Combining Drift Analysis and Generalized Schema Theory to Design Efficient Hybrid and/or Mixed Strategy EAs

Boris Mitavskiy and Jun He

Abstract

Hybrid and mixed strategy EAs have become rather popular for tackling various complex and NP-hard optimization problems. While empirical evidence suggests that such algorithms are successful in practice, rather little theoretical support for their success is available, not mentioning a solid mathematical foundation that would provide guidance towards an efficient design of this type of EAs. In the current paper we develop a rigorous mathematical framework that suggests such designs based on generalized schema theory, fitness levels and drift analysis. An example-application for tackling one of the classical NP-hard problems, the "single-machine scheduling problem" is presented.

I. INTRODUCTION

In recent years hybrid and/or mixed strategy EAs are frequently applied to tackle various NP-hard optimization problems. Such algorithms exploit a variety of different recombination, mutation, and selection operators where these operators are chosen with distinct probabilities depending on the current population, the individuals that are selected for recombination or mutation, and, sometimes, the time when the population arises. According to the empirical evidence, many such algorithms are rather successful in practice (see, for instance, [1] and [2] and [3]). At the same time, rather little theoretical support or, the more so, general guidance for the design of such EAs, exists in the literature. This work is largely motivated by a special case design of a hybrid $1 + 1$ EA on a single machine scheduling problem in [4], however, in this paper, we aim to investigate possible methodology for the design of population-based EAs of this type. In the current article we unify the theory of hybrid and mixed strategy EAs into a common mathematical framework. This opens the door to various existent and well-developed mathematical tools such as generalized schema theory, drift analysis and tail inequalities to design hybrid and mixed strategy EAs for various specific problems with polynomial runtime guarantees to encounter a satisfactory solution (such as a solution up to a desirable or allowable approximation ratio). The paper is organized in a straightforward fashion: In section II we set up a rigorous mathematical framework that incorporates a wide class of hybrid and mixed strategy EAs. Next, in section III, generalized schema theory is presented. In section IV the central idea of the article, namely the design of hybrid and mixed strategy EAs for specific optimization problems, based on the notion of auxiliary fitness levels and schemata is provided. These ideas can then be linked with classical tools from applied probability to analyze runtime complexity: for instance, in section V, drift analysis methodology has been applied to analyze the conditions under which the expected runtime bounds are polynomial. Finally, this approach is illustrated with a specific example application: designing a family of hybrid and mixed strategy population-based EAs for the "single machine scheduling problem" (see [5] and [4] for a detailed description) with expected polynomial time approximation ratio guarantees.

II. MATHEMATICAL DESCRIPTION OF HYBRID AND MIXED STRATEGY EAs

While the families of recombination, mutation and selection operators are typically independent of population and the iteration time at which the population is encountered, hybrid and mixed strategy EAs exploit several families of recombination, mutation and selection operators. Furthermore, each pair (P, t) where P is the population encountered at t^{th} iteration of a given hybrid or mixed strategy EA, is equipped with a probability distribution over the various families of recombination, mutation and selection transformations. Therefore, mathematically, mixed and hybrid EAs are fully determined in terms of the following parameters:

1. A finite set Ω of candidate solutions that we call the search space.

2. A tuple of indexed families

$$(\{\mathcal{F}_i\}_{i \in I}, \{\mathcal{M}_j\}_{j \in J}, Sel_q\}_{q \in Q}, \{f_l\}_{l \in L \cup \{0\}})$$

where I, J, Q and L are various indexing sets (usually finite subsets of $\mathbb{N}$) while $\mathcal{F}_i$, $\mathcal{M}_j$, Sel_q are various families of recombination, mutation and selection operators respectively and $\{f_l\}_{l \in L}$ is the family of fitness functions. Among the fitness functions, only the function f_0 is the objective function to be optimized. We will say that the remaining fitness functions are *auxiliary fitness functions*. In practice these may be given implicitly, motivated by certain deterministic algorithms (such as the

Boris Mitavskiy and Jun He are with Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, UK.

Jackson rule in case of the single machine scheduling problem). Recombination operators are usually maps $F : \Omega^2 \to \Omega$ that take a pair of individuals $(x, y) \in \Omega^2$ to a single offspring $F(x, y) \in \Omega,$[1] while mutation operators are functions $M : \Omega \to \Omega$. A selection operator is a function $Sel : (\Omega^m)^2 \to \Omega^m$ on the set of pairs of populations of size m such that $\forall$ two populations $\vec{x} = (x_1, x_2, \ldots x_m)$ and $\vec{y} = (y_1, y_2, \ldots y_m) \in \Omega^m$ all the individuals of the population $Sel(\vec{x}, \vec{y})$ are also the individuals of the population $\vec{x}$ or of the population $\vec{y}$: $Sel(\vec{x}, \vec{y}) = (z_1, z_2, \ldots z_m) \in \Omega^m$ and $\forall i$ with $1 \le i \le m$ $\exists j$ with $1 \le j \le m$ such that $z_i = x_j$ or $z_i = y_j$. It is also reasonable to assume that given any subset $S \in \Omega$ such that all the individuals in S appear either in the population $\vec{x}$ or $\vec{y}$, then the number of individuals from the subset S that appear in the population $Sel(\vec{x}, \vec{y})$ does not depend on the specific location (indexing) of the elements from S in the populations $\vec{x}$ and $\vec{y}$. Most certainly it may depend on all other parameters such as the fitness of the various individuals in S and even on whether or not these individuals occur in the population $\vec{x}$ or in the population $\vec{y}$.

3. To every pair $(\vec{x}, t)$ where $\vec{x} = (x_1, x_2, \ldots x_m) \in \Omega^m$ is a population at t^{th} iteration of the algorithm and to every pair of individuals (x_i, x_j) in $\vec{x}$ assign probability distributions $Pr^{\text{RecFamily}}_{(\vec{x}, t), (x_i, x_j)}$ on the set indexing the families of recombination transformations. and for every index $w \in I$, a probability distribution $Pr^{\text{Rec}, w}_{(\vec{x}, t), (x_i, x_j)}$ on the set of recombination transformations $\mathcal{F}_w$. For simplicity we shall assume that the pairs (x_i, x_j) are sampled uniformly at random (either with replacement or without replacement) from the population $\vec{x}$. Once a pair (x_i, x_j) has been selected for recombination, it first selects a family of recombination operators to use according to the probability distribution $Pr^{\text{RecFamily}}_{(\vec{x}, t), (x_i, x_j)}$ and then, once the index w has been chosen, it selects a specified transformation to use according to the probability distribution $Pr^{\text{Rec}, w}_{(\vec{x}, t), (x_i, x_j)}$ on the family of recombination transformations $\mathcal{F}_w$. Mutation operators are selected analogously except that this time only a single individual, say x_i, is selected uniformly at random from the population $\vec{x}$ and selects a family of mutation transformations $\mathcal{M}_w$ according to a probability distribution $Pr^{\text{MutFamily}}_{(\vec{x}, t), x_i}$ on the indexing set J of the families of mutation transformations. Afterwards, it selects mutation transformations from the family $\mathcal{M}_w$ according to the probability distribution $Pr^{\text{Mut}, w}_{(\vec{x}, t), x_i}$ on the family of mutation transformations $\mathcal{M}_w$. Likewise, to every pair $(\vec{x}, t)$ and a population $\vec{y} \in \Omega^m$ we associate a probability distribution $Pr^{\text{SelFamily}}_{(\vec{x}, t), \vec{y}}$ on the indexing set Q of the families of selection transformations, and to every family of selection transformations Sel_w we assign a probability distribution $Pr^{\text{Sel}, w}_{(\vec{x}, t), \vec{y}}$ on the family Sel_w of selection transformations. Once an "intermediate" population $\vec{y}$ has been obtained from the population $\vec{x}$ upon completion of recombination stage followed by mutation stage, an appropriate family of selection transformations S_w is selected through sampling its index via the probability distribution $Pr^{\text{SelFamily}}_{(\vec{x}, t), \vec{y}}$. Afterwards, an appropriate selection transformation Sel is chosen from the family of selection transformations Sel_w via the probability distribution $Pr^{\text{Sel}, w}_{(\vec{x}, t), \vec{y}}$.

A hybrid/mixed strategy EA cycles through the recombination, mutation and selection stages sufficiently long to encounter a satisfactory solution. A single cycle consisting of these three consecutive stages is typically called a single iteration of the algorithm that produces the next generation from the previous one.

III. GENERALIZED SCHEMA THEORY

In the current section we will establish a very general version of the schema theorem that applies to the types of EAs fitting the framework in the previous section. Suppose we are given any subset $S \subseteq \Omega$ of the search space of an EA and a population $\vec{x}$ at t^{th} iteration of the algorithm. Recall from the previous section that various individuals from the population $\vec{x}$ are paired up for recombination independently m times with the aim of producing exactly m offsprings. Thus, the probability that an individual from the set S appears at the ith position of the "intermediate" population $\vec{x}^{\text{rec}}$ obtained from the population $\vec{x}$ upon completion of recombination can be computed as follows:

Lemma III.1. *Continuing with the notation in the preceding paragraph, for a given pair of individuals $(x_i, x_j) \in \vec{x}^2$, let*

$$Pr(S \,|\, (x_i, x_j)) = \sum_{w \in I} \sum_{F \in \mathcal{F}_w}^{F(x_i, x_j) \in S} Pr^{\text{RecFamily}}_{(\vec{x}, t), (x_i, x_j)}(i = w) \cdot Pr^{\text{Rec}, w}_{(\vec{x}, t), (x_i, x_j)}(F).$$

Furthermore, let

$$Pr^{nonRepl}(S \,|\, \vec{x}, t) = \frac{\sum_{i \ne j} Pr(S \,|\, (x_i, x_j))}{m(m - 1)}$$

and

$$Pr^{Repl}(S \,|\, \vec{x}, t) = \frac{\sum_{(i, j) \in \{1, 2, \ldots, m\}} Pr(S \,|\, (x_i, x_j))}{m^2}$$

Then, in case when pairs of individuals are selected for recombination independently without replacement, $\forall i$ with $1 \le i \le m$, the probability that the ith individual in the intermediate population $\vec{x}^{rec}$ is in the set S is $Pr^{nonRepl}(S \,|\, \vec{x}, t)$. Likewise, in case when pairs of individuals are selected for recombination independently with replacement, this probability is $Pr^{Repl}(S \,|\, \vec{x}, t)$.

---

[1]One may also allow general m-ary operators. In case when $m = 1$, i.e. when the recombination operators are unary, they are usually known as mutation operators.

Proof: According to the general framework in the previous section, if the individuals x_i and x_j have been selected for recombination, $Pr(S \,|\, (x_i, x_j))$ is the probability that the offspring individual is an element of the subset $S \subseteq \Omega$ is precisely $Pr(S \,|\, (x_i, x_j))$. Since every pair is selected for recombination uniformly at random, the probability that the pair (x_i, x_j) has been chosen for recombination and their offspring is in S is $\frac{Pr(S \,|\, (x_i, x_j))}{m(m-1)}$ in case of sampling pairs without replacement and $\frac{Pr(S \,|\, (x_i, x_j))}{m^2}$ in case of sampling with replacement. The desired conclusion follows now by summing the probabilities of pairwise disjoint events. ∎

Since recombination takes place independently, the number of individuals in the intermediate population $\vec{x}^{\text{rec}}$ is distributed binomially with success probabilities

$Pr^{nonRepl}(S \,|\, \vec{x}, t)$ in case of sampling recombination pairs without replacement and $Pr^{Repl}(S \,|\, \vec{x}, t)$ in case of sampling recombination pairs with replacement respectively, Chernoff tail inequality (see, for instance, chapter 1 of [6]) applies and readily tells us the following.

Lemma III.2. *Continuing with the notation in lemma III.1, let $N(S, \vec{y})$ denote the random variable counting the total number of individuals in the population $\vec{y}$ of size m that are in the set S. Then $\forall \, \delta \in [0, 1]$*

$$Pr(N(S, \vec{x}^{\text{rec}}) < (1 - \delta)mP) \leq \exp\left(-\delta^2 \frac{mP}{2}\right)$$

while

$$Pr(N(S, \vec{x}^{\text{rec}}) > (1 + \delta)mP) \leq \exp\left(-\delta^2 \frac{mP}{3}\right)$$

where

$$P = \begin{cases} \dfrac{Pr(S \,|\, (x_i, x_j))}{m(m-1)} & \text{\textit{in case of sampling pairs}} \\ & \text{\textit{without replacement}} \\ \dfrac{Pr(S \,|\, (x_i, x_j))}{m^2} & \text{\textit{in case of sampling pairs}} \\ & \text{\textit{with replacement}} \end{cases}$$

In some simplified cases (such as one in the current paper), more informative bounds may be used when the total number $N(S, \vec{x}^{\text{rec}})$ is expected to be small but bigger than 1.

Lemma III.3. *Given a population $\vec{x} = (x_1, x_2, \ldots x_m)$ and a schema S, suppose $\exists$ a schema S_0 such that $\forall i$ and $j \in \{1, 2, \ldots m\}$, as long as $x_i \in S_0$, $Pr(S, \,|x_i, x_j) \geq \alpha$. Then $Pr(N(S, \vec{x}^{\text{rec}}) \geq 1) \geq \left(1 - \left(1 - \frac{N(S_0, \vec{x}^{\text{rec}})}{m}\right)^m\right)\alpha$.*

Proof: According to the assumption, a sufficient condition to obtain an individual fitting the schema S upon completion of recombination is to select an individual fitting the schema S_0 to be the first one in a recombination pair at least once after m consecutive trials and, afterwards, to apply an appropriate recombination transformation with probability at least as large as α. An individual fitting the schema S_0 is selected at least once with probability $1 - \left(1 - \frac{N(S_0, \vec{x}^{\text{rec}})}{m}\right)^m$ via considering the complementary event implying the desired conclusion. ∎

We now proceed to analyze the context of the next intermediate population $\vec{x}^{\text{mut}}$ obtained from the population

$$\vec{x}^{\text{rec}} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m)$$

upon completion of mutation. When applying mutation operator to an individual in position i (i.e. to $\hat{x}_i$), in order to obtain an individual from the set S in the ith position of the population $\vec{x}^{\text{mut}}$, we must select a mutation operator that sends the individual $\hat{x}_i$ to an element of the set S. This event happens with probability

$$Pr(S \,|\, x_i) = \sum_{w \in J} \sum_{\substack{M \in \mathcal{M}_w \\ M(x_i) \in S}} Pr^{\text{MutFamily}}_{(\vec{x}, t), \, x_i}(i = w) \cdot Pr^{\text{Mut}, \, w}_{(\vec{x}, t), \, x_i}(M). \tag{1}$$

The indices of the individuals in the population $\vec{x}^{\text{rec}}$ can be partitioned into two disjoint subsets:

$$S \cap \vec{x}^{\text{rec}} = \{i \,|\, \hat{x}_i \in S\} \text{ and } \overline{S} \cap \vec{x}^{\text{rec}} = \{i \,|\, \hat{x}_i \notin S\}. \tag{2}$$

We now introduce the following probabilities:

$$\underline{Pr^{mut}(S \,|\, S)} = \min\{Pr(S \,|\, x_i) \,|\, i \in S \cap \vec{x}^{\text{rec}}\} \tag{3}$$

to be the minimal probability of preserving the ith individual that is already in the set S upon completion of mutation and

$$\underline{Pr^{mut}(S \,|\, \overline{S})} = \min\{Pr(S \,|\, x_i) \,|\, i \in S \cap \vec{x}^{\text{rec}}\} \tag{4}$$

to be the minimal probability of mutating the ith individual that is not in the set S into one that is in S. Notice that the random variable $N(S, \vec{x}^{\text{mut}})$ measuring the total number of individuals in the population $\vec{x}^{\text{mut}}$ from the set S (recall that this random variable has been introduced in the statement of lemma III.2) is the sum of independent indicator random variables

$$
\mathcal{X}_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ individual of } \vec{x}^{\text{mut}} \in S \\ 0 & \text{otherwise:} \end{cases}
$$

$$
N(S, \vec{x}^{\text{mut}}) = \sum_{i=1}^{m} \mathcal{X}_i = \sum_{i \in S \cap \vec{x}^{\text{rec}}} \mathcal{X}_i + \sum_{i \in \overline{S} \cap \vec{x}^{\text{rec}}} \mathcal{X}_i. \tag{5}
$$

From the discussion preceding equation 1,

$$
E(\mathcal{X}_i) = Pr(\mathcal{X}_i = 1) = Pr(S \mid x_i)
$$

so that, by linearity of expectation, we have

$$
\begin{aligned}
E(N(S, \vec{x}^{\text{mut}})) &= \sum_{i=1}^{m} E(\mathcal{X}_i) = \sum_{i=1}^{m} Pr(S \mid x_i) \\
&= \sum_{i \in S \cap \vec{x}^{\text{rec}}} Pr(S \mid x_i) + \sum_{i \in \overline{S} \cap \vec{x}^{\text{rec}}} Pr(S \mid x_i) \\
&\geq \underline{Pr^{mut}(S \mid S)} \cdot |S \cap \vec{x}^{\text{rec}}| + \underline{Pr^{mut}(S \mid \overline{S})} \cdot |\overline{S} \cap \vec{x}^{\text{rec}}| \\
&= \underline{Pr^{mut}(S \mid S)} \cdot |S \cap \vec{x}^{\text{rec}}| + \underline{Pr^{mut}(S \mid \overline{S})} \cdot (m - |S \cap \vec{x}^{\text{rec}}|).
\end{aligned}
$$

In summary, we have deduced that if

$$
\mu = \underline{Pr^{mut}(S \mid S)} \cdot |S \cap \vec{x}^{\text{rec}}| + \underline{Pr^{mut}(S \mid \overline{S})} \cdot (m - |S \cap \vec{x}^{\text{rec}}|)
$$

then

$$
\mu \leq E(N(S, \vec{x}^{\text{mut}})). \tag{6}
$$

The classical Chernoff bound applies again now and tells us that $\forall\, \delta \in [0, 1]$

$$
Pr\left(N(S, \vec{x}^{\text{mut}}) < (1-\delta)\mu\right) \overset{\text{thanks to inequality 6}}{\leq} Pr\left(N(S, \vec{x}^{\text{mut}}) < (1-\delta) \cdot E(N(S, \vec{x}^{\text{mut}}))\right)
$$

$$
\leq \exp\left(-\frac{\delta^2}{2} \cdot E(N(S, \vec{x}^{\text{mut}}))\right) \leq \exp\left(-\frac{\delta^2}{2}\mu\right). \tag{7}
$$

Observe that $|S \cap \vec{x}^{\text{rec}}| = N(S, \vec{x}^{\text{rec}})$ (see lemma III.2), so that, according to lemma III.2, we can bound μ (see equation-definition preceding equation 6) below as follows: $\forall\, \epsilon \in [0, 1]$

$$
Pr(\mu \geq \underline{Pr^{mut}(S \mid S)} \cdot (1-\epsilon)mP + +\underline{Pr^{mut}(S \mid \overline{S})} \cdot (m - (1+\epsilon)mP)) \geq 1 - \exp\left(-\epsilon^2 \frac{mP}{2}\right) - \exp\left(-\epsilon^2 \frac{mP}{3}\right) \tag{8}
$$

where P is the average probability of obtaining an element in the set S upon completion of recombination as introduced in the statement of lemma III.2. Combining inequalities 7 and 8 we finally deduce the following lower bound on the probability of the number of occurrences of individuals from the set S occurring in the population $\vec{x}^{\text{mut}}$: Let

$$
\underline{\mu} = m\left(\underline{Pr^{mut}(S \mid S)} \cdot (1-\epsilon)P + \underline{Pr^{mut}(S \mid \overline{S})}(1 - P(1+\epsilon))\right) \tag{9}
$$

Then we have

$$
Pr\left(N(S, \vec{x}^{\text{mut}}) \geq (1-\delta)\underline{\mu}\right) \geq Pr\left(N(S, \vec{x}^{\text{mut}}) \geq (1-\delta)\mu \mid \mu \geq \underline{\mu}\right) \cdot Pr(\mu \geq \underline{\mu})
$$

$$
\overset{\text{via inequalities 7 and 8}}{\geq} \left(1 - \exp\left(-\frac{\delta^2}{2}\underline{\mu}\right)\right) \times \left(1 - \exp\left(-\epsilon^2 \frac{mP}{2}\right) - \exp\left(-\epsilon^2 \frac{mP}{3}\right)\right). \tag{10}
$$

We summarize ineqaulity 10 in the following lemma.

Lemma III.4. *Given a pair $(\vec{x}, t)$ where $\vec{x}$ is a population at t^{th} generation of an EA and any subset $S \in \Omega$, continuing with the notation in lemmas III.1 and III.2, as well as equation-definitions 3, 4 and 9, select a pair of small numbers $(\delta, \epsilon) \in [0, 1]^2$. Then the probability that the total number of individuals in the "intermediate" population obtained from the population $\vec{x}$ upon completion of recombination followed by mutation is above the threshold $(1-\delta)\underline{\mu}$ is at least*

$$
\left(1 - \exp\left(-\frac{\delta^2}{2}\underline{\mu}\right)\right) \times \left(1 - \exp\left(-\epsilon^2 \frac{mP}{2}\right) - \exp\left(-\epsilon^2 \frac{mP}{3}\right)\right).
$$

Once again, in simplified constructions such as one presented in the current preliminary work, the following alternative lemma is an immediate corollary of lemma III.3:

Lemma III.5. *Continuing with the assumptions of lemma III.3, suppose, in addition, that $\forall\, x_i \in S$ the probability $Pr(S \,|\, x_i) \geq \beta$. Then*

$$Pr(N(S, \vec{x}^{mut}) \geq 1) \geq \left(1 - \left(1 - \frac{N(S_0, \vec{x}^{rec})}{m}\right)^m\right) \alpha \cdot \beta.$$

The following generalized schema theorem is nearly a restatement of lemma III.4 that takes into account generalized selection as described in the previous section.

Theorem III.6. *Let $\vec{z}$ denote the population obtained from the population $\vec{x}$ upon completion of the recombination $\rightarrow$ mutation $\rightarrow$ selection cycle (equivalently, the population $\vec{z}$ is obtained from the populations $\vec{x}$ and $\vec{x}^{mut}$ after selection). Recall from the statement of lemma III.2 that the random variable $N(S, \vec{z})$ counts the total number of individuals from the set S that appear in the population $\vec{z}$. Repeat verbatim the first sentence of lemma III.4. Then $\forall\, n \in \{1, 2, \ldots, m\}$*

$$Pr(N(S, \vec{z}) \geq n)$$
$$\geq Pr\left(N(S, \vec{z}) \geq n \,|\, N(S, \vec{x}^{mut}) \geq (1-\delta)\underline{\mu}\right)\left(1 - \exp\left(-\frac{\delta^2}{2}\underline{\mu}\right)\right)\left(1 - \exp\left(-\epsilon^2 \frac{mP}{2}\right) - \exp\left(-\epsilon^2 \frac{mP}{3}\right)\right).$$

Applying the classical Markov inequality (see, for instance, [6]), we immediately deduce the following.

Corollary III.7. *Continuing with the notation and assumptions in theorem III.6, $\forall$ real $k \in \left[0, \frac{m}{(1-\delta)\underline{\mu}}\right]$*

$$E(N(S, \vec{z})) \geq \lceil k(1-\delta)\underline{\mu}\rceil \cdot \left(1 - \exp\left(-\frac{\delta^2}{2}\underline{\mu}\right)\right) \times \left(1 - \exp\left(-\epsilon^2 \frac{mP}{2}\right) - \exp\left(-\epsilon^2 \frac{mP}{3}\right)\right)$$
$$\times Pr\left(N(S, \vec{z}) \geq \lceil k(1-\delta)\underline{\mu}\rceil \,|\, N(S, \vec{x}^{mut}) \geq (1-\delta)\underline{\mu}\right).$$

The corresponding simplified schema theorem is a direct consequence of lemma III.5:

Corollary III.8. *Continuing with the notation and assumptions in theorem III.6,*

$$Pr(N(S, \vec{z}) \geq 1) \geq \alpha\beta Pr\left(N(S, \vec{z}) \geq 1 \,|\, N(S, \vec{x}^{mut}) \geq 1\right)\left(1 - \left(1 - \frac{N(S_0, \vec{x}^{rec})}{m}\right)^m\right).$$

IV. Recombination-invariant Subsets, Fitness Levels and Mutation-invariant Subsets

While theorem III.6 and corollary III.7 are valid for arbitrary subsets $S \subseteq \Omega$, it is not in vain that most notions of schemata (see, for instance, [7] and [8]) happen to be *recombination-invariant* subsets of the search space Ω as defined precisely below:

Definition IV.1. *Given a family of recombination transformations $\mathcal{F}$ on a search space Ω, a* recombination invariant subset *or, alternatively, a* generalized schema *with respect to the family of recombination transformations $\mathcal{F}$ is a subset $H \subseteq \Omega$ having the property that $\forall\, x$ and $y \in H$ and $\forall$ transformation $T \in \mathcal{F}$ the child $T(x, y) \in H$.*

General mathematical properties of recombination-invariant subsets have been studied by several authors: see, for instance, [9], [10], [11], [12], [13] and [14]. First of all, we list a few basic properties of families of recombination-invariant subsets (see [12] for a detailed exposition and an in-depth analysis of the relationship between the collections of recombination-invariant subsets of the search space and the corresponding families of recombination transformations).

Proposition IV.1. *Given any family of recombination transformations $\mathcal{F}$ on the search space Ω, the corresponding family of invariant subsets with respect to the family $\mathcal{F}$, call it $Sel_\mathcal{F}$, is closed under arbitrary intersections, contains the $\emptyset$ and the whole search space Ω^2. Furthermore, given any collection of subsets $\mathcal{S} \subseteq \mathcal{P}(\Omega)$ of the search space Ω, the family $\underline{\mathcal{S}} = \{\bigcap_{S \in \mathcal{T}} \,|\, \mathcal{T} \subseteq \mathcal{S}\} \cup \{\emptyset, \Omega\}$ is closed under arbitrary intersections, contains the $\emptyset$ and the entire space, and $\exists$ a family of recombination transformations $\mathcal{F}$ such that $\mathcal{S}_\mathcal{F} = \underline{\mathcal{S}}$. Consequently, the union of all families of recombination transformations with the above property is the unique maximal (in the sense of containment) family of recombination transformations, call it $\widetilde{\mathcal{F}}$, such that $\mathcal{S}_{\widetilde{\mathcal{F}}} = \underline{\mathcal{S}}$. We will say that the collection of recombination-invariant subsets $\underline{\mathcal{S}}$ is* generated *by the collection of subsets $\mathcal{S}$ or, alternatively, that the collection of subsets $\mathcal{S}$* generates *the collection of recombination invariant subsets of the search space $\underline{\mathcal{S}}$.*

The correspondence summarized in proposition IV.1 is known in mathematics as a Galois connection[3]. One of the central ideas of the current article is that recombination transformations should be designed based on the suitable families of recombination-invariant subsets and below we will suggest how such families of recombination-invariant subsets may be selected to design efficient algorithms. This design is largely based on the notion of a fitness level introduced below.

---

[2] such collections of subsets are also known as pre-topologies: see [10] and [11]

[3] see [15] for the notions of natural transformations, adjunctions and Galois connections i.e. natural transformations between posets considered as categories. No knowledge of category theory is necessary to understand the current paper though.

Definition IV.2. *Given a fitness function $f : \Omega \to [0, M]$, the kth fitness level of f is the pre-image*

$$f^{-1}([k,\, M]) = \{\omega \in \Omega \mid f(\omega) \geq k\}.$$

Recall from section II that hybrid and mixed strategy EAs may have a large number of auxiliary fitness functions. The auxiliary fitness functions are often defined implicitly in terms of a certain incremental deterministic algorithm to find a satisfactory solution for a specific NP-hard optimization problem. We impose the following conditions on our Hybrid or mixed strategy EA:

Condition 1. The total number of auxiliary fitness functions is bounded above by a polynomial of degree ρ in the size of the problem instance (in other wards, $L = O(n^\rho)$ where L is the indexing set of the auxiliary fitness functions as in section II and n is the size of the problem instance.

Condition 2. All of the auxiliary fitness functions are non-negative, integer valued[4] and have a common range[5] and there are polynomially many auxiliary fitness levels: $\forall l \in L\ f_l : \Omega \to \{0, 1, \ldots, M\}$ and $M \leq O(n^\tau)$ where n is the size of an instance of a specific optimization problem.

Condition 3. $\exists l \in L$ such that the Mth fitness level of the auxiliary fitness function f_l consists of "satisfactory solutions" (for instance, up to a specified approximation ratio) for the objective fitness function f_0.

There is a number of ways to design the collections of recombination invariant subsets based on the fitness levels of various auxiliary fitness functions to guarantee that the expected time (i.e. the expected number of iterations) an EA requires to encounter a satisfactory solution is polynomial in the size of the input instance. Since the aim of the current paper is to illustrate the general ideas for such designs, we present what is, perhaps, one of the simplest and the shortest methodologies. For $k \in \{1, 2, \ldots M\}$, let $S_k = \bigcup_{l \in L} f_l^{-1}([k, M])$ and let $\mathcal{S}_k = \{S_j \mid j \geq k\} \cup \{\emptyset, \Omega\}$. Observe that the collection $\mathcal{S}_k$ of the unions of fitness levels at least as high as k is a collection of nested sets so that, in particular, it is closed under arbitrary intersections. According to proposition IV.1 we may select families of recombination transformations $\mathcal{F}_k$ such that the corresponding families of invariant subsets $\mathcal{S}_{\mathcal{F}_k} = \mathcal{S}_k$. In fact, all that we require is that $\mathcal{S}_{\mathcal{F}_k} \supseteq \mathcal{S}_k$ i.e. that the family of recombination transformations $\mathcal{F}_k$ preserves the unions of lth fitness levels across all of the auxiliary fitness functions.

We now turn our attention to mutation transformations. Invariant subsets for mutation transformations are defined in the same fashion.

Definition IV.3. *Given a family $\mathcal{M}$ of mutation transformations, a subset $S \in \Omega$ is invariant under the family of mutation transformations $\mathcal{M}$ if $\forall x \in S$ and $\forall M \in \mathcal{M}$ the individual $M(x) \in S$ as well. We write $\mathcal{S}_\mathcal{M}$ to denote the collection of all subsets that are invariant under the family of mutation transformations $\mathcal{M}$.*

Families of mutation-invariant subsets enjoy the same properties as these of recombination-invariant subsets as described in proposition IV.1.[6] Once again, we design the families of mutation transformations based on preferable family of mutation-invariant subsets. Just as the case with recombination, a vast number of designs are possible, yet, for illustrative purposes, we select one of the simplest in the current article. We let the indexing family of our hybrid or mixed strategy EA $J = \{1, 2, \ldots M\}$ and for each $j \in J$ we select a family of mutation transformations $\mathcal{M}_j$ the collection of mutation-invariant subsets of which is, just as in case of recombination, $\mathcal{S}_j = \{S_q \mid q \geq j\} \cup \{\emptyset, \Omega\}$ with $S_q = \bigcup_{l \in L} f_l^{-1}([q, M])$. There is a further requirement on the families of mutation transformations though:

Condition 4. We require that the family of mutation transformations $\mathcal{M}_q$ for $q = M$ (i.e. at the highest common auxiliary fitness level) possesses the following property: whenever l and $k \leq L$, $\forall x \in f_l^{-1}(\{M\})\ \exists y \in f_k^{-1}(\{M\})$ and a sequence of mutation transformations $T_1, T_2, \ldots, T_i \in \mathcal{M}_l$ such that $y = T_i \circ T_{i-1} \circ \ldots \circ T_1(x)$. We further require that $\exists$ polynomials n^γ and n^λ such that $\forall l$ and $k \leq L$ and $\forall x \in f_l^{-1}(\{M\})\ \exists y \in f_k^{-1}(\{M\})$ such that the probability that y is encountered after $O(n^\gamma)$ applications of the mutation transformations from the families $\mathcal{M}_l$ is at least $\Omega(n^{-\lambda})$ regardless of the population in which the individual x appears and the iteration time at which the population arises.

The following simple lemma hints at the motivation for condition 4 in our design.

Lemma IV.2. *Consider any Markov chain on the state space $S_M = \bigcup_{l \in L} f_l^{-1}(\{M\})$ with the transition matrix $\{p_{x \to y}\}_{x,\, y \in S_M}$ defined as*

$$p_{x \to y} = \mu_x \{F \mid F \in \mathcal{M}_M \text{ and } F(x) = y\}$$

where $\{\mu_x\}_{x \in S_M}$ is the collection of probability measures on $\mathcal{M}_M$ satisfying condition 4 above. For an $x \in S_M$ let T_x denote the random waiting time to encounter a "satisfactory" solution with respect to the objective fitness function f_0 for the first time. Then $\forall x \in S_M\ E(T_x) \leq O\left(n^{\gamma+\lambda}\right)$.

Proof: According to condition 3, $\exists l_0 \in L$ such that any $y \in f_{l_0}^{-1}(\{M\})$ is a satisfactory "satisfactory" solution with respect to the objective fitness function f_0 so that for any given $x \in S_M$ the random variable T_x is bounded above by the

---

[4]This assumption does not reduce the generality since there are finitely many auxiliary fitness functions and one can always "shift all of them up" by an additive positive constant.

[5]The assumption of having a common range can be alleviated at the cost of technical complications that divert attention away from the mainstream idea of the current article.

[6]In fact, this applies to arbitrary families of m-ary transformations on Ω and their corresponding families of invariant subsets.

random time $T_{x \to l_0}$ of encountering an individual $y \in f_{l_0}^{-1}(\{M\})$ for the first time. Comparing the random time $T_{x \to l_0}$ with a geometric random variable T with a unit step size $O(n^{\gamma})$ and success probability $\Omega(n^{-\lambda})$, thanks to condition 4 we deduce that $E(T_{x \to l_0}) \leq n^{\gamma} E(T) = O(n^{\gamma}) \cdot \frac{1}{\Omega(n^{-\lambda})} = O\left(n^{\gamma + \lambda}\right)$ as claimed. ∎

The type of hybrid EA's design suggested in the current article is largely based on the notion of an individual's maximal auxiliary fitness level introduced below:

Definition IV.4. *For an individual $x \in \Omega$ the* maximal auxiliary fitness level *of x is $auxFit(x) = \max_{l \in L} f_l(x)$.*

In other words, the maximal auxiliary fitness level of an individual x is the largest auxiliary fitness level q of x so that x "fits" the schema S_q that is invariant under the family of recombination transformations $\mathcal{F}_j$ for $j \geq q$. A rather simple complexity analysis presented in the current paper relies on applying the simplified version of schema theory from section III to the special schemata introduced in the following definition:

Definition IV.5. *Consider a hybrid or mixed strategy EA fitting the framework of the current article that uses populations of size $m \geq 2$, and a population $\vec{x} = \{x_1, x_2, \ldots x_m\}$. Let $AuxMax(\vec{x}) = \max\{auxFit(x_i) \mid 1 \leq i \leq m\}$ denote the maximal auxiliary fitness level present in the population $\vec{x}$. We say that the schema $S_{AuxMax(\vec{x})}$ is the* leading current schema *while the schema $S_{AuxMax(\vec{x})+1}$ is the* leading future schema *of the population $\vec{x}$.*

One of the simplest (but not the only possible) designs of "efficient" hybrid or mixed strategy EAs is to concentrate the probability distributions $Pr^{\text{RecFamily}}_{(\vec{x}, t), (x_i, x_j)}$ on the set indexing the families of recombination transformations (recall that there are as many of these as there are auxiliary fitness levels) on the indices $q \geq auxFit(x_i)$ whenever $auxFit(x_i) = AuxMax(\vec{x})$. Likewise, the probability of selecting the families of mutation transformations, $Pr^{\text{MutFamily}}_{(\vec{x}, t), x_i}$, is also concentrated on the indices $q \geq auxFit(x_i)$ for the individuals x_i fitting the leading current schema. For $q \in \{1, 2, \ldots M\}$ let

$$P^{Rec}_{Imp}(q) = \min\{Pr^{\text{RecFamily}}_{(\vec{x}, t), (x_i, x_j)}(w > q) \mid auxFit(x_i) = q, \, x_j \in \Omega\}$$

and

$$P^{Mut}_{Imp}(q) = \min\{Pr^{\text{MutFamily}}_{(\vec{x}, t), x_i}(w \geq q) \mid auxFit(x_i) = q\}. \tag{11}$$

denote the minimal probabilities of improving the auxiliary fitness level q after applying recombination and mutation transformations respectively. We assume that all the probabilities of types $P^{Rec}_{Imp}(q)$ and $P^{Mut}_{Imp}(q)$ are positive.

In the current paper we do not assume much about selection apart from preserving the highest auxiliary as well as the highest objective fitness levels. Formally this can be defined as follows.

Definition IV.6. *We say that a selection transformation $Sel : (\Omega^m)^2 \to \Omega^m$ is hybrid-elitist if $\forall \vec{x}$ and $\vec{y} \in \Omega^m$*

$$AuxMax(Sel(\vec{x}, \vec{y})) = \max\{AuxMax(\vec{x}), \, AuxMax(\vec{y})\}$$

and

$$\max\{f_0(Sel(\vec{x}, \vec{y})) \mid 1 \leq i \leq m\} = \max\{\max\{f_0(x_i) \mid 1 \leq i \leq m\}, \, \max\{f_0(y_i) \mid 1 \leq i \leq m\}\}$$

Applying corollary III.8 with S_0 being the leading current schema and S being the leading future schema as in definition IV.5 and observing that $N(S_{AuxMax(\vec{x})}, \vec{x}^{\text{rec}}) \geq 1$ by definition IV.5, we immediately deduce the following fact.

Lemma IV.3. *Suppose we are given a hybrid or mixed strategy EA with constant population size m that exploits hybrid-elitist selection. Then*

$$Pr(N(S_{AuxMax(\vec{x})+1}, \vec{z}) \geq 1) \geq \left(1 - \left(1 - \frac{1}{m}\right)^m\right) \times P^{Rec}_{Imp}(AuxMax(\vec{x})) \cdot P^{Mut}_{Imp}(AuxMax(\vec{x}) + 1)$$

$$> (1 - \exp(-1)) \cdot P^{Rec}_{Imp}(AuxMax(\vec{x})) \cdot P^{Mut}_{Imp}(AuxMax(\vec{x}) + 1)$$

where, as usual, $\vec{z}$ denotes the population obtained from the population $\vec{x}$ upon completion of a single recombination $\longrightarrow$ mutation $\longrightarrow$ selection cycle.

V. DRIFT ANALYSIS AND SIMPLE RUNTIME COMPLEXITY BOUNDS

Drift analysis methodology invented in [16] has been introduced into the evolutionary computation theory for estimating the expected run-time complexity in [17], quickly gained popularity and has been modified and enhanced in a number of ways. In the current article we will use the additive variable drift analysis version established in [18]. For the sake of completeness, the necessary definition and the relevant lemma are stated below.

Definition V.1. *Let $(\mathcal{X}, \{p_{x \to y}\}_{x, y \in \mathcal{X}})$ denote a Markov chain with finite state space $\mathcal{X}$ and transition probabilities $p_{x \to y}$ for x and $y \in \mathcal{X}$. Let $A \subseteq \mathcal{X}$. A distance function D on $\mathcal{X}$ with respect to A is any function $D : \mathcal{X} \to [0, \infty)$ with the property*

that $D(x) = 0$ if and only if $x \in A$. Let $\{X_t\}_{t=0}^{\infty}$ denote the stochastic process associated with the Markov chain $\mathcal{X}$. We are interested in the following waiting time random variable:

$$T(\mathbf{x} \,|\, X_0 = \xi_0) = \min\{t \,|\, X_t(\mathbf{x}) \in A\}$$

under the assumption that $X_0(\mathbf{x}) = \xi_0$ with probability 1 (i.e. the chain starts at a specified $\xi_0 \in \mathcal{X}$).

A simple complexity bound appearing in the current paper is based on the following additive variable drift lemma from [18].

Lemma V.1. *Suppose we are given a Markov chain*
$(\mathcal{X}, \{p_{x \to y}\}_{x,y \in \mathcal{X}})$, *a subset* $A \subseteq \mathcal{X}$ *and a distance function* $D : \mathcal{X} \to [0, \infty)$ *as described in definition V.1. Suppose also that for every integer* $k \in \mathbb{N} \cup \{0\}$ $\exists$ *a constant* $l_k \in (0, \infty)$ *such that* $\forall\, x \in A^c$ *with* $\lceil D(x) \rceil \geq k$ *(here* A^c *denotes the complement of A in $\mathcal{X}$) we have* $D(x) - \sum_{y \in \mathcal{X}} p_{x \to y} D(y) \geq l_k$. *Then*

$$E(T(\mathbf{x} \,|\, X_0 = \xi_0)) \leq \sum_{k=1}^{\lceil D(\xi_0) \rceil} \frac{1}{l_k}.$$

Given a hybrid or mixed strategy EA with the highest auxiliary fitness level M satisfying conditions 1, 2 and 3 that exploits hybrid-elitist selection, we apply lemma V.1 to the Markov chain $\mathcal{X}$ of all populations of size m, Ω^m with the probability $p_{\vec{x} \to \vec{z}}$ being the probability that the population $\vec{z}$ is obtained from the population $\vec{x}$ upon completion of a recombination $\longrightarrow$ mutation $\longrightarrow$ selection cycle. The set

$$A = \{\vec{x} \,|\, \vec{x} \in \Omega^m \text{ and } AuxMax(\vec{x}) = M\}$$

is the set of all populations containing an individual of the highest auxiliary fitness level and the distance function $D : \mathcal{X} \to \{0, 1, \ldots, M\}$ defined as $D(\vec{x}) = M - AuxMax(\vec{x})$. According to lemma IV.3, whenever $AuxMax(\vec{x}) = q$

$$Pr(D(\vec{x}) - D(\vec{z}) \geq 1 \,|\, p_{\vec{x} \to \vec{z}} \neq 0) \geq \left(1 - \left(1 - \frac{1}{m}\right)^m\right) \cdot P_{Imp}^{Rec}(q) \cdot P_{Imp}^{Mut}(q + 1)$$
$$> (1 - \exp(-1)) P_{Imp}^{Rec}(q) \cdot P_{Imp}^{Mut}(q + 1)$$

Furthermore, thanks to the assumption that our EA exploits a hybrid-elitist selection, whenever $p_{\vec{x} \to \vec{z}} \neq 0$ $D(\vec{x}) - D(\vec{z}) \geq 0$. Now, letting

$$l_k(m) = \left(1 - \left(1 - \frac{1}{m}\right)^m\right) \cdot P_{Imp}^{Rec}(k) \cdot P_{Imp}^{Mut}(k + 1) \tag{12}$$

and

$$l_k = (1 - \exp(-1)) \cdot P_{Imp}^{Rec}(k) \cdot P_{Imp}^{Mut}(k + 1) \tag{13}$$

we deduce that the worst case expected runtime complexity upper bound to reach the highest auxiliary fitness level for a hybrid or mixed strategy EA that exploits hybrid-elitist selection (and hence must have a population size $m \geq 2$) is $\sum_{k=0}^{M-1} \frac{1}{l_k(m)} \leq \sum_{k=0}^{M-1} \frac{1}{l_k}$. If we assume, in addition, that our EA satisfies condition 2 in section IV then $M = O(n^\tau)$. Furthermore, conditions 1, 3 and 4 allow us to apply lemma IV.2 and to deduce the following expected runtime result.

Theorem V.2. *Suppose a given hybrid or mixed strategy EA that employs hybrid-elitist selection satisfies conditions* 1, 2, 3 *and* 4 *in section IV. Then the worst-case expected runtime for the EA to reach a satisfactory solution is bounded above by*

$$\sum_{k=0}^{M-1} \frac{1}{l_k(m)} + n^{\gamma+\lambda} \leq \sum_{k=0}^{M-1} \frac{1}{l_k} + n^{\gamma+\lambda}.$$

In the next section we illustrate an application of theorem V.2 and, more importantly, the methodology developed in sections III and IV as well as in the current section with a single machine scheduling problem.

VI. AN EXAMPLE APPLICATION: SINGLE MACHINE SCHEDULING PROBLEM

One of many classical NP-hard combinatorial optimization problems is the single machine scheduling problem (see chapter on scheduling problems by Leslie Hall in [5]). The instance of the problem of size n consists of a sequence of ordered triples $\{j_i\}_{i=1}^n$ where $j_i = (r_i, p_i, q_i)$ with r_i, p_i and $q_i \in [0, \infty)$ standing for "release time", "processing time" and "delivery time" of the job j_i respectively. Each of the jobs has to be processed on a single machine, call it M, without interruption and it starts getting delivered immediately after being processed. There is no restriction on the total number of jobs being delivered simultaneously, yet a job can start getting processed no earlier than its release time r_i and only when the machine is available (i.e. not processing another job). The jobs can be processed in any order and the objective is to minimize the "maximal lateness" of the schedule: i.e. the time instant when the last job has been just delivered. Thus, the search space $\Omega = \{\pi \,|\, \pi : I_n \to$

I_n is a permutation} where $I_n = \{1, 2, \ldots, n\}$ so that every $\pi \in \Omega$ determines the schedule $(j_{\pi(1)}, j_{\pi(2)}, \ldots, j_{\pi(n)})$. Let $s_{\pi(i)}$ denote the time when the job $j_{\pi(i)}$ starts processing. Then the *maximal lateness* of the schedule π is $J_\pi = \max_{1 \leq i \leq n}\{s_{\pi(i)} + p_{\pi(i)} + q_{\pi(i)}\}$. and the objective is to find a permutation $\sigma \in \Omega$ such that $J_\sigma = J^* = \min\{J_\pi \mid \pi \in \Omega\}$. Let now $P = \sum_{i=1}^{n} p_i$ and let $\epsilon > 0$ be given. Let $\delta = \epsilon P$ and let $B_\epsilon = \{i \mid p_i \geq \delta\}$. Then $|B_\epsilon| \leq \frac{P}{\epsilon P} = \frac{1}{\epsilon}$. Let $\Phi = \{\phi : B_\epsilon \to I_n \mid \phi \text{ is one-to-one}\}$ denote the set of *repositioning maps* and notice that the total number of such repositioning maps is bounded above as $|\Phi| = n^{|B_\epsilon|} \leq n^{\frac{1}{\epsilon}}$ (this verifies condition 1 in the upcoming design). The following notion is crucial in determining the auxiliary fitness functions (see [4] for a significantly more detailed exposition).

Definition VI.1. *We say that a schedule $\pi \in \Omega$ is (k, ϵ, ϕ)-Jackson if $\forall h \leq k$ if $h = \phi(i)$ for some $i \in I_n$ then $\pi(i) = h = \phi(i)$ or else let $a(h-1)$ denote the time when the job $\pi(h-1)$ has just finished being processed and consider the set of all the jobs $A_h = \{u \mid u \notin B_\epsilon \text{ and } r_u \geq a(h-1)\}$. It is then required that $q_{\pi(h)} = \max\{q_u \mid u \in A_h\}$.*

Plainly speaking, the idea behind definition VI.1 is the stepwise implementation of the "partial Jackson rule": whenever the machine is available, as long as a long-processing job (with $p_i \geq \delta$) is not one booked to be scheduled at time h, then schedule a job with "short" processing time that has the longest delivery time next. Our set of auxiliary fitness functions is indexed by the family of repositioning maps Φ and $\forall \phi \in \Phi$ and $\pi \in \Omega$, $f_\phi(\pi) = \max\{k \mid k \in I_n \text{ and } \pi \text{ is } (k, \epsilon, \phi) - \text{Jackson}\}$. The following clever result that is implicitly established in [5] and enforces condition 3 to hold in our design appears below:

Theorem VI.1. $\exists$ *a repositioning map $\phi \in \Phi$ such that $f_\phi(\pi) = n \implies f_0(\pi) = J_\pi \leq J^*(1 + \epsilon)$.*

Since there are totally $n = n^1$ auxiliary fitness levels, condition 2 is fulfilled automatically. The only remaining part is to design the families of recombination and mutation transformations preserving the "auxiliary cross-fitness level schemata" and the highest auxiliary fitness level mutation transformations according to the recipe in section IV. This can be done in a vast number of ways (see [12] for a detailed analysis of the relationship between families of recombination-invariant subsets and the families of recombination transformations fixing them). Here is one possibility. For an auxiliary fitness level i we define the family of hybrid recombination transformations $\mathcal{F}_i = \{F_\zeta^i \mid \zeta \in \mathcal{S}_{I_{n-i}}\}$ where I_{n-i} denotes the indexing set $\{1, 2, \ldots, n-i\}$ while $\mathcal{S}_{I_{n-i}}$ denotes the group of all permutations on the set $\mathcal{S}_{I_{n-i}}$ as follows. Select a permutation ζ on I_{n-i}. Given a pair of permutations $(\pi, \sigma) \in \Omega^2$ with $\pi = (\pi(1), \pi(2), \ldots, \pi(n))$ and $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$, let $F_\zeta^i(\pi, \sigma) = \eta = (\eta(1), \eta(2), \ldots, \eta(n))$ with $\eta(l) = \pi(l)$ whenever $1 \leq l \leq i$. Now extract a subsequence of jobs in σ that do not appear among the first i jobs in π and notice that there must be exactly $n - i$ such jobs. The ordering of these jobs in σ can be represented by a permutation $\omega \in \mathcal{S}_{I_{n-i}}$ and the composition $\zeta \circ \omega$ produces another ordering of these remaining jobs. We schedule them all right after the job $\eta(i) = \pi(i)$ in the schedule η in the ordering $\zeta \circ \omega$. Observe that if $auxFit(\pi) = i$ and a transformation $F_\zeta^i \in \mathcal{F}_i$ is selected uniformly at random, then $\forall \sigma \in \Omega$ the probability that $auxFit(F_\zeta^i(\pi, \sigma)) > i$ is at least $\frac{1}{n-i}$ since at least one of the $n - i$ jobs that do not appear among the first i jobs in π, when scheduled after the job $j_{\pi(i)}$ must improve the auxiliary fitness level for at least one of the auxiliary fitness functions. The next step is to design mutation transformations and, once again, the number of ways to do so is countless. Here we present the following very simple design: for an auxiliary fitness level i, let $\mathcal{M}_i = \{M_{a,b}^i \mid a \text{ and } b \in I_n\}$ where $M_{a,b}^i(\pi) = \tilde{\pi}$ and $\tilde{\pi} = \pi$ unless $\pi(a)$ or $\pi(b) \in B_\epsilon$ in which case the positions of these jobs are swapped and then, if at least one of the jobs that has been swapped appears below the ith position, the partial Jackson rule with respect to the new positioning $\tilde{\phi}$ of the jobs in B_ϵ is applied starting with the lowest index of one of the repositioned job up to the ith fitness level of the auxiliary fitness function $f_{\tilde{\phi}}$ thereby obtaining a new schedule $\tilde{\pi}$. It follows then that $f_{\tilde{\phi}}(\pi) \geq i = auxFit(\pi)$. We equip each family of mutation transformations $\mathcal{M}_i$ with the uniform probability distribution. To apply theorem V.2, all that remains now is to check condition 4 in section IV. Here we use the classical fact "about card shuffling via random transpositions", the simplest analysis of which is presented as an elegant illustration of the Markov chain coupling methodology in Chapter 4-3, section 1.7 of [19], it easily follows that if we are given a schedule π with $auxFit(\pi) = n$ (the highest auxiliary fitness level) and another schedule σ with $auxFit(\sigma) = n$, after $O(n^2)$ time steps, the probability that the schedule σ has been encountered after repeated application of the mutation transformations from the family $\mathcal{M}_n$ is at least $\Omega(n^{\frac{1}{\epsilon}})$, thereby establishing the desired condition 4. We are now in a position to apply theorem V.2 to deduce that the expected runtime until encountering a population containing a schedule π with $f_0(\pi) = J_\pi \leq J^*(1 + \epsilon)$ is no bigger than $\sum_{i=0}^{n-1} \frac{1}{n-i} + O(n^{2+\frac{1}{\epsilon}}) = \Theta(\ln(n)) + O(n^{2+\frac{1}{\epsilon}}) = O(n^{2+\frac{1}{\epsilon}})$.

VII. Conclusion

While classical schema theory has been widely criticized in the setting of traditional EAs: see, for instance, section 3.2 of [20], it's quite remarkable to observe that in case of mixed strategy and hybrid EAs it can be used for intelligent design guidance as well as to understand the success behind this novel kind of EAs. The current paper presents only preliminary and highly simplified analysis that may be altered and improved in a number of ways. For instance, the generalized schema theorem III.6 motivates runtime analysis based on the ideas in [4] in place of drift analysis methodology to design and analyze hybrid and/or mixed strategy EAs where the runtime to encounter a satisfactory solution is polynomial with overwhelmingly high probability. This work is postponed for the future research. Nonetheless, the authors believe that the core ideas of designing the collections of generalized schemata (see definition IV.1) based on the auxiliary fitness levels in a similar manner to the

way it's been done in section IV and then designing the families of recombination and mutation transformations based on the corresponding families of generalized schemata, opens the door to understanding and designing efficient hybrid and mixed strategy EAs.

REFERENCES

[1] P. Galinier and J. Hao. Hybrid evolutionary algorithms for graph coloring. *Journal of Combinatorial Optimization*, 3(4):379–397, 1999.

[2] P. Preux and E. Talbi. Towards hybrid evolutionary algorithms. *International transactions in operational research*, 6(6):557–570, 1999.

[3] H. Dong, J. He, J. Huang, and W. Hou. Evolutionary programming using a mixed mutation strategy. *Information Sciences*, 177(1):312–327, 2007.

[4] B. Mitavskiy and J. He. A polynomial time approximation scheme for a single machine scheduling problem using a hybrid evolutionary algorithm. In *World Congress on Computational Intelligence (WCCI-2012)*, pages 156–175. Springer, lecture Notes in Computer Science 3469, 2012.

[5] D. Hochbaum. *Approximation algorithms for NP-hard problems*. PWS Publishing, Reading, Massachusetts, 1997.

[6] A. Auger and B. Doerr. *Theory of Randomized Search Heuristics*. Series on Theoretical Computer Science,Reading, Massachusetts, 2011.

[7] J. Antonisse. A new interpretation of schema notation that overturns the binary encoding constraint. In *Procedings of the Third International Conference on Genetic Algorithms*, pages 84–89, 1989.

[8] R. Poli and B. Langdon. Schema theory for genetic programming with one-point crossover and point mutation. *Evolutionary Computation*, 6(3):231–252, 1998.

[9] N. Radcliffe. The algebra of genetic algorithms. *Annals of mathematics and artificial intelligence*, 10(3):339–384, 1992.

[10] B. Stadler and P. Stadler. Generalized topological spaces in evolutionary theory and combinatorial chemistry. *Journal of Chemical Information and Computer Sciences*, 42(2):577–585, 2002.

[11] P. Stadler and G. Wagner. Algebraic theory of recombination spaces. *Evolutionary Computation*, 5(3):241–275, 1997.

[12] B. Mitavskiy. Crossover invariant subsets of the search space for evolutionary algorithms. *Evolutionary Computation*, 12(1):19–46, 2004.

[13] B. Mitavskiy. Comparing evolutionary computation techniques via their representation. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2003)*, pages 156–175. Springer, lecture Notes in Computer Science 3469, 2003.

[14] B. Mitavskiy. A category theoretic method for comparing evolutionary computation techniques via their representation. In *Proceedings of the 15th European Summer School in Logic Language and Information (ESSLLI-2003)*, pages 156–175. Springer, lecture Notes in Computer Science 3469, 2003.

[15] M. Barr and C. Wells. *Category Theory for Computing Science*. Prentice Hall, Reading, Massachusetts, 1998.

[16] B. Hajek. Hitting time and occupation time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(1):502–525, 1982.

[17] J. He and X. Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127(1):57–85, 2001.

[18] B. Mitavskiy, J. Rowe, and C. Cannings. Theoretical analysis of local search strategies to optimize network communication subject to preserving the total number of links. *International Journal of Intelligent Computing and Cybernetics*, 2(2):243–284, 2009.

[19] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs. unpublished monograph: http://www.stat.berkeley.edu/∼aldous/RWG/book.html, 2001.

[20] T. Jansen. *Analyzing Evolutionary Algorithms*. Springer, Reading, Massachusetts, 2013.