On the Convergence of Multi-Parent Genetic Algorithms

Chuan-Kang Ting
International Graduate School of Dynamic Intelligent Systems
University of Paderborn, 33100 Paderborn, Germany
ckting@upb.de

Abstract— This paper presents a Markov model for the convergence of multi-parent genetic algorithms (MPGAs). The proposed model formulates the variation of gene frequency caused by selection, multi-parent crossover, and mutation. In addition, it reveals the pairwise equivalence phenomenon in the number of parents and identifies the correlation between this number and the mean fitness in the OneMax problem. The good fit between theoretical and experimental results demonstrate the capability of this model. Moreover, the superiority of multi-parent crossover in convergence fitness over 2-parent crossover is validated theoretically as well as empirically.

I. INTRODUCTION

Multi-parent genetic algorithms (MPGAs) are genetic algorithms using multi-parent crossovers. Traditionally, genetic algorithms (GAs) adopt two parents in crossover to reproduce offspring. This idea is reasonable because, to the best of our knowledge, the form of sexual reproduction on the Earth is absolutely of two parents. Multi-parent crossovers break through this natural limitation by allowing more than two parents in the process of crossover. In a sense, multi-parent genetic algorithms are said to be multi-parent generalization of genetic algorithms.

Several multi-parent crossovers have been proposed for MPGAs and shown their power in a variety of optimization problems [8], [10], [21]. However, most of these crossovers are validated empirically. Theoretical analysis on the effect of multiple parents upon crossover or upon the whole MPGA is lacking. Thus, the role of parent numbers in MPGAs is still open.

This paper presents a Markov model for exploring the influence of raising parents in MPGAs. Specifically, we investigate *occurrence based scanning crossover* (OB-Scan) [11] — a multi-parent generalization of *uniform crossover* [19]. In addition, this paper focuses on the OneMax problem. Even though the OneMax problem is relatively easy, the analysis on it founds a basic understanding of how the GA operators work and interact in the course of evolution. Furthermore, the analytical models for the OneMax problems are promisingly applicable to other problem domains.

In the procedures for analysis, first we investigate the gene frequency affected by selection, multi-parent crossovers, and mutation individually. Based on gene frequency, we model by Markov chains the behavior of MPGAs. This Markov model formulates the variation of gene frequency in the course of MPGA's evolution. Additionally, the mean convergence fitness for MPGAs in the OneMax problem is derived accordingly. These theoretical results are further verified by a series of experiments.

The rest of this paper is organized as follows. Section 2 describes OB-Scan and Section 3 analyzes the gene frequency affected by the MPGA operators. Next, in Section 4 we model MPGAs with Markov chains. Theoretical results and experimental validation are presented in Section 5. Finally, conclusions are drawn in Section 6.

II. OCCURRENCE BASED SCANNING CROSSOVER (OB-SCAN)

Occurrence based scanning crossover is one of the *scanning crossovers* [11] — a multi-parent generalization of *uniform crossover*. In uniform crossover, the donor for each offspring gene is randomly picked from two parents. Extended to more than two parents, scanning crossovers choose the donor at random or using heuristics. According to different strategies, Eiben et al. [11] proposed three variations of scanning crossovers: *uniform scanning crossover* (U-Scan), *occurrence based scanning crossover* (OB-Scan), and *fitness based scanning crossover* (FB-Scan). In this paper, we only discuss OB-Scan.

Rather than random, OB-Scan determines offspring genes depending on the occurrence of parental genes at that locus. Specifically, it picks the *majority* of parental gene values as the offspring gene for each locus. Note that in this paper OB-Scan is defined to break ties by randomly[1] choosing a binary. Examples of 2-parent OB-Scan (corresponding to uniform crossover) and 4-parent OB-Scan are given in Fig. 1. The formal definition of the components of GAs and OB-Scan are drawn below.
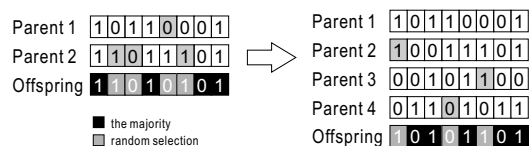

Fig. 1.   Examples of 2-parent OB-Scan (left) and 4-parent OB-Scan (right)

[1]The original OB-Scan [11] breaks ties by directly inheriting the genes of the first selected parent. However, random tie break conforms to generalization of uniform crossover.

Definition 1 (Chromosome and Population):

1) A *chromosome* $\vec{c}$ is encoded as a bit string, i.e. $\vec{c} \stackrel{\text{def}}{=} (c_1, \ldots, c_l) \in \{0,1\}^l$, where c_i denotes a *gene* and l is the *chromosome length*.
2) A *population* C is a set of chromosomes: $C \stackrel{\text{def}}{=} \{\vec{c}_1, \ldots, \vec{c}_m\}$, where $\vec{c}_i \in \{0,1\}^l$ and m is the *population size*.

Definition 2 (OB-Scan): Given n parents $\vec{c}_1, \ldots, \vec{c}_n \in C$ selected from population C, OB-Scan reproduces the offspring $\vec{c}' = \mathcal{X}_{\text{ob}}(\vec{c}_1, \ldots, \vec{c}_n) = (c'_1, \ldots, c'_l)$ by

$$c'_i = \begin{cases} 0 & \text{if } \sum_{j=1}^{n} (\vec{c}_j)_i < \frac{n}{2} \\ 1 & \text{if } \sum_{j=1}^{n} (\vec{c}_j)_i > \frac{n}{2} \quad \text{for } i = 1, \ldots, l, \\ \text{rand}(0,1) & \text{otherwise} \end{cases}$$

where $(\vec{c}_j)_i$ denotes the ith gene of the chromosome $\vec{c}_j$, and $\text{rand}(0,1)$ is a binary random function.

III. VARIATION OF GENE FREQUENCY

Gene frequency is widely used as a quantitative measure of genetic variation in population genetics [15]. It also suffices to clue us in on the course of evolution in GAs. In this section we analyze the variation of gene frequency caused by selection, OB-Scan, and bit-flip mutation.

Definition 3 (Gene Frequency): The *gene frequency* $p_k(\alpha, t)$ is defined as the proportion of allele α at locus k in the population at time t. Let $C = \{\vec{c}_1, \ldots, \vec{c}_m\}$ be a population at time t and let $C_k(\alpha) = \{\vec{c} \in C \mid c_k = \alpha\}$ be the subset in which chromosomes possess allele α at locus k. The gene frequency

$$p_k(\alpha, t) \stackrel{\text{def}}{=} \frac{|C_k(\alpha)|}{|C|},$$

where $|C|$ and $|C_k(\alpha)|$ represent the cardinality of set C and $C_k(\alpha)$, respectively.

In GAs chromosomes are represented as binary strings. Thus there exist two gene frequencies $p_k(1,t)$ and $p_k(0,t)$ with $p_k(1,t) = 1 - p_k(0,t)$ for every locus k and time t. For simplicity, we refer to the gene frequency $p_k(1,t)$ as $p_k(t)$ and refer to $p_k(0,t)$ as $(1 - p_k(t))$. Incidentally, the symbol $p_k(t)$ is referred to as p_k while the indication of time t is not in effect.

Remark 1: Let c_k be the kth gene of chromosome $\vec{c}$. In GAs, for a population C at time t

$$p_k(t) \stackrel{\text{def}}{=} \mathrm{E}[c_k] = \frac{1}{m} \sum_{c \in C} c_k, \tag{1}$$

$$\sigma_k^2(t) \stackrel{\text{def}}{=} \text{Var}(c_k) = p_k(1 - p_k). \tag{2}$$

Definition 4 (Variation of Gene Frequency in GAs): Let s, x, m, and r denote the selection, crossover, mutation, and replacement (survivor), respectively. The process of GA with respect to gene frequency can be expressed as

$$p_k(t) \stackrel{s}{\longrightarrow} p_k^s(t) \stackrel{x}{\longrightarrow} p_k^x(t) \stackrel{m}{\longrightarrow} p_k^m(t) \stackrel{r}{\longrightarrow} p_k(t+1). \tag{3}$$

The successive sections will investigate the influence of selection, multi-parent crossovers, and mutation on the gene frequency. In the analysis we will make use of the well-known *binomial distribution*. The probability mass function (p.m.f) of binomial distribution $B(n,p)$ is

$$B(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

A. Variation by Selection

This section examines how selection affects gene frequency. Recall the fact that most of the selection operators in GAs depend on fitness — a phenotypic property of chromosomes. Conversely, gene frequency concerns the genotypes of chromosomes. The mapping from genotypes to phenotype is problem-dependent in essence. In this paper we focus on the OneMax problem. The fitness of the OneMax problem provides a direct way to bridge this gap between phenotype and genotype.

Definition 5 (The OneMax Problem): The OneMax problem is to find the binary string $\vec{c} = (c_1, \ldots, c_l) \in \{0,1\}^l$, which maximizes the following function $f : \{0,1\}^l \to \{0, 1, \ldots, l\}$:

$$f(\vec{c}) = \sum_{k=1}^{l} c_i.$$

Lemma 1: In GAs, the covariance of the genes at any two loci $u, v \in \{1, \ldots, l\}$ has

$$\text{Cov}(c_u, c_v) = 0.$$

Proof: Let C_k be the subset in which chromosomes possess allele α at locus k. Since

$$\mathrm{E}[c_u c_v] = \frac{1}{m^2} |C_u| |C_v| = p_u p_v,$$

The covariance

$$\text{Cov}(c_u, c_v) = \mathrm{E}[c_u c_v] - \mathrm{E}[c_u]\mathrm{E}[c_v] = p_u p_v - p_u p_v = 0.$$

$\blacksquare$

Next, we give a definition of mean fitness, which will be extensively discussed in this paper. The relation between mean fitness and gene frequency will be presented afterwards.

Definition 6 (Mean Fitness): The *mean fitness* of a population C is defined as

$$\bar{f} = \frac{1}{|C|} \sum_{\vec{c} \in C} f(\vec{c}).$$

Lemma 2: Let p_k be the gene frequency at locus k. In the OneMax problem we have the mean $\bar{f}$ and variance σ_F^2 of fitness:

$$\bar{f} = \sum_{k=1}^{l} p_k$$

$$\sigma_F^2 = \sum_{k=1}^{l} \sigma_k^2 = \sum_{k=1}^{l} p_k(1 - p_k)$$

Proof: The fitness mean can be simply derived by (1):

$$\bar{f} = \mathrm{E}[f] = \mathrm{E}\left[\sum_{k=1}^{l} c_k\right] = \sum_{k=1}^{l} \mathrm{E}[c_k] = \sum_{k=1}^{l} p_k.$$

Using Lemma 1, we have the fitness variance

$$\sigma_F^2 = \mathrm{Var}(f) = \mathrm{Var}\left(\sum_{k=1}^{l} c_k\right)$$

$$= \sum_{k=1}^{l} \mathrm{Var}(c_k) + \sum_{u \neq v} \mathrm{Cov}(c_u, c_v) = \sum_{k=1}^{l} \sigma_k^2.$$

■

To compute the variation of gene frequency caused by selection, here we introduce the concept of *selection intensity*, which originates from quantitative genetic [12] but now is widely used in evolutionary computation community as a quantitative measure of selection pressure [5], [16]. The definitions of selection differential and selection intensity are given as follows.

Definition 7 (Selection Differential): Let $\bar{f}(t)$ and $\bar{f}^s(t)$ be the mean fitness of the population and that of selected parents at generation t, respectively. The *selection differential* $S(t)$ is defined as

$$S(t) = \bar{f}^s(t) - \bar{f}(t).$$

Definition 8 (Selection Intensity): Let $S(t)$ be the selection differential and $\sigma_F(t)$ be the standard deviation of fitness in a population at generation t. The *selection intensity* is defined as

$$\mathcal{I}(t) = \frac{S(t)}{\sigma_F(t)} = \frac{\bar{f}^s(t) - \bar{f}(t)}{\sigma_F(t)}. \qquad (4)$$

In this paper we additionally define a *selection intensity for gene frequency* to measure the effect of selection upon gene frequency.

Definition 9 (Selection Intensity for Gene Frequency): The selection intensity $\mathcal{I}_k^{\mathrm{p}}$ for gene frequency at locus $k \in \{1, \ldots, l\}$ is defined as

$$\mathcal{I}_k^{\mathrm{p}} \overset{\text{def}}{=} \frac{p_k^s - p_k}{\sigma_k}.$$

Lemma 3: In the OneMax problem, the expected selection intensity for gene frequency

$$\mathrm{E}\left[\mathcal{I}^{\mathrm{p}}\right] = \frac{\mathcal{I}}{\sqrt{l}}\sqrt{1 + \frac{\mathrm{Var}(\sigma)}{\mathrm{E}^2[\sigma]}}. \qquad (5)$$

Proof: The definition of selection intensity for gene frequency tells

$$p_k^s - p_k = \mathcal{I}_k^{\mathrm{p}}\sigma_k.$$

Since $\mathcal{I}_k^{\mathrm{p}}$ and σ_k are independent, averaging all loci gives

$$\mathrm{E}[p^s] - \mathrm{E}[p] = \mathrm{E}\left[\mathcal{I}^{\mathrm{p}}\right]\mathrm{E}[\sigma]. \qquad (6)$$

According to Lemma 2 and eq. (4), the above equation can be rewritten as

$$\mathrm{E}\left[\mathcal{I}^{\mathrm{p}}\right] = \frac{\mathrm{E}[p^s] - \mathrm{E}[p]}{\mathrm{E}[\sigma]} = \frac{\bar{f}^s - \bar{f}}{l\mathrm{E}[\sigma]} = \frac{\mathcal{I}\sigma_F}{l\mathrm{E}[\sigma]}$$

$$= \frac{\mathcal{I}\sqrt{\sum_{k=1}^{l}\sigma_k^2}}{l\mathrm{E}[\sigma]} = \frac{\mathcal{I}\sqrt{l\mathrm{E}[\sigma^2]}}{l\mathrm{E}[\sigma]} = \frac{\mathcal{I}}{\sqrt{l}}\sqrt{\frac{\mathrm{E}[\sigma^2]}{\mathrm{E}^2[\sigma]}}.$$

The variance $\mathrm{Var}(\sigma) = \mathrm{E}[\sigma^2] - \mathrm{E}^2[\sigma]$. Thus,

$$\mathrm{E}\left[\mathcal{I}^{\mathrm{p}}\right] = \frac{\mathcal{I}}{\sqrt{l}}\sqrt{\frac{\mathrm{E}^2[\sigma] + \mathrm{Var}(\sigma)}{\mathrm{E}^2[\sigma]}} = \frac{\mathcal{I}}{\sqrt{l}}\sqrt{1 + \frac{\mathrm{Var}(\sigma)}{\mathrm{E}^2[\sigma]}}.$$

■

Remark 2: As $\mathrm{Var}(\sigma) \ll \mathrm{E}^2[\sigma]$, the expected selection intensity for gene frequency becomes

$$\mathrm{E}\left[\mathcal{I}^{\mathrm{p}}\right] \approx \frac{\mathcal{I}}{\sqrt{l}}.$$

In this paper we utilize the above expectation as the selection intensity for gene frequency at any locus k. Consequently, the variation of gene frequency caused by selection approximates to

$$p_k^s \approx p_k + \left(\frac{\mathcal{I}}{\sqrt{l}}\right)\sigma_k. \qquad (7)$$

A practical advantage of (7) is its correlation to selection intensity $\mathcal{I}$. Accordingly, we can utilize the existing analytical results on the selection intensity $\mathcal{I}$ of diverse selection operators in GAs, e.g. [5]. Note that the selection intensity for gene frequency in (7) is based on the assumption $\mathrm{Var}(\sigma) \ll \mathrm{E}^2[\sigma]$. Even though this assumption does not necessarily hold in the course of evolution, we will empirically show this approximation can work well in Section 5.

B. Variation by OB-Scan

Here we analyze the gene frequency affected by OB-Scan. Since OB-Scan is a multi-parent generalization of uniform crossover, the analyses and results presented here are applicable to uniform crossover as well.

Before conducting the analysis of OB-Scan, we introduce the incomplete beta function for simplifying the expression of equations .

Definition 10 (Incomplete Beta Function): The *incomplete beta function* is defined as

$$I_x(a, b) \overset{\text{def}}{=} \frac{1}{\mathrm{Beta}(a, b)}\int_0^x t^{a-1}(1-t)^{b-1}dt,$$

where $a, b > 0$ and $\mathrm{Beta}(a, b)$ is the *beta function.*
The incomplete beta function holds the following properties:
1) (26.5.24 [1]) For binomial distribution $B(n, p)$,

$$\sum_{i=a}^{n} B(i; n, p) = I_p(a, n - a + 1). \qquad (8)$$

2) (26.5.16 [1])

$$I_x(a, b) = \frac{1}{a \cdot \mathrm{Beta}(a, b)}x^a(1-x)^b + I_x(a+1, b). \qquad (9)$$

Now we embark on the analysis of OB-Scan's influence on gene frequency.

Lemma 4: Suppose we have the gene frequency p_k^s of the selected parents. The gene frequency, denoted by p_k^{ob}, of the offspring reproduced by n-parent OB-Scan $\mathcal{X}_{\mathrm{ob}}$ with $n \in \mathbb{N}_{>1}$ has

$$p_k^{\mathrm{ob}} = I_{p_k^s}(a, a),$$

where I_p denotes the incomplete beta function and $a = \lceil\frac{n}{2}\rceil$.

Proof: Let X be the number of parents possessing the allele 1 at locus k among n selected parents. Since the process

of random selection is independent, it is a Bernoulli process. Performing this selection n times, the number X holds a binomial distribution with p.m.f.

$$\Pr(X = x) = B(x; n, p_k^s) = \binom{n}{x} (p_k^s)^x (1 - p_k^s)^{n-x}.$$

Let $\mathfrak{D}_1$ denote the event that OB-Scan assigns the allele 1 to the offspring locus k. According to Definition 2, OB-Scan yields

$$\Pr(\mathfrak{D}_1 \mid X = x) = \begin{cases} 1 & \text{if } x > n/2, \\ 0 & \text{if } x < n/2, \\ \frac{1}{2} & \text{if } x = n/2. \end{cases}$$

For OB-Scan with odd number of parents ($n = 2a - 1$ with $a \in \mathbb{N}_{>1}$), from (8) we have

$$p_k^{\mathrm{ob}} = \Pr(\mathfrak{D}_1) = \sum_{x=0}^{n} \Pr(\mathfrak{D}_1 \mid X = x) \cdot \Pr(X = x)$$
$$= \sum_{x=a}^{2a-1} B(x; 2a - 1, p_k^s) = I_{p_k^s}(a, a).$$

Similarly, for OB-Scan with even number of parents ($n = 2a$ with $a \in \mathbb{N}$), we have

$$p_k^{\mathrm{ob}} = \sum_{x=a+1}^{2a} B(x; 2a, p_k^s) + \frac{1}{2} B(a; 2a, p_k^s)$$
$$= I_{p_k}(a, a) - \frac{1}{a\mathrm{Beta}(a, a)} (p_k^s)^a (1 - p_k^s)^a$$
$$+ \frac{1}{2} \binom{2a}{a} (p_k^s)^a (1 - p_k^s)^a \qquad \text{(from (9))}$$
$$= I_{p_k^s}(a, a) + \left[-\frac{\Gamma(2a)}{a\Gamma(a)\Gamma(a)} + \frac{1}{2} \frac{(2a)!}{a!a!} \right] (p_k^s)^a (1 - p_k^s)^a$$
$$= I_{p_k^s}(a, a).$$

∎

Corollary 1 (Pairwise Equivalence): Let $p_k^{\mathrm{ob}(n)}$ be the gene frequency p_k^{ob} corresponding to n-parent OB-Scan. For $n \in 2\mathbb{N}$ and $n \geq 4$,

$$p_k^{\mathrm{ob}(n)} = p_k^{\mathrm{ob}(n-1)}.$$

Proof: Trivial (since $\lceil \frac{n}{2} \rceil = \lceil \frac{n-1}{2} \rceil$ in Lemma 4). ∎

C. Variation by Mutation

In this work, the analysis of mutation focuses on the most popular mutation — *bit-flip mutation*. Bit-flip mutation randomly chooses a locus and then *flips* the gene at this locus, i.e. $0 \to 1$ and $1 \to 0$. A parameter γ_m, called *mutation rate*, was introduced to determine the probability for a gene to be mutated. The variation of gene frequency caused by bit-flip mutation is calculated in the following lemma.

Lemma 5: Suppose we have the gene frequency p_k^x. Given the mutation rate γ_m, the gene frequency of the offspring mutated by bit-flip mutation has

$$p_k^m = p_k^x + \gamma_m(1 - 2p_k^x).$$

Proof: Considering bit-flip mutation, the cases to yield the gene 1 at locus k are: $0 \to 1$ (mutated) and $1 \to 1$ (not

mutated). Let c_k be the gene at locus k before mutation and c_k' be the gene after mutation. The gene frequency

$$p_k^m = \Pr\{c_k' = 1 \mid c_k = 0\} \Pr\{c_k = 0\}$$
$$+ \Pr\{c_k' = 1 \mid c_k = 1\} \Pr\{c_k = 1\}$$
$$= \gamma_m(1 - p_k^x) + (1 - \gamma_m)p_k^x$$
$$= p_k^x + \gamma_m(1 - 2p_k^x).$$

∎

IV. MODELING WITH MARKOV CHAINS

Markov chains have been used to model the exact behavior of GAs [2], [13], [20] and to analyze the convergence of GAs [7], [9], [17], [18]. In this paper, we use Markov chains to model the evolution of gene frequency. Furthermore, the mean fitness in the OneMax problem is derived from this Markov model.

A. The Model for Gene Frequency

In light of gene frequency, GA can be viewed as a stochastic process manipulating the number of allele 1 (or 0) in the population: Let random variables $G_k(t) \in \{0, 1, \ldots, m\}$ be the number of allele 1 at locus k at generation t. The process of GAs on gene frequency can be represented as $\{G_k(t) : t \in \mathbb{Z}_*\}$. Since for every $i_0, i_1, \ldots, i_{t+1} \in \{0, 1, \ldots, m\}$ the process $\{G_k(t)\}$ satisfies

$$\Pr\{G_k(t + 1) = i_{t+1} \mid G_k(t) = i_t, \ldots, G_k(0) = i_0\}$$
$$= \Pr\{G_k(t + 1) = i_{t+1} \mid G_k(t) = i_t\},$$

the process $\{G_k(t)\}$ is a Markov chain. The formal definition of the Markov chain for gene frequency is given as follows.

Definition 11 (Markov Chain for Gene Frequency): In the Markov chain $\{G_k(t)\}$ for gene frequency at locus $k \in \{1, \ldots, l\}$ in GAs,

1) the *state* is defined as the number of allele 1 at locus k in the population; thereby the state space is $\{0, 1, \ldots, m\}$. A state i in $\{G_k(t)\}$ gives the gene frequency

$$p_k = \frac{i}{m}.$$

2) The *transition matrix* of $\{G_k(t)\}$ is defined as $\mathbf{P} \overset{\text{def}}{=} (\rho_{ij})$, where ρ_{ij} is the *transition probability* of state i to state j.

$$\rho_{ij} \overset{\text{def}}{=} \Pr\{G_k(t + 1) = j \mid G_k(t) = i\}.$$

The previous section has shown how selection, OB-Scan, and mutation affect gene frequency. From those formulae, we can derive the transition probabilities of the Markov chain $\{G_k(t)\}$ for the evolution of gene frequency in MPGA.

Theorem 1: For a GA using selection with selection intensity $\mathcal{I}$, n-parent OB-Scan, and bit-flip mutation with γ_m, the transition probability ρ_{ij} of the Markov chain $\{G_k(t)\}$ corresponding to that GA is

$$\rho_{ij} = B(j; m, p_k')$$

with

$$p'_k = I_{\frac{i}{m}+\mathcal{I}\sqrt{\frac{i(m-i)}{m^2 l}}}\left(\left\lceil\frac{n}{2}\right\rceil,\left\lceil\frac{n}{2}\right\rceil\right)$$
$$+ \gamma_m\left(1-2I_{\frac{i}{m}+\mathcal{I}\sqrt{\frac{i(m-i)}{m^2 l}}}\left(\left\lceil\frac{n}{2}\right\rceil,\left\lceil\frac{n}{2}\right\rceil\right)\right)$$

Proof: The state i of transition probability ρ_{ij} suggests the gene frequency $p_k = \frac{i}{m}$. Given this frequency p_k, from eq. (7), Lemma 4, and Lemma 5, we can compute the gene frequency p'_k of the offspring reproduced by the predefined GA:

$$p'_k = p_k^x + \gamma_m(1-2p_k^x)$$
$$= I_{p_k^s}\left(\left\lceil\frac{n}{2}\right\rceil,\left\lceil\frac{n}{2}\right\rceil\right) + \gamma_m\left(1-2I_{p_k^s}\left(\left\lceil\frac{n}{2}\right\rceil,\left\lceil\frac{n}{2}\right\rceil\right)\right)$$
$$= I_{\frac{i}{m}+\mathcal{I}\sqrt{\frac{i(m-i)}{m^2 l}}}\left(\left\lceil\frac{n}{2}\right\rceil,\left\lceil\frac{n}{2}\right\rceil\right)$$
$$+ \gamma_m\left(1-2I_{\frac{i}{m}+\mathcal{I}\sqrt{\frac{i(m-i)}{m^2 l}}}\left(\left\lceil\frac{n}{2}\right\rceil,\left\lceil\frac{n}{2}\right\rceil\right)\right)$$

In generational GAs, population is completely replaced with the subpopulation consisting of m offspring reproduced by m times of *selection-crossover-mutation* process. Since this process is independent, the number of allele 1 holds a binomial distribution $B(m, p'_k)$. Therefore, the transition probability

$$\rho_{ij} = \Pr\{G_k(t+1) = j \mid G_k(t) = i\}$$
$$= B(j; m, p'_k).$$

∎

A Markov chain consists of a family of random variables; therefore it has a probability distribution of states for each random variable. In terms of Markov processes, this probability distribution is called *state distribution*.

Definition 12 (State Distribution):

1) Denoted by a probability vector $\vec{\pi}(t) = (\pi_0(t), \ldots, \pi_N(t))$, the *state distribution* of a Markov chain $\{X(t)\}$ with state space $S = \{0, \ldots, N\}$ represents the unconditional probabilities of states. Precisely, for $t \in \mathbb{Z}_*$ and $i \in S$

$$\pi_i(t) \stackrel{\text{def}}{=} \Pr\{X(t) = i\} \quad \text{and} \quad \sum_{i \in S}\pi_i(t) = 1.$$

2) The vector $\vec{\pi}(0)$ is called the *initial distribution* of $\{X(t)\}$.

Next, we introduce the time homogeneity of Markov chains. The state distribution of time homogeneous Markov chain can be computed by a direct way given in the following theorem.

Definition 13 (Time Homogeneous): A Markov chain is said to be *time homogeneous* if its transition probabilities are independent of time. That is, for all $t \in \mathbb{Z}_*$

$$\rho_{ij} = \Pr\{X(t+1) = j \mid X(t) = i\}.$$

Theorem 2 ([6, pp.57]): Suppose we have a time homogeneous Markov chain with transition matrix $\mathbf{P}$. Given the initial distribution $\vec{\pi}(0)$, the state distribution at generation t can be obtained by

$$\vec{\pi}(t) = \vec{\pi}(0)\mathbf{P}^t. \tag{10}$$

Proposition 1: The Markov chain $\{G_k(t)\}$ for the GA given in Theorem 1 is time homogeneous.

Proof: According to Theorem 1, the transition probability ρ_{ij} of the Markov chain $\{G_k(t)\}$ is independent of time t. Therefore, the Markov chain $\{G_k(t)\}$ is time homogeneous. ∎

Now we can utilize (10) to compute the state distribution and further the mean fitness at generation t.

Theorem 3: For the GA defined in Theorem 1, given the state distribution $\vec{\pi}(t)$, we have the mean fitness $\bar{f}(t)$ and variance $\sigma_F(t)$ in the OneMax problem at generation t:

$$\bar{f}(t) = \frac{l}{m}\sum_{j=0}^{m}j\pi_j(t)$$

$$\sigma_F^2(t) = \frac{l}{m^2}\sum_{j=0}^{m}j(m-j)\pi_j(t)$$

Proof: Given the state distribution $\vec{\pi}(t)$ at generation t. The gene frequency $p_k(t)$ at any locus $k \in \{1, \ldots, k\}$

$$p_k(t) = \sum_{j=0}^{m}\left(\frac{j}{m}\right)\Pr\{G_k(t) = j\} = \frac{1}{m}\sum_{j=0}^{m}j\pi_j(t).$$

With the above equation, from Lemma 2 we obtain the mean fitness at generation t

$$\bar{f}(t) = \sum_{k=1}^{l}p_k(t) = \frac{l}{m}\sum_{j=0}^{m}j\pi_j(t),$$

and the variance

$$\sigma_F^2(t) = \sum_{k=1}^{l}\left(p_k(t)-p_k^2(t)\right) = \frac{l}{m^2}\sum_{j=0}^{m}j(m-j)\pi_j(t).$$

∎

B. Convergence

Theorem 1 gave the transition matrix of the Markov chain $\{G_k(t)\}$ for the GA using selection with selection intensity $\mathcal{I}$, n-parent OB-Scan, and bit-flip mutation with mutation rate γ_m. Of particular interest to us is, if at all, the *convergence* of $\{G_k(t)\}$ — at that time the state distribution will stay invariant once the Markov chain achieves it. Such a state distribution is referred to as *stationary distribution* (or *equilibrium distribution*) and this chain is called *stationary* [6]. In this section we will show that the MPGA given in Theorem 1 has this convergence property. Furthermore, the mean convergence fitness in the OneMax problem will be derived.

Definition 14 (Stationary Distribution): A state distribution $\vec{\pi}$ is said to be the *stationary distribution* for the Markov chain with transition matrix $\mathbf{P}$, if

$$\vec{\pi}\mathbf{P} = \vec{\pi}. \tag{11}$$

Lemma 6 (Ergodicity [4, pp.50]): A Markov chain is *ergodic* if and only if

$$\exists t \in \mathbb{Z}_+ : \mathbf{P}^t > \mathbf{0},$$

where $\mathbf{0}$ denotes *zero matrix*, i.e. all components in $\mathbf{0}$ are zeros.

Theorem 4 (Existence of Stationarity [4, pp.50]): For an ergodic Markov chain with state space $\{0, 1, \ldots, m\}$,

1) the powers $\mathbf{P}^t$ of the transition matrix $\mathbf{P}$ converge componentwise to a matrix whose all rows are equal. If we denote a typical row by $\vec{\pi} = (\pi_0, \ldots, \pi_m)$, then we have $\pi_i > 0$ and $\sum_i \pi_i = 1$ for all $i \in \{0, \ldots, m\}$.
2) $\vec{\pi}$ is the unique vector such that $\vec{\pi}\mathbf{P} = \vec{\pi}$.

This unique $\vec{\pi}$ is called the *stationary distribution* associated with the chain.

To prove the existence of convergence for the Markov chain $\{G_k(t)\}$, we need to show this chain is ergodic.

Proposition 2: With mutation rate $0 < \gamma_m < 1$, the Markov chain corresponding to the GA given in Theorem 1 is ergodic.

Proof: Considering the case $t = 1$, from Theorem 1 we have

$$\rho_{ij} = 0 \iff B(j; m, p'_k) = 0$$

with

$$p'_k = p^x_k + \gamma_m(1 - 2p^x_k). \tag{12}$$

The above binomial distribution equals zero if and only if the probability $p'_k = 0$ or 1. In case of $0 < \gamma_m < 1$, the solutions for $p'_k = 0$ or $p'_k = 1$ in (12) are

$$p^x_k < 0 \quad \text{or} \quad p^x_k > 1.$$

Both solutions violate the Probability Axiom for p^x_k, that is, $0 \le p^x_k \le 1$. This implies for all $i, j \in \{0, \ldots, m\}$

$$\rho_{ij} > 0.$$

According to Theorem 4, the chain $\{G_k(t)\}$ is proven to be ergodic. ∎

Theorem 5: Suppose we have the GA defined in Theorem 1 with $0 < \gamma_m < 1$. Let $\{G_k(t)\}$ be the corresponding Markov chain at locus $k \in \{1, \ldots, l\}$. We have the stationary distribution $\vec{\pi} = (\pi_1, \ldots, \pi_l)$, which gives for all $j \in \{0, \ldots, m\}$

$$\lim_{t \to \infty} \Pr\{G_k(t) = j\} = \pi_j.$$

Proof: Proposition 2 shows the Markov chain $\{G_k(t)\}$ is ergodic. According to Theorem 4, there exists the unique stationary distribution $\vec{\pi}$ giving for all $j \in \{0, \ldots, m\}$

$$\pi_j = \lim_{t \to \infty} \Pr\{G_k(t) = j\}.$$

∎

Corollary 2 (Mean Convergence Fitness): For the GA defined in Theorem 1 with $0 < \gamma_m < 1$, the mean and variance of convergence fitness in the OneMax problem are

$$\bar{f}^* = \frac{l}{m} \sum_{j=0}^{m} j\pi_j$$

$$\sigma_F^{2*} = \frac{l}{m^2} \sum_{j=0}^{m} j(m-j)\pi_j$$

Proof: Substitute stationary distribution $\vec{\pi}$ given in Theorem 5 into the equations in Theorem 3; then we complete the proof. ∎

V. THEORETICAL RESULTS AND EXPERIMENTAL VALIDATION

This section demonstrates the theoretical results obtained from the above theorems. Moreover, we conduct several experiments on the OneMax problem to verify our theoretical arguments. The setting of MPGA used in our experiments is generational GA, bit-string representation, population size $m = 128$, OB-Scan, bit-flip mutation, and 1000 generations for terminal condition. The size of OneMax problem is set to be 100 bits, i.e. $l = 100$. Each experiment setting includes 100 independent runs.

In this paper we adopt linear ranking selection [14] as the selection operator. The linear ranking selection, as its name tells, assigns the probability for a chromosome to be selected according to a linear formula based on the rank of this chromosome. Let $i \in \{1, \ldots, m\}$ be the rank (1 for the worst chromosome and m for the best one), the probability

$$\Pr\{\vec{c}_i \text{ to be selected}\} = \frac{1}{m}\left[\eta^- + (\eta^+ - \eta^-)\frac{i-1}{m-1}\right],$$

where $1 \le \eta^+ \le 2$ and $\eta^- = 2 - \eta^+$ are two parameters used to control the linear relation [3]. Precisely, $\frac{\eta^-}{m}$ and $\frac{\eta^+}{m}$ are the expected probabilities for the worst and the best chromosome to be selected, respectively. For the linear ranking selection, Blickle and Thiele [5] derived its selection intensity. In the following text we will use their formula to compute the selection intensity of linear ranking selection:

$$\mathcal{I}_{\mathrm{R}} = (1 - \eta^-)\frac{1}{\sqrt{\pi}} = (\eta^+ - 1)\frac{1}{\sqrt{\pi}}. \tag{13}$$

Figure 2 compares the variation of mean fitness computed by Theorem 3 with the mean fitness averaged over 100 experiments. First, this figure shows the theoretical and the experimental results fit very well. Second, we can see the profiles of OB-Scan with even number n are very close to those of OB-Scan with the number $n-1$. This confirms the *pairwise equivalence* claimed in Corollary 1. In addition, for mutation rates $\gamma_m \ge 0.05$, OB-Scan with $n > 2$ achieves higher fitness and even faster convergence than OB-Scan with $n = 2$, viz uniform crossover. Nevertheless, there is no winner for all these four mutation rates; the best number of parents is dependent upon the mutation rate. It is noteworthy that the profiles for 7- and 8-parent OB-Scan at $\gamma_m = 0.2$ continues climbing after 20 generations while others turn into steady then. This condition also occurs for 3- and 4-parent OB-Scan at $\gamma_m = 0.1$. Although the mutation rates in these two cases are relatively strong to the common setting $\gamma_m = \frac{1}{l}$, they keep the MPGA using OB-Scan on advancing in fitness. These results indicate the important role of mutation in the performance of MPGA using OB-Scan.

Furthermore, we examine the solution quality for 1000 generations. Figure 3 depicts the experimental and the theoretical fitness means in 1000 generations ($t = 1000$); additionally it plots the theoretical mean convergence fitness ($t \to \infty$) according to Corollary 2. First, we can see that the theoretical and the experimental results fit very well except two cases:
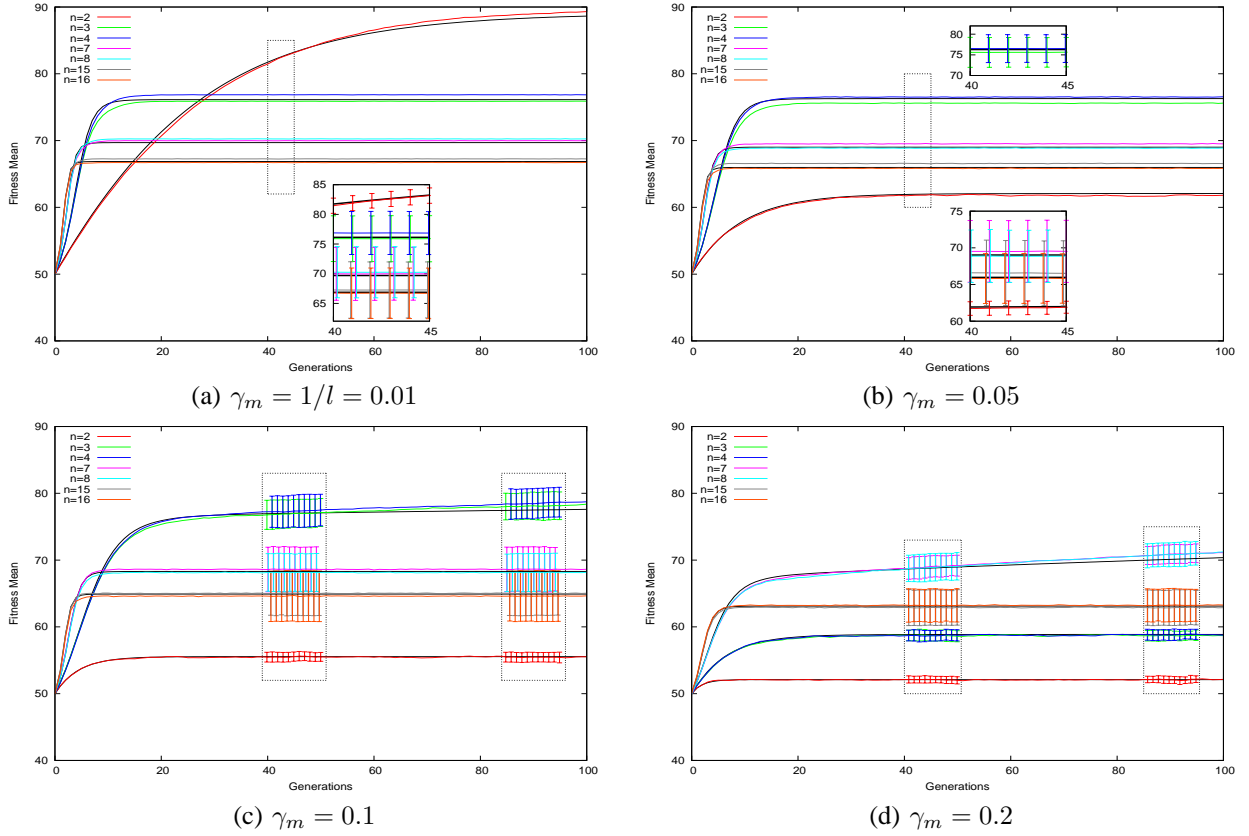
Fig. 2. Comparison of mean fitness obtained from theory (black bold lines) and experiment (color thin lines) for the MPGA using linear ranking selection with $\eta^+ = 1.5$, n-parent OB-Scan, and bit-flip mutation with $\gamma_m = 0.01, 0.05, 0.1,$ and 0.2

$n = 3, 4$ at $\gamma_m = 0.1$ and $n = 9, 10$ at $\gamma_m = 0.2$. In these two cases experimental fitness is higher than theoretical one. We reason this discrepancy results from violation of the assumption $\mathrm{Var}(\sigma) \ll \mathrm{E}^2[\sigma]$ in the approximation of the intensity $\mathcal{I}_k^{\mathrm{p}}$. Once this assumption does not hold, eq. (5) tells that the consequent gene frequency p_k^s will be higher than the value computed by (7). In the OneMax problem it leads to a higher fitness, which is reflected in the discrepancy between theoretical and experimental fitness in those two cases. Second, as shown in Fig. 3, the closeness in mean fitness between n-parent and $(n-1)$-parent OB-Scan for $n \in 2\mathbb{N}$ reconfirms the *pairwise equivalence*. Third, in comparison of the experimental mean fitness at $t = 1000$, the GA using 2-parent OB-Scan performs best at $\gamma_m = 0.01$ but does worst at $\gamma_m \geq 0.05$. Concerning the influence of mutation, the mean fitness for 2-parent OB-Scan decreases monotonically with the increase of mutation rate. Contrary, experimental results at $t = 1000$ show that putting mutation rate up may improve the GA using more than two parents in OB-Scan. However, as $t \to \infty$, the mean fitness for OB-Scan with more than two parents turns to decrease with mutation rate absolutely, which is the same tendency with 2-parent OB-Scan. Fig. 3 further points out that, in theory, all the GAs using OB-Scan with more than two parents will converge to higher fitness than the GA using 2-parent does. Yet, there exists a gap in fitness

between $t = 1000$ and $t \to \infty$. A reason for this gap is that our prescribed terminal condition is not long enough for some of the test GAs to converge.

VI. CONCLUSIONS

This paper proposes a Markov model to investigate the convergence of multi-parent genetic algorithms. Specifically, we probe into the mean fitness of the GA using *occurrence based scanning crossover* (OB-Scan) — a multi-parent generalization of *uniform crossover*.

The theoretical results demonstrate several interesting points. First, the analysis reveals the *pairwise equivalence* in OB-Scan. That is, for an even number n, the performance of n-parent OB-Scan equals to that of $(n-1)$-parent OB-Scan. Second, it manifests the critical role of mutation in the performance of MPGAs using OB-Scan. In the OneMax problem, the theoretical results show that the MPGA using more parents in OB-Scan together with mutation is capable of higher mean convergence fitness than the GA using 2-parent OB-Scan, viz uniform crossover. All these theoretical results are examined by a series of experiments on the OneMax problem. It shows that the theoretical and the experimental results fit very well. This consistency validates the capability of our proposed model. Furthermore, the superiority of n-parent $(n > 2)$ OB-Scan over 2-parent OB-Scan in the OneMax problem is verified theoretically as well as empirically.
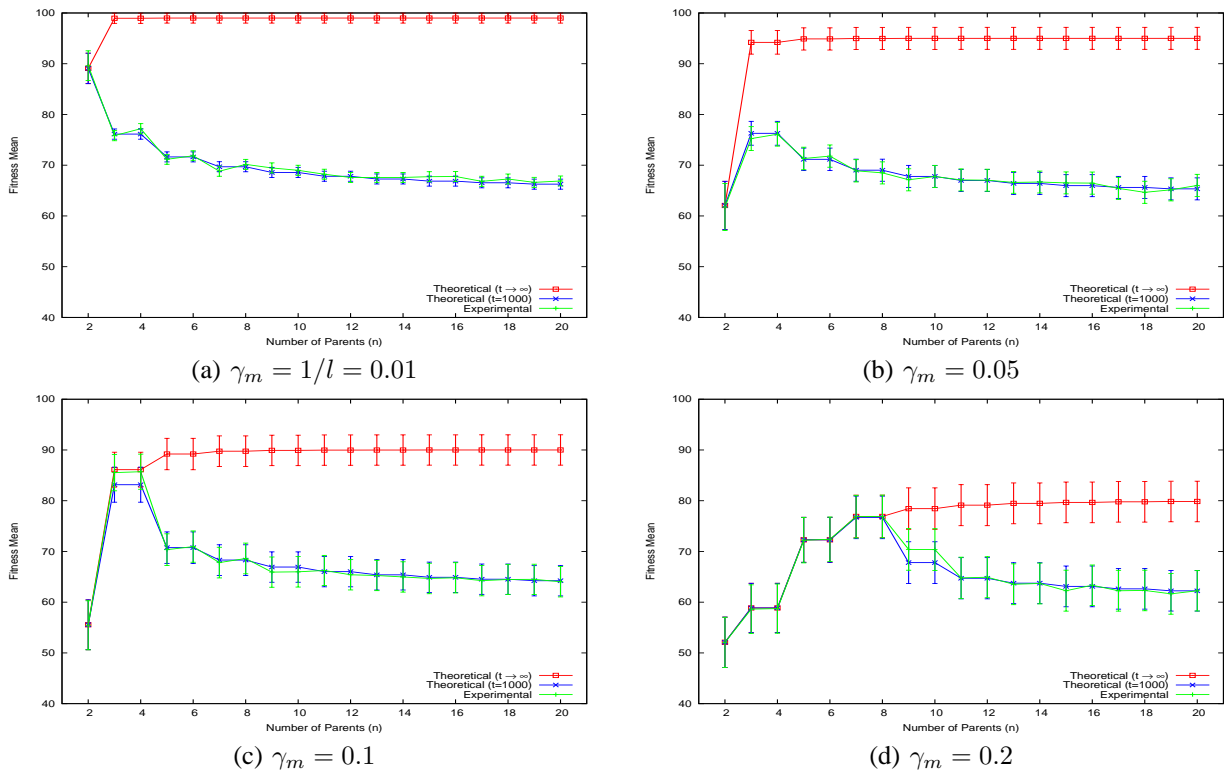
Fig. 3. The mean fitness of the MPGA using linear ranking selection with $\eta^+ = 1.50$, n-parent OB-Scan, and mutation with $\gamma_m = 0.01, 0.05, 0.1,$ and 0.2

The proposed model established a theory concerning the individual as well as the integral influence of multi-parent crossover with selection and mutation on the performance of MPGA. It also showed the power and the limit of OB-Scan, and manifested the key role of mutation in this crossover. However, this analysis focuses only on the MPGA using OB-Scan in the OneMax problem. An extension to other crossovers and problem domains remains to be investigated in the future.

REFERENCES

[1] M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1972. ninth Dover printing.

[2] H. Asoh and H. Mühlenbein. On the mean convergence time of evolutionary algorithms without selection and mutation. In *Parallel Problem Solving from Nature – PPSN III*, volume 866 of *LNCS*, pages 88–97, Berlin, 1994. Springer.

[3] T. Bäck. Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, pages 57–62, 1994.

[4] E. Behrends. *Introduction to Markov Chains*. Vieweg, 2000.

[5] T. Blickle and L. Thiele. A comparison of selection schemes used in genetic algorithms. TIK-Report 11, TIK Institut fur Technische Informatik und Kommunikationsnetze, Computer Engineering and Networks Laboratory, ETH, Swiss Federal Institute of Technology, 1995.

[6] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer Verlag, 1999.

[7] T.E. Davis and J.C. Principe. A markov chain framework for the simple genetic algorithm. *Evolutionary Computation*, 1(3):269–288, 1993.

[8] A.E. Eiben. Multiparent recombination in evolutionary computing. In *Advances in Evolutionary Computing*, pages 175–192. Springer, 2002.

[9] A.E. Eiben, E.H.L. Aarts, and K.M. Van Hee. Global convergence of genetic algorithms: A markov chain analysis. In *Parallel Problem*

Solving from Nature (PPSN 1), volume 496 of *LNCS*, pages 4–12. Springer-Verlag, 1991.

[10] A.E. Eiben and T. Bäck. An empirical investigation of multiparent recombination operators in evolution strategies. *Evolutionary Computation*, 5(3):347–365, 1997.

[11] A.E. Eiben, P.-E. Rau'e, and Zs. Ruttkay. Genetic algorithms with multiparent recombination. In *Parallel Problem Solving from Nature – PPSN III*, volume 866 of *LNCS*, pages 78–87. Springer, 1994.

[12] D.S. Falconer and T.F.C. Mackay. *Introduction to Quantitative Genetics*. Prentice Hall, 4 edition, 1996.

[13] D.E. Goldberg and P. Segrest. Finite markov chain analysis of genetic algorithms. In *Proceedings of the 2nd International Conference on Genetic Algorithms and their Applications*, pages 1–8. Lawrence Erlbaum Associates, 1987.

[14] J.J. Grefenstette and J.E. Baker. How genetic algorithms work: A critical look at implicit parallelism. In *Proceedings of the Third International Conference on Genetic Algorithms*, pages 20–27. Morgan Kaufmann Publishers, 1989.

[15] D.L. Hartl and A.G. Clark. *Principles of Poppulation Genetics*. Sinauer Associates, 1 edition, 1989.

[16] H. Mühlenbein. The equation for the response to selection and its use for prediction. *Evolutionary Computation*, 5(3):303–346, 1998.

[17] A.E. Nix and M.D. Vose. Modeling genetic algorithms with markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:78–88, 1992.

[18] Günter Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101, 1994.

[19] G. Syswerda. Uniform crossover in genetic algorithms. In *Proceedings of the 3rd International Conference on Genetic Algorithms*, pages 2–9. Morgan Kaufmann, 1989.

[20] D. Thierens, D.E. Goldberg, and Â.G. Pereira. Domino convergence, drift, and the temporal-salience structure of problems. In *Proceedings of 1998 IEEE International Conference on Evolutionary Computation*, pages 535–540, 1998.

[21] S. Tsutsui and A. Ghosh. A study on the effect of multi-parent recombination in real coded genetic algorithms. In *Proceedings of International Conference on Evolutionary Computation*, pages 828–833, 1998.