^2\rangle_1^i\\ &\quad -r\left\langle\begin{bmatrix}1&0\end{bmatrix}\begin{bmatrix}(1-c)\langle q\rangle_{\sigma_1} & c\langle q^2\rangle_{\sigma_1}\\ c\langle q\rangle_{\sigma_1} & (1-c)\langle q^2\rangle_{\sigma_1}\end{bmatrix}\cdots\begin{bmatrix}(1-c)\langle q\rangle_{\sigma_I} & c\langle q^2\rangle_{\sigma_I}\\ c\langle q\rangle_{\sigma_I} & (1-c)\langle q^2\rangle_{\sigma_I}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\right\rangle_{\sum\sigma_j=i}\\ &\quad -r\left\langle\begin{bmatrix}1&0\end{bmatrix}\begin{bmatrix}(1-c)\langle q^2\rangle_{\sigma_1} & c\langle q\rangle_{\sigma_1}\\ c\langle q^2\rangle_{\sigma_1} & (1-c)\langle q\rangle_{\sigma_1}\end{bmatrix}\cdots\begin{bmatrix}(1-c)\langle q^2\rangle_{\sigma_I} & c\langle q\rangle_{\sigma_I}\\ c\langle q^2\rangle_{\sigma_I} & (1-c)\langle q\rangle_{\sigma_I}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\right\rangle_{\sum\sigma_j=i},\end{aligned} \tag{I1b}$$

In above derivation, we used the assumption of independent distribution at each locus. $\langle\ \rangle_0$ and $\langle\ \rangle_1$ mean the average under the partial distribution around $q\approx 0$ and $q\approx 1$ respectively. These partial distributions are proportional to $q^{2\beta-1}$ and $(1-q)^{2\beta-1}$ respectively (see the definition of $h(q)$ in Eq. (2.5)), so that, considering the normalization factor, $\langle q^k\rangle_0$ and $\langle q^k\rangle_1$ are calculated from

$$\langle q^k \rangle_0 = \frac{\int_0^1 (q^k) q^{2\beta-1} dq}{\int_0^1 q^{2\beta-1} dq} = \frac{2\beta}{2\beta + k},$$

$$\langle q^k \rangle_1 = \frac{\int_0^1 (q^k)(1-q)^{2\beta-1} dq}{\int_0^1 (1-q)^{2\beta-1} dq} = \frac{k!}{(2\beta+1)_k} .$$

σ_i's in Eq. (I1b) are binary variables and $\langle\ \rangle_{\sum \sigma_j = i}$ means the average for various $(\sigma_1, \cdots, \sigma_I)$ s satisfying $\sum_{j=1}^{I} \sigma_j = i$. Although when I is large and i has an intermediate value around $I/2$, the number of $\{\sigma_i\}$ vectors satisfying the condition becomes extraordinarily large, we calculate an average by using only one thousand sampling vectors. We confirmed this limitation exerted no influence upon the results from some numerical experiments.

Appendix J: EVALUATION OF $\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{Q\approx i}$ AND $\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{Q\approx i}$

Substituting Eqs. (H1a), (H1b), and (E10), $\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{Q\approx i}$ and $\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{Q\approx i}$ are transformed as

$$\begin{aligned}
&\langle p_{\mathrm{d}}^{(\mathrm{mut.})}\rangle_{Q\approx i} \\
&\quad = \langle (\prod\{u+(1-2u)q_i\} - \prod q_i)(1-N) + \{1-(1-u)^I\}(1-N\prod q_i)\rangle_{Q\approx i} \\
&\quad = \{u+(1-2u)\langle q\rangle_0\}^{I-i}\{u+(1-2u)\langle q\rangle_1\}^{i}(1-N) \\
&\qquad -\{1-N(1-u)^I\}\langle q\rangle_0^{I-i}\langle q\rangle_1^{i} + \{1-(1-u)^I\} \quad , \qquad \text{(J1a)}
\end{aligned}$$

$$\begin{aligned}
&\langle p_{\mathrm{d}}^{(\mathrm{cross.})}\rangle_{Q\approx i} \\
&\quad \approx r\cdot\langle P_{\mathrm{des}}(1-N\prod q_i)\rangle_{Q\approx i} \\
&\quad = r\{1-(N-1)\langle q\rangle_0^{I-i}\langle q\rangle_1^{i} - N\langle q^2\rangle_0^{I-i}\langle q^2\rangle_1^{i}\} \\
&\qquad -r\langle \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}1-c & c\langle q\rangle_{\sigma_1}\\ c & (1-c)\langle q\rangle_{\sigma_1}\end{bmatrix}\cdots\begin{bmatrix}1-c & c\langle q\rangle_{\sigma_I}\\ c & (1-c)\langle q\rangle_{\sigma_I}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\rangle_{\sum\sigma_j=i} \\
&\qquad -r\langle \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}(1-c)\langle q\rangle_{\sigma_1} & c\\ c\langle q\rangle_{\sigma_1} & 1-c\end{bmatrix}\cdots\begin{bmatrix}(1-c)\langle q\rangle_{\sigma_I} & c\\ c\langle q\rangle_{\sigma_I} & 1-c\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\rangle_{\sum\sigma_j=i} \\
&\qquad +rN\langle \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}(1-c)\langle q\rangle_{\sigma_1} & c\langle q^2\rangle_{\sigma_1}\\ c\langle q\rangle_{\sigma_1} & (1-c)\langle q^2\rangle_{\sigma_1}\end{bmatrix}\cdots\begin{bmatrix}(1-c)\langle q\rangle_{\sigma_I} & c\langle q^2\rangle_{\sigma_I}\\ c\langle q\rangle_{\sigma_I} & (1-c)\langle q^2\rangle_{\sigma_I}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\rangle_{\sum\sigma_j=i} \\
&\qquad +rN\langle \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}(1-c)\langle q^2\rangle_{\sigma_1} & c\langle q\rangle_{\sigma_1}\\ c\langle q^2\rangle_{\sigma_1} & (1-c)\langle q\rangle_{\sigma_1}\end{bmatrix}\cdots\begin{bmatrix}(1-c)\langle q^2\rangle_{\sigma_I} & c\langle q\rangle_{\sigma_I}\\ c\langle q^2\rangle_{\sigma_I} & (1-c)\langle q\rangle_{\sigma_I}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\rangle_{\sum\sigma_j=i} \\
&\qquad , \qquad \text{(J1b)}
\end{aligned}$$

The definition of binary variables σ_i's, average operations $\langle\ \rangle_0$, $\langle\ \rangle_1$, and $\langle\ \rangle_{\sum\sigma_j=i}$

are the same as those in Appendix E. See Appendix E for detailed methods to evaluate those quantities.

Appendix K: MODIFIED METHOD TO CALCULATE CROSSOVER TENSOR (C_{jki})

C_{jki} is the transition probability from j-string to i-string after recombining with k-string, and was calculated by use of the recursion formula in Chapter 1 (Appendix B). We here only describe the modified point for calculating C_{jki} using the same notation as in Chapter 1.

Let b_j be the $h+1$-th bit on the original j-string and b_k be the $h+1$-th bit on the original k-string. In Chapter 1, the probability of b_j being 1 was formulated as

$$p_{jh}(j') \equiv P(b_j = 1) = \frac{j - j'}{I - h}. \tag{K1}$$

Although this equation is right under the assumption that $j - j'$ bits of 1's are equally distributed in remaining $I - h$ loci, when $H < 0.5$, this equation must be modified considering the occurrence probability of bit pairs. If we consider H as the probability of the heterogenous bit pair occurring and $1 - H$ as that of the homogeneous bit pair occurring, the formula for $p_{jh}(j')$ is modified as

$$\begin{aligned} p_{jh}(j') &= [\text{the sum of occurrence probabilities of bit distributions which satisfy } b_j = 1] / [\text{the sum of occurrence probabilities of all bit distributions}] \\ &= [\text{ditto}] / p(I - h,\ j - j',\ k - k',\ H), \end{aligned} \tag{K2}$$

where

$$\begin{aligned} p(x, y, z, H) \equiv\ & [\text{the sum of occurrence probabilities of all bit distributions in } x\text{-loci pair which includes } y \text{ 1's on one side and } z \text{ 1's on the other side}] \\ =\ & \sum_{l = \max(0,\ y + z - x)}^{\min(y,\ z)} \binom{x}{l}\binom{x - l}{y - l}\binom{x - y}{z - l} H^{y + z - 2l}(1 - H)^{x - y - z + 2l}. \end{aligned} \tag{K3}$$

The final expression in Eq. (K3) was derived by use of the parameter l defined as the number of pairs of bit 1's positioned at the same locus.

In order to get the modified recursion formula for $P_h(e', j', k', i')$, it is convenient to consider both j-string and k-string rather than $p_{jh}(j')$ only. Using binary parameters α and β,

$$P((b_j = \alpha) \wedge (b_k = \beta)) = [\text{the sum of occurrence probabilities of bit distributions which satisfy } b_j = \alpha \text{ and } b_k = \beta] / p(I - h,\ j - j',\ k - k',\ H)$$

$$= \frac{p(I-h-1,\ j-j'-\alpha,\ k-k'-\beta,\ H)\cdot H^{\,1-\delta_{\alpha\beta}}(1-H)^{\delta_{\alpha\beta}}}{p(I-h,\ j-j',\ k-k',\ H)}\,. \qquad \text{(K4)}$$

We can get the modified formula for $\mathrm{Prob}_{h,\,h+1}(e'',\ j'',\ k'',\ i''|e',\ j',\ k',\ i')$ by substituting $\mathrm{Prob}_h(j''|j')\cdot\mathrm{Prob}_h(k''|k')$ in Appendix B with Eq. (K4).

Figure 2.1

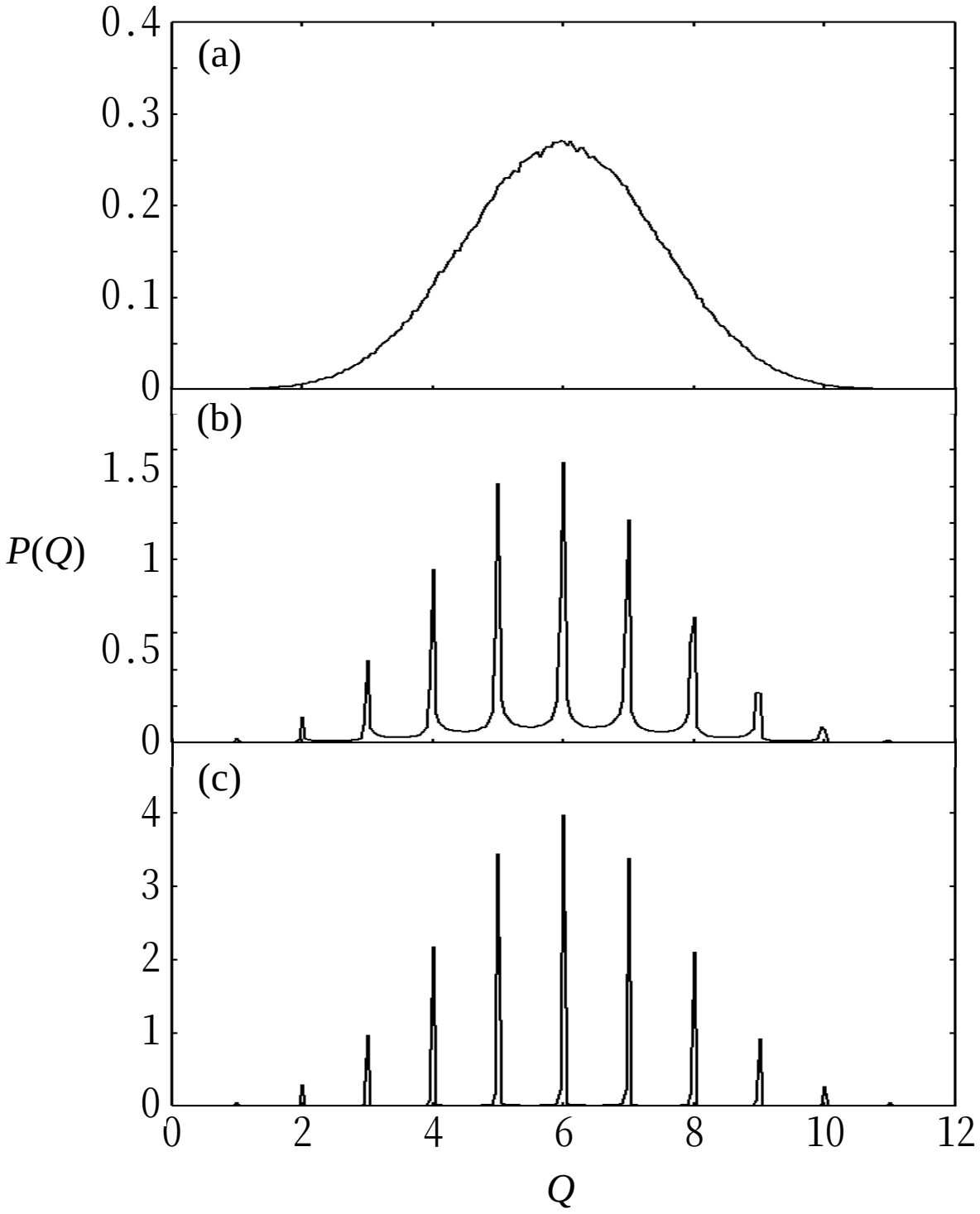


Figure 2.2

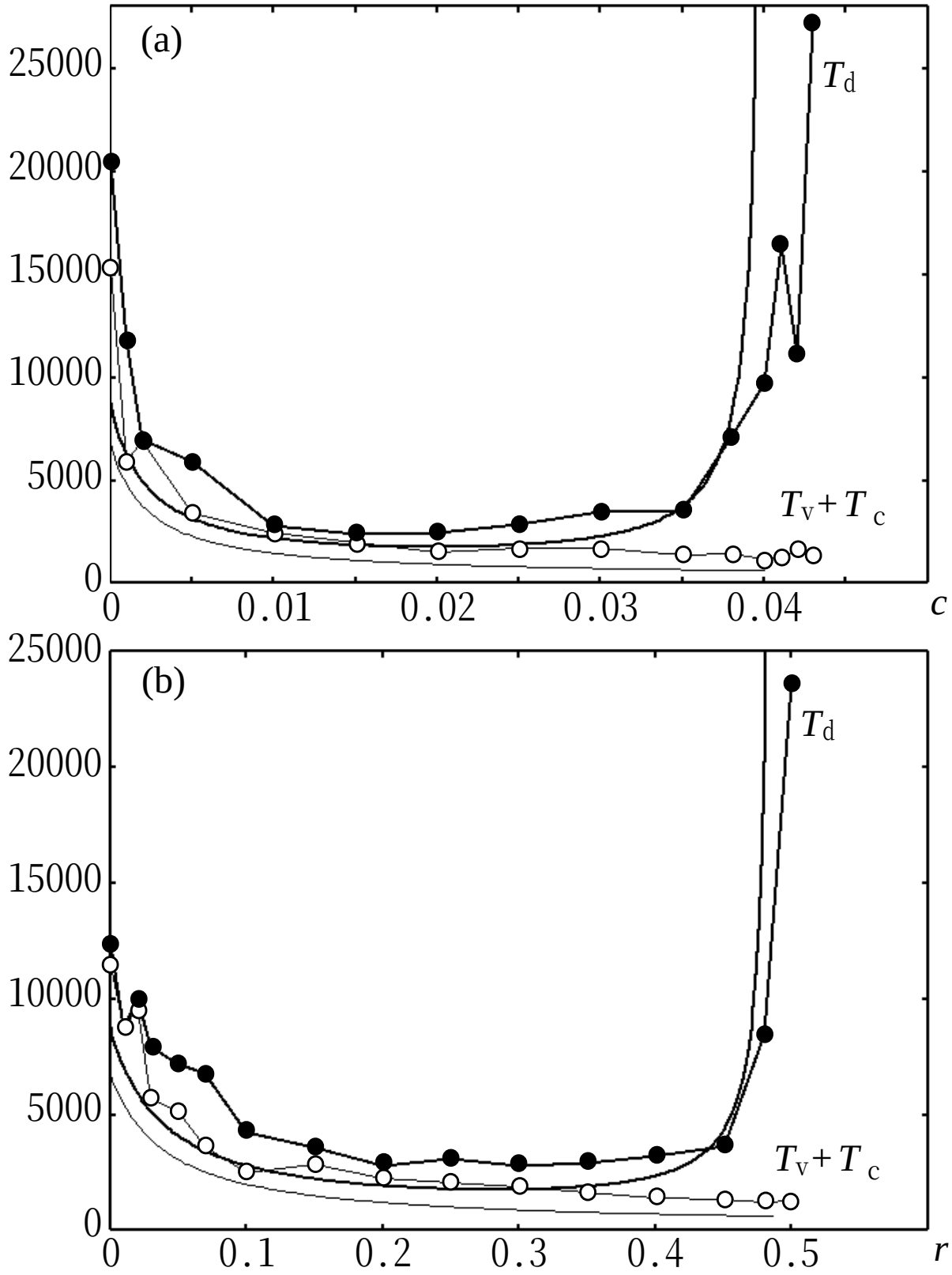

(a)
25000
20000
15000
10000
5000
0
0
0.01
0.02
0.03
0.04
c
T_d
$T_v + T_c$
(b)
25000
20000
15000
10000
5000
0
0
0.1
0.2
0.3
0.4
0.5
r
T_d
$T_v + T_c$

Figure 2.3

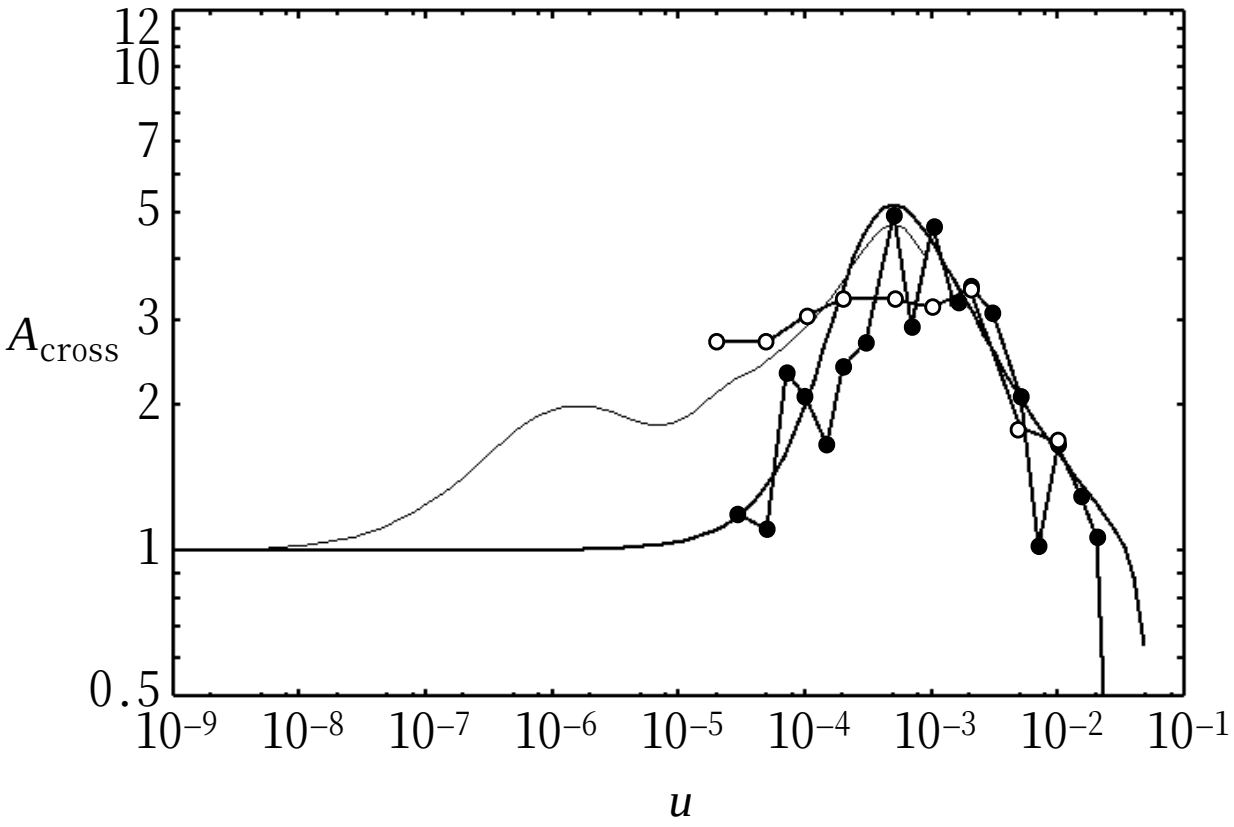


Figure 2.4

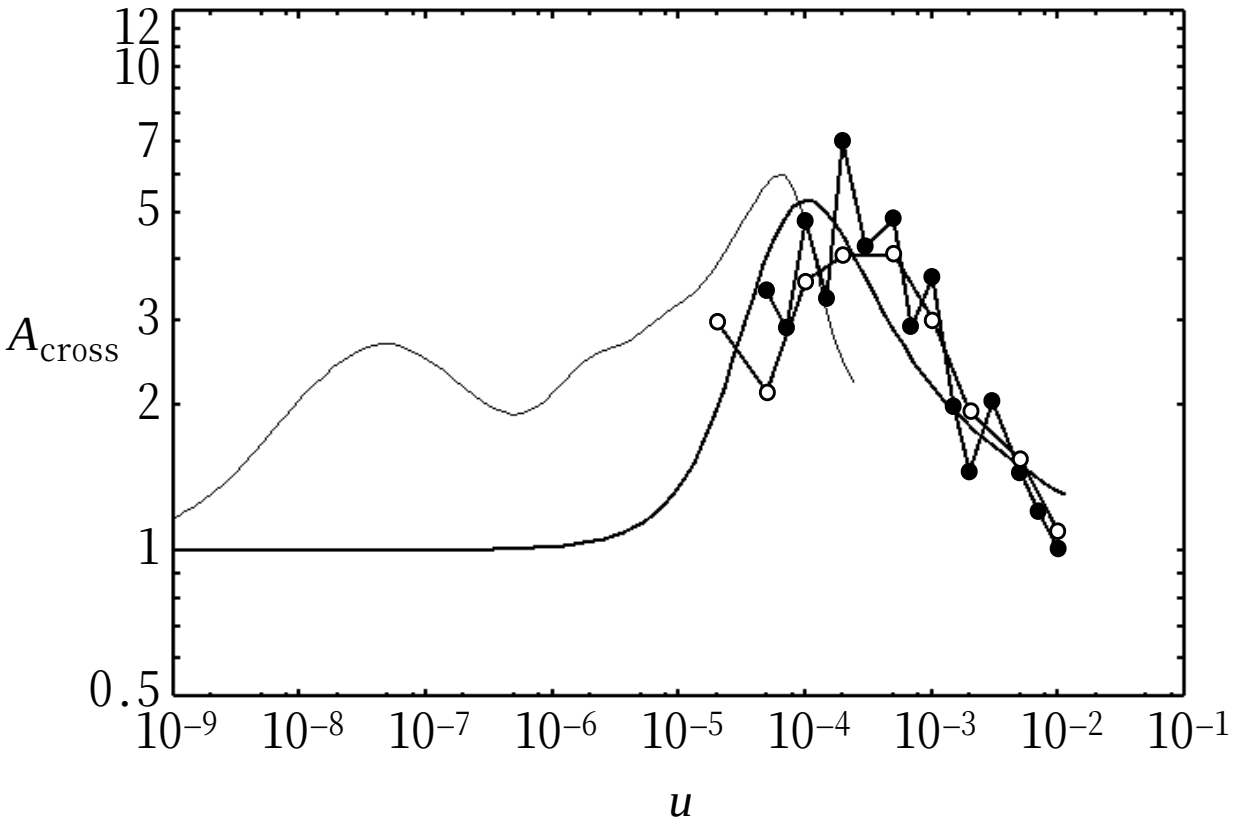


Figure 2.5

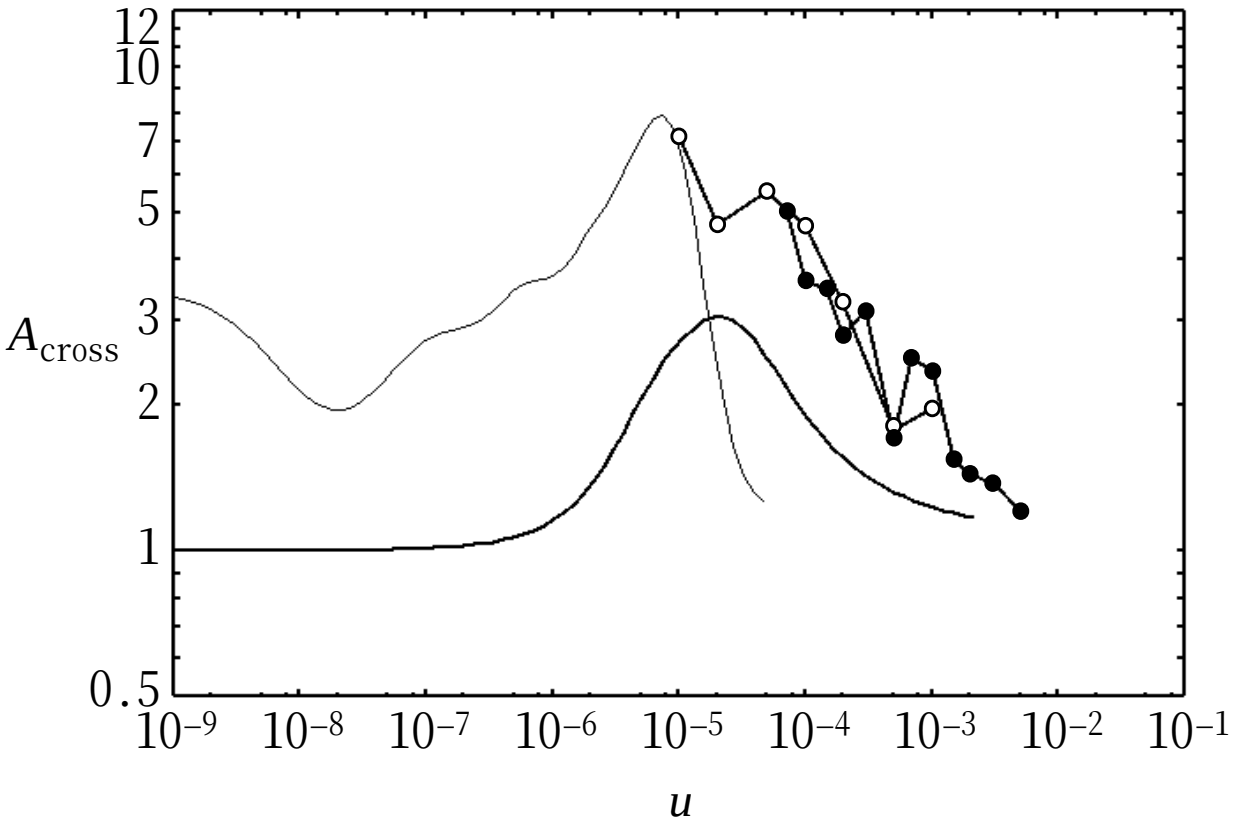


Figure 2.6

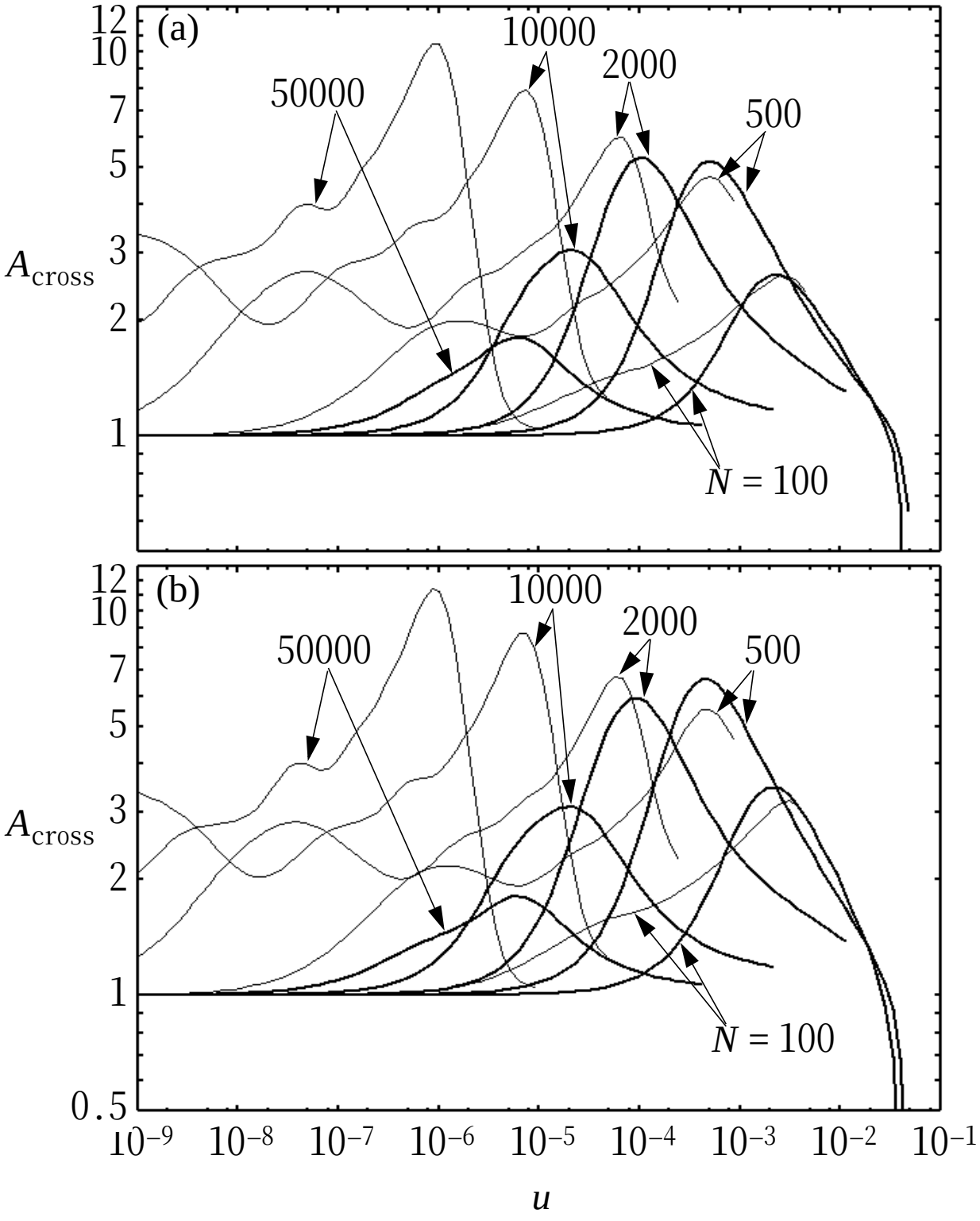

(a)
(b)
50000
10000
2000
500
N = 100
A_cross
u
12
10
7
5
3
2
1
0.5
10^−9
10^−8
10^−7
10^−6
10^−5
10^−4
10^−3
10^−2
10^−1

Figure 2.7

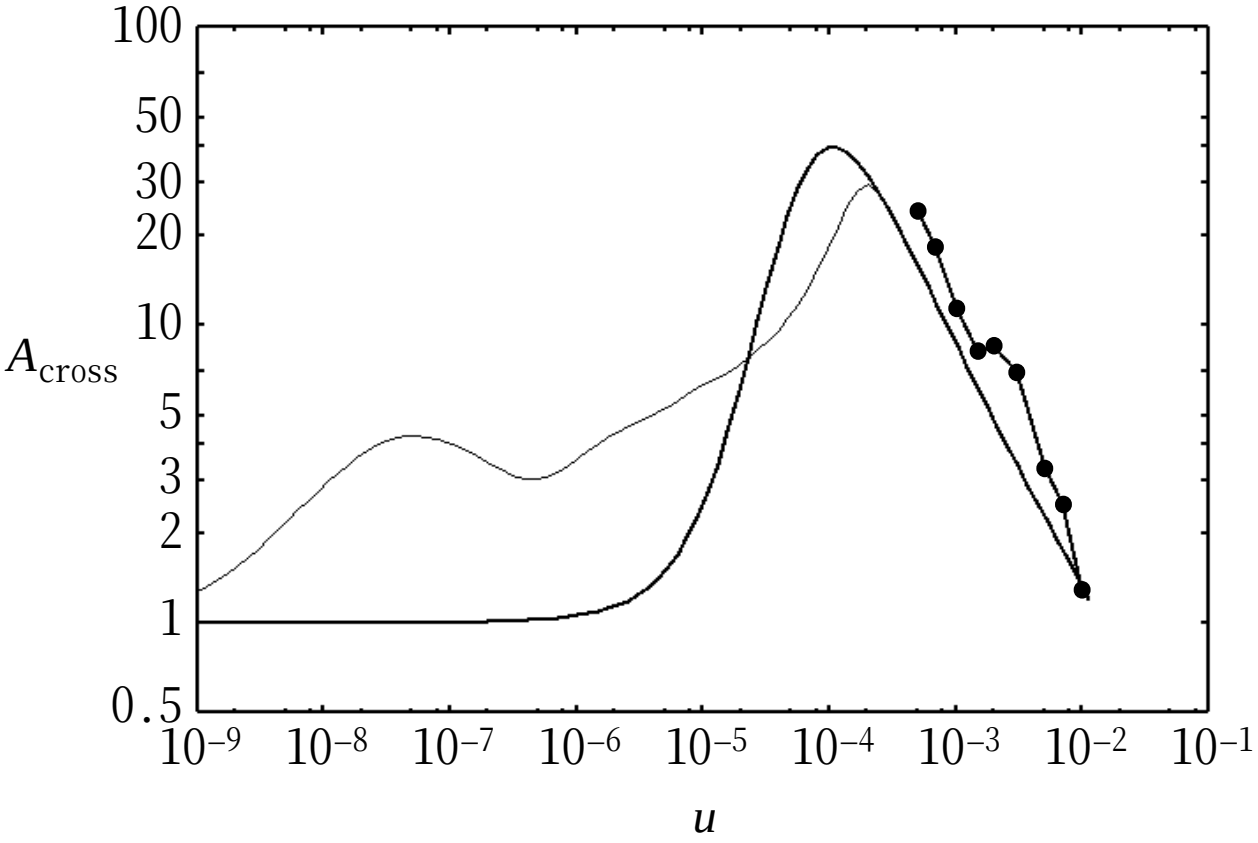

100
50
30
20
10
5
3
2
1
0.5
A_{cross}
10^{-9}
10^{-8}
10^{-7}
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
u

Figure 2.8

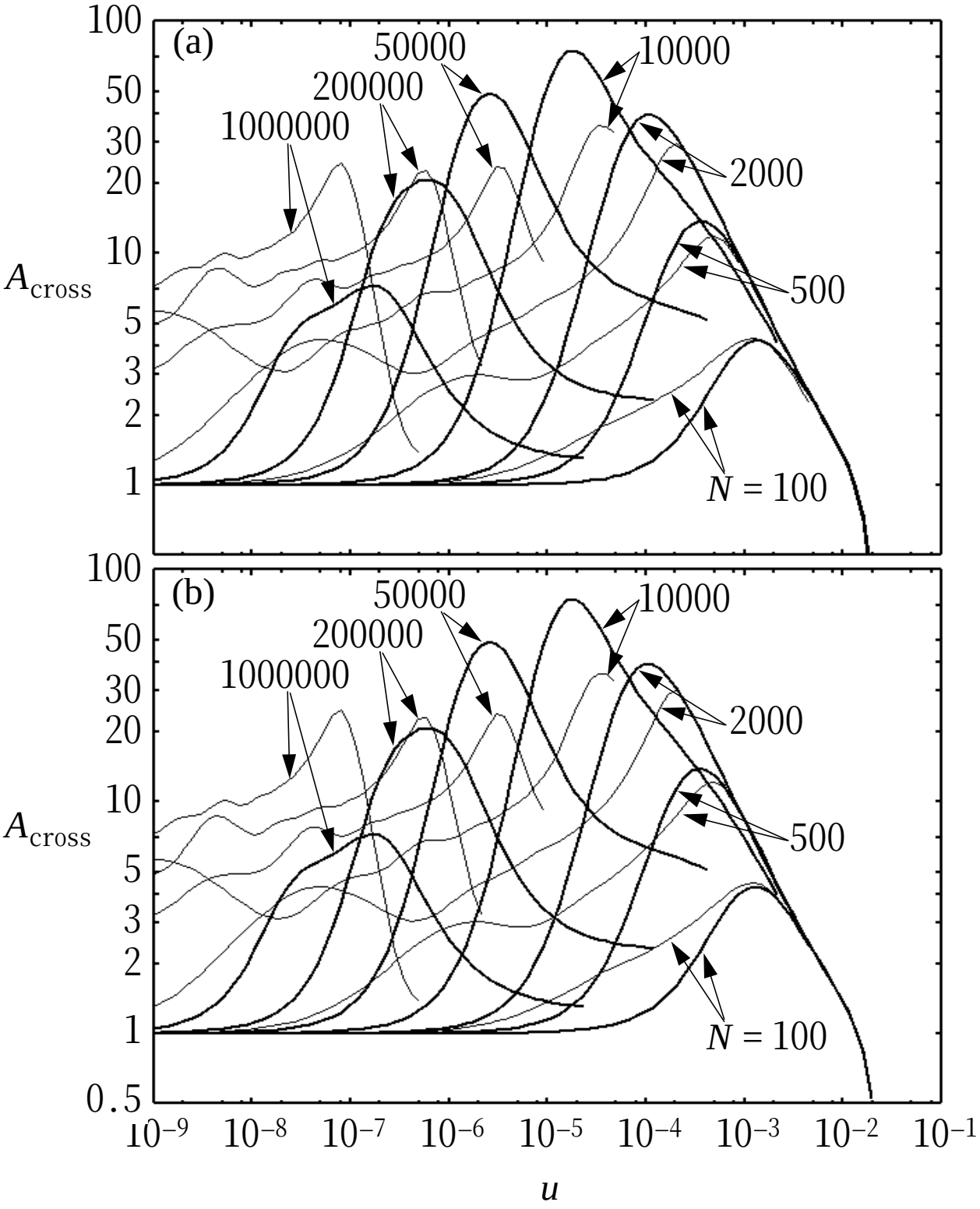

(a)
(b)
50000
10000
200000
1000000
2000
500
N = 100
A_cross
u
100
50
30
20
10
5
3
2
1
0.5
10^−9
10^−8
10^−7
10^−6
10^−5
10^−4
10^−3
10^−2
10^−1

Figure 2.9

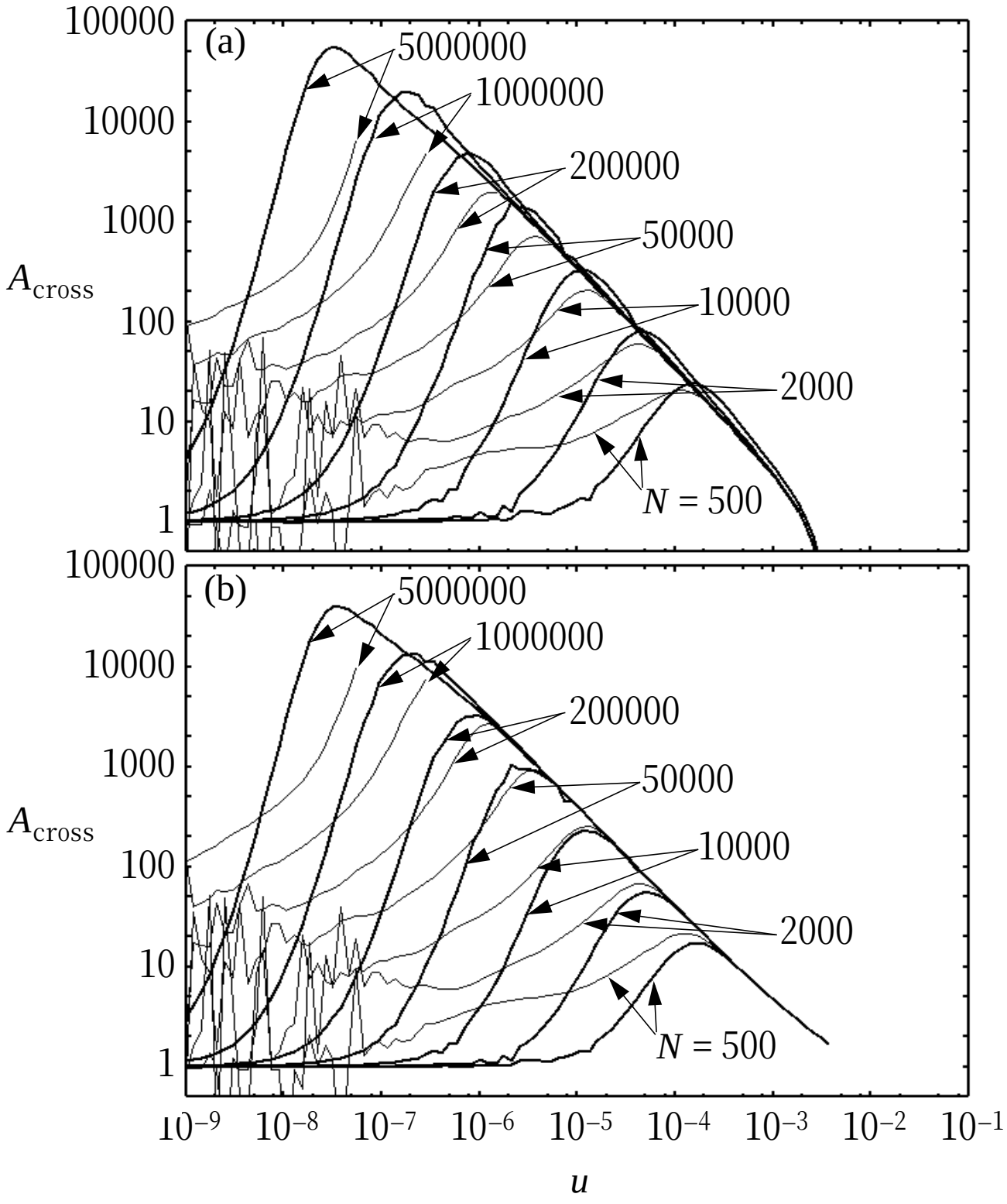

(a)
(b)
5000000
1000000
200000
50000
10000
2000
N = 500
100000
10000
1000
100
10
1
A_cross
10^−9
10^−8
10^−7
10^−6
10^−5
10^−4
10^−3
10^−2
10^−1
u

Figure 2.10

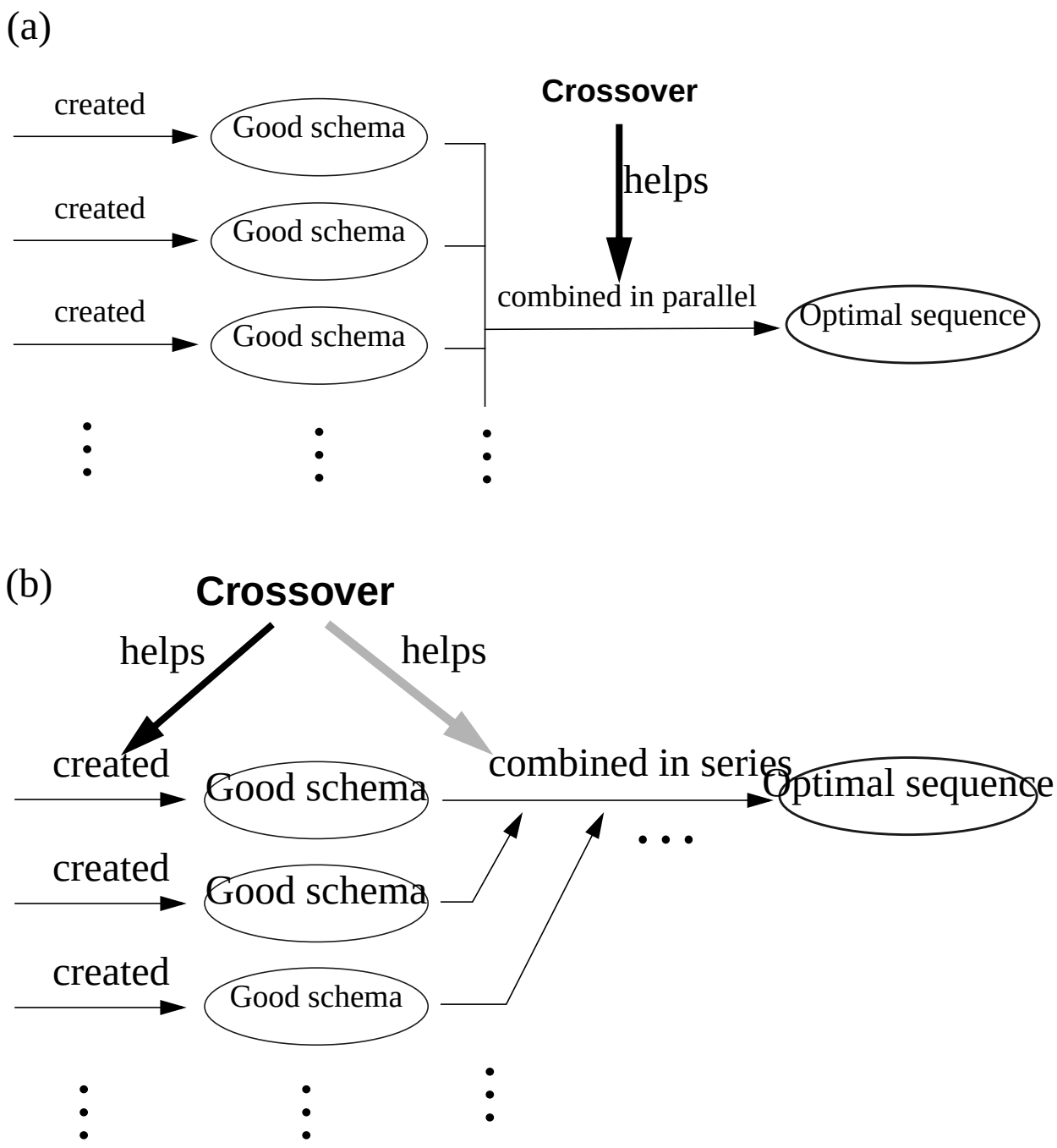

(a)
created
Good schema
created
Good schema
created
Good schema
Crossover
helps
combined in parallel
Optimal sequence
(b)
Crossover
helps
helps
created
Good schema
created
Good schema
created
Good schema
combined in series
Optimal sequence

Figure Legend

Fig. 2.1. Probability distribution $P(Q)$ defined by Eq. (2.6). The epistatic number is $I = 12$, and $\beta = N_e u$ is (a) 0.1, (b) 0.01, or (c) 0.001. The Monte Carlo method (Suzuki & Iwasa, 1997) was adopted to calculate Eq. (2.6). Using the beta distribution, Eq. (2.5), the sample vector $\{q_i\}$ was generated and the value of $Q = \sum q_i$ was calculated. Each graph was made from the histogram created with one million sample vectors.

Fig. 2.2. Mean number of generations until the advantageous sequence first appears in the population ($T_v + T_c$) or until the advantageous sequence dominates the population (T_d). Horizontal axis is (a) c for FPAP mode, or (b) r for MPIP mode. The line with open circles (○) show the results for $T_v + T_c$, and the line with solid circles (●) show the results for T_d given by the direct GA simulation. The lines without circles are the theoretical results given by the formula for large β (the fine curved lines are for $T_v + T_c$ and the thick ones are for T_d). The values of the other parameters are $I = 20$, $N = N_e = 4096$ (in the simulation) or 4000 (in theoretical estimation), and $u = 0.002$.

Fig. 2.3. The crossover's acceleration rate A_{cross} for FPAP mode ($c = 0.015$ and $r = 1$) as a function of the mutation rate u. The values of the other parameters are $I = 12$ and $N = N_e = 512$ (in the simulations) or 500 (in theoretical estimation). Here and in the subsequent figures (Figs. 2.4~2.9), the open circles (○) show the results obtained from the vector representation simulation, the solid circles (●) show the results obtained from the GA simulation, thick curved lines show the theoretical results obtained from the large-β analysis, and fine curved lines show those obtained from the small-β analysis.

Fig. 2.4. The crossover's acceleration rate A_{cross} for FPAP mode ($c = 0.015$ and $r = 1$). Horizontal axis is the mutation rate u. Parameters are $I = 12$ and $N = N_e = 2048$ (in the simulations) or 2000 (in theoretical estimation).

Fig. 2.5. The crossover's acceleration rate A_{cross} for FPAP mode ($c = 0.015$ and

$r = 1$) as a function of the mutation rate u. Parameters are $I = 12$ and $N = N_e = 10000$.

Fig. 2.6. The crossover's acceleration rate A_{cross} for (a) FPAP mode ($c = 0.015$ and $r = 1$) or (b) MPIP mode ($c = 0.5$ and $r = 0.2$). Horizontal axis is the mutation rate u. Numbers in figures are population sizes ($N = N_e$). The epistatic number is $I = 12$.

Fig. 2.7. The crossover's acceleration rate A_{cross} for FPAP mode ($c = 0.015$ and $r = 1$) as a function of the mutation rate u. Parameters are $I = 20$ and $N = N_e = 2048$ (in the simulation) or 2000 (in theoretical estimation).

Fig. 2.8. The crossover's acceleration rate A_{cross} for (a) FPAP mode ($c = 0.015$ and $r = 1$) or (b) MPIP mode ($c = 0.5$ and $r = 0.2$). Horizontal axis is the mutation rate u. Numbers in figures are population sizes ($N = N_e$). The epistatic number is $I = 20$.

Fig. 2.9. The crossover's acceleration rate A_{cross} for (a) FPAP mode ($c = 0.015$ and $r = 1$) or (b) MPIP mode ($c = 0.5$ and $r = 0.2$). Horizontal axis is the mutation rate u. Numbers in figures are population sizes ($N = N_e$). The epistatic number is $I = 40$. The fluctuation of A_{cross} for low u in the results obtained from the small-β analysis is caused by the numerical error which happens in calculating a product of matrices of a large dimension (40×40).

Fig. 2.10. Illustration of the roles of crossover in GAs. (a) Building block hypothesis and implicit parallelism (Conventional). (b) Explicit serialism (Proposed).

Acknowledgments

It is a pleasure to acknowledge Professor Y. Iwasa, Kyushu University, for his advice and valuable comments on this paper. Especially, the author is greatly indebted to him for a helpful discussion about the mathematical analyses of the simulation results. Thanks are also due to Professor M. Feldman, Stanford University, for his guidance to literature on genetic recombination and many valuable comments on the early version of the manuscript on Chapter 1. Those were very helpful for improving the quality of this chapter. Assist. Professor M. Kawata, Shizuoka University, gave me information about the literature on paleontologic studies.

References

ANDEREASEN, V. & CHRISTIANSEN, F. B. (1995). Slow co-evolution of a viral pathogen and its diploid host. *Philos. Trans. R. Soc. London Ser. B* **348,** 341-354.

BERGMAN, A., OTTO, S. P., & FELDMAN, M. W. (1995a). On the evolution of recombination in haploids and diploids I. Deterministic models. *Complexity* **1(1),** 57-67.

BERGMAN, A., OTTO, S. P., & FELDMAN, M. W. (1995b). On the evolution of recombination in haploids and diploids II. Stochastic models. *Complexity* **1(2),** 49-57.

BELL, G. & MAYNARD-SMITH, J. (1987). Short-term selection for recombination among mutually antagonistic species. *Nature* **328,** 66-68.

BERGMAN, A., & FELDMAN, M. W. (1990). More on selection for and against recombination. *Theor. Pop. Biol.* **38,** 68-92.

BERGMAN, A., & FELDMAN, M. W. (1992). Recombination dynamics and the fitness landscape. *Physica D.* **56,** 57-67.

BODMER, W. F. (1970). *The evolutionary significance of recombination in prokaryotes. Prokaryotic and eukaryotic cells*. Edited by H. P. Charles and B. C. J. G. Knight. (Symposia of the Society for General Microbiology No.XX) Cambridge: Cambridge Univ. Press. p.279-294.

BREMERMANN, H. J. (1980). Sex and polymorphism as strategies in host-pathogen interactions. *J. theor. Biol.* **87,** 671-702.

CHARLESWORTH, B. (1993). Directional selection and evolution of sex and recombination. *Genet. Res.* **61,** 205-224.

CROW, J. F. & KIMURA, M. (1965). Evolution in sexual and asexual populations. *Amer. Natur.* **99,** 439-450.

CROW, J. F. & KIMURA, M. (1969). Evolution in sexual and asexual populations: a reply. *Amer. Natur.* **103,** 89-91.

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory.* New York: Harper and Row.

EIGEN, M., GARDINER, W., SCHUSTER, P. & WINKLER, R. (1981). The origin of genetic information. *Sci. Amer.* **244,** 88-118.

ESHEL, I. & FELDMAN, M. (1970). On the evolutionary effect of recombination. *Theor. Pop. Biol.* **1,** 88-100.

EWENS, W. J. (1979). *Mathematical Population Genetics.* New York: Springer-Verlag.

FELDMAN, M. W. (1972). Selection for linkage modification: I. Random mating

populations. *Theor. Pop. Biol.* **3,** 234-346.

FELDMAN, M. W., CHRISTIANSEN, F. B. & BROOKS, L. D. (1980). Evolution of recombination in a constant environment. *Proc. Nat. Acad. Sci. USA* **77,** 4838-4841.

FELDMAN, M. W. & LIBERMAN, U. (1986). An evolutionary reduction principle for genetic modifiers. *Proc. Nat. Acad. Sci. USA* **83,** 4824-4827.

FELDMAN, M. W., OTTO, S. P. & CHRISTIANSEN, F. B. (1997). Population genetic perspectives on the evolution of recombination. *Annu. Rev. Genet.* **30,** 261-295.

FELSENSTEIN, J. (1974). The evolutionary advantage of recombination. *Genetics* **78,** 737-756.

FELSENSTEIN, J. & YOKOYOMA, S. (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83,** 845-859.

FISHER, R. A. (1930). *The Genetic Theory of Natural Selection*. New York: Dover Publications. (2nd Ed. 1958) p.291

FORREST, S. (1993) Genetic algorithms: Principles of natural selection applied to computation. *Science* **261,** 872-878.

FORREST, S. & MITCHELL, M. (1993) What makes a problem hard for a genetic algorithm? Some anomalous results and their explanation. *Machine Learning* **13,** 285-319.

FORREST, S. & MITCHELL, M. (1994) Genetic algorithms and artificial life. *Artificial Life,* **1***,* 267-289.

GOLDBERG, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. New York: Addison-Wesley.

HAIGH, J. (1978). The accumulation of deleterious genes in a population -- Muller's ratchet. *Theor. Pop. Biol.* **14,** 251-267.

HAMILTON, W. D. (1980). Sex versus non-sex versus parasite. *Oikos* **35,** 282-290.

HAMILTON, W. D., AXELROD, R. & TANESE, R. (1990). Sexual reproduction as an adaptation to resist parasites (A Review). *Proc. Nat. Acad. Sci. USA* **87,** 3566-3573.

HILL, W. G. & ROBERTSON, A. (1966). The effect of linkage on limits to artificial selection. *Genet. Res.* **8,** 269-294.

HOLLAND, J. H. (1992). *Adaptation in Natural and Artificial Systems*. Boston: MIT Press.

JAENIKE, J. (1978). An hypothesis to account for the maintenance of sex within populations. *Evolutionary Theory* **3,** 191-194.

JAYAKAR, S. (1970). A mathematical model for interaction of gene frequencies in a

parasite and its host. *Theor. Pop. Biol.* **1,** 140-164.

JONES, T. (1994). A description of Holland's Royal Road function. *Evolutionary Computation* **2** (4)**,** 409-415.

KAPLAN, N. L., HUDSON, R. R. & LANGLEY, C. H. (1989). The "hitch-hiking effect" revisited. *Genetics* **123,** 887-899.

KARLIN, S. (1973). *Sex and infinity: A mathematical analysis of the advantages and disadvantages of genetic recombination. The Mathematical Theory of the Dynamics of Biological Populations*. Edited by M. S. Bartlett AND R. W. Hiorns. London/New York: Academic Press. p.155-194

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge Univ. Press.

KONDRASHOV, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336,** 435-440.

KONDRASHOV, A. S. (1994). The asexual ploidy cycle and the origin of sex. *Nature* **370,** 213-216.

LIBERMAN, U. & FELDMAN, M. W. (1986). Modifiers of mutation rate: A general reduction principle. *Theor. Pop. Biol.* **30,** 341-371.

LLOYD, D. G. (1980). Benefits and handicaps of sexual reproduction. *Evol. Biol.* **13,** 69-111.

LUKE, Y. L. (1969). *The special functions and their applications.* New York and London: Academic Press.

MALMGREN, B. A., BERGGREN, W. A. & LOHMAN, G. P. (1983). Evidence for punctuated gradualism in the Late Neogene Globorotalia tumida lineage of planktonic foraminifera. *Paleobiology* **9,** 377-389.

MATHAI, A. M. & SAXENA, R. K. (1973). *Generalized hypergeometric functions with applications in statistics and physical sciences.* Lecture Notes in Mathematics (Edited by A. Dold & B. Eckmann). Berlin, Heidelberg, and New York: Springer-Verlag.

MAYNARD-SMITH, J. (1971). What use is sex? *J. theor. Biol.* **30,** 319-335.

MAYNARD-SMITH, J. (1978). *The Evolution of Sex*. Cambridge: Cambridge Univ. Press.

MAYNARD-SMITH, J. (1980). Selection for recombination in a polygenic model. *Genet. Res.* **35,** 269-277.

MAYNARD-SMITH, J. (1988). Selection for recombination in a polygenic model - the mechanism. *Genet. Res.* **51,** 59-63.

MAYNARD-SMITH, J. & HAIGH, J. (1974). The hitchhiking effect of a favourable gene.

Genet. Res. **23,** 23-35.

MICHOD, R. E. & LEVIN, B. R. (1988). *The Evolution of Sex : an examination of current ideas.* Sunderland, MA: Sinauer Associates.

MITCHELL, M. (1996). *An Introduction to Genetic Algorithms.* Complex Adaptive Systems series. Boston: MIT Press.

MITCHELL, M., FORREST, S., & HOLLAND, J. H. (1991) The royal road for genetic algorithms: fitness landscapes and GA performance. Edited by F. J. Varela et al. *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* Cambridge, MA: MIT Press. p.245-254.

MULLER, H. J. (1932). Some genetic aspects of sex. *Amer. Natur.* **66,** 118-138.

MULLER, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.* **1,** 2-9.

NEI, M. (1967). Modification of linkage intensity by natural selection. *Genetics.* **57,** 625-626.

NEI, M. (1969). Linkage modification and sex difference in recombination. *Genetics.* **63,** 681-699.

NIMWEGEN, E., CRUTCHFIELD, J. P., & MITCHELL, M. (1996) Finite populations induce metastability in evolutionary search. *Physics Letters A.* **229** (2), 144-150.

NIMWEGEN, E., CRUTCHFIELD, J. P., & MITCHELL, M. (1997) Statistical dynamics of the royal road genetic algorithm. *SFI Working Paper 97-04-035.* submitted to *the Special Issue on Evolutionary Computation of the journal Theoretical Computer Science,* A. E. Eiben and G. Rudolph (Eds.)

OTTO, S., FELDMAN, M. W., & CHRISTIANSEN, F. B. (1994). Some advantages and disadvantages of recombination. Edited by S. Levin *Frontiers in Mathematical Biology,* Vol. 100 p.198-211.

PRUDNIKOV, A. P. & MARICHEV, O. I. (1990). *Integrals and series. Vol. 3 More special functions.* (Traslated by G. G. Gould from the original edition which was published in Russian in 1986) New York, Philadelphia, London, Paris, Montreux, Tokyo, and Melbourne: Gordon and Breach Science Publisher.

PRUGEL-BENNETT, A. & SHAPIRO, J. L. (1994) Analysis of genetic algorithms using statistical mechanics. *Physical Review Letters.* **72** (9), 1305-1309.

PRUGEL-BENNETT, A. & SHAPIRO, J. L. (1996) The dynamics of a genetic algorithm for simple random ising systems. *Physica D.* in press.

RATTRAY, M. & SHAPIRO, J. L. (1996) The dynamics of a genetic algorithm for a simple learning problem. *J. of Phys. A.* **29** (23), 7451-7473.

REDFIELD, R. J. (1994). Male mutation rates and the cost of sex for females. *Nature* **369,** 145-147.

SASAKI, A. & IWASA, Y. (1987). Optimal recombination rate in fluctuating environments. *Genetics* **115,** 377-388.

SUZUKI, H. (1996) Functional Emergence with Multiple von Neumann Computers. Edited by C. Langton and K. Shimohara. *Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living System*s Cambridge, MA: MIT Press. p.108-115.

SUZUKI, H. (1997). Multiple von Neumann Computers: An Evolutionary Approach to Functional Emergence. *Artificial Life.* **3,** 121-142.

SUZUKI, H. & IWASA, Y. (1997). GA Performance in a Babel-like Fitness Landscape. *Proceedings of the Nineth IEEE International Conference on Tools with Artificial Intelligence*. Cambridge, MA: MIT Press. pp. 357-366.

TAKAHATA, N. (1982). Sexual recombination under the joint effects of mutation, selection and random sampling drift. *Theor. Pop. Biol.* **22,** 258-277.

TOOBY, J. (1982). Pathogens, polymorphism, and the evolution of sex. *J. theor. Biol.* **97,** 557-576.

VOSE, M. D. & WRIGHT, A. H. (1995) Simple genetic algorithms with linear fitness. *Evol. Computation* **2(4),** 347-368.

WILLIAMS, G. C. (1975). *Sex and Evolution*. Princeton: Princeton Univ. Press.

WILLIAMSON, P. G. (1981). Paleontological documentation of speciation in Venozoic molluscs from Turkana basis. *Nature* **293,** 437-443.

WU, A. S. & LINDSAY, R. K. (1996) A comparison of the fixed and floating building block representation in the genetic algorithm. *Evol. Computation* **4(2),** 169-193.

ZEYL, C. & BELL, G. (1997). The advantage of sex in evolving yeast populations. *Nature* **388,** 465-468.

ZHIVOTOVSKY, L. A., FELDMAN, M. W., & CHRISTIANSEN, F. B. (1994). Evolution of recombination among multiple selected loci: A generalized reduction principle. *Proc. Nat. Acad. Sci. USA* **91,** 1079-1083.