On the Analysis of the Simple Genetic Algorithm

Pietro S. Oliveto*
School of Computer Science
University of Birmingham
United Kingdom
p.s.oliveto@cs.bham.ac.uk

Carsten Witt†
DTU Informatics
Technical University of Denmark
Denmark
cawi@imm.dtu.dk

ABSTRACT

For many years it has been a challenge to analyze the time complexity of Genetic Algorithms (GAs) using stochastic selection together with crossover and mutation. This paper presents a rigorous runtime analysis of the well-known Simple Genetic Algorithm (SGA) for OneMax. It is proved that the SGA has exponential runtime with overwhelming probability for population sizes up to $\mu \leq n^{1/8-\varepsilon}$ for some arbitrarily small constant ε and problem size n. To the best of our knowledge, this is the first time non-trivial lower bounds are obtained on the runtime of a standard crossover-based GA for a standard benchmark function. The presented techniques might serve as a first basis towards systematic runtime analyses of GAs.

Categories and Subject Descriptors

F.2.0 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity

General Terms

Algorithms, Performance, Theory

Keywords

Simple Genetic Algorithm, Crossover, Runtime Analysis

1. INTRODUCTION

In recent years significant progress has been made in the runtime analysis of Evolutionary Algorithms (EAs) [1]. On one hand, nowadays it is possible to analyse the performance of simple EAs on well known combinatorial optimization problems [15]. On the other hand new techniques have enabled the analysis of more realistic EAs using populations and stochastic selection mechansisms. The introduction of the *simplified negative-drift theorem* [17] has allowed the analysis of population-based EAs using fitness proportional selection while the *negative-drift in populations* theorem [9] combined with the *fitness levels for non-elitist populations* technique [10], permit the analysis of population based EAs with several stochastic selection mechanisms such as *comma*, *tournament* or *linear ranking* selection. However, the above mentioned techniques cannot be directly applied to algorithms using crossover operators which are at the heart of Genetic Algorithms (GAs). In fact, it has been a challenge for many years to analyse the runtime of standard GAs using stochastic selection together with mutation and crossover. Many examples using toy problems [7, 20, 8] and classical optimisation problems [16, 2, 11] proving that crossover is useful are available. Nevertheless, the algorithms considered in the quoted literature rely heavily on elitism and partly on diversity mechanisms. Moreover, mostly upper bounds on the running time of crossover-based algorithms are obtained. In particular, no lower bounds on the standard benchmark function OneMax were available so far.

In this paper we present a first step towards a systematic analysis of GAs by performing a runtime analysis of the *Simple Genetic Algorithm* (**SGA**) on OneMax. Although the **SGA** is the most elementary example of GA, it has often been considered in literature. For instance Goldberg uses the algorithm in several chapters of his book [4], Vose has used it for building his infinite population model [19] and Rudolph used it for his Markov chain convergence analysis [18]. The results that will be presented herein are a continuation of previous work. Happ et al. performed the first runtime analysis of fitness proportional selection (f.p.s.) by considering only one individual and bitwise mutation [5]. This work was extended by Neumann et al. to consider arbitrary population sizes again on a mutation-only EA [12]. In particular, it was proved that the runtime of an EA using f.p.s. and bitwise mutation for OneMax is exponential with overwhelming probability (w. o. p.) whatever the polynomial population size. Also if the population is not too large (i. e., logarithmic in the problem size), then the algorithm cannot optimize any function with unique optimum in polynomial time w. o. p. The main result in this paper is the rigorous proof that even by adding crossover the obtained **SGA** cannot optimize OneMax.

The well-known **SGA** is displayed in Figure 1. For analyses convenience we will also consider simplified versions of the **SGA** along the way, by either removing mutation or even by changing the selection operator. If no mutation is used and the rest of the **SGA** is kept the same we will

*Supported by EPSRC grant EP/H028900/1.

†Supported by DFG grant WI 3552/1.

Algorithm (SGA) [4]

1. Create a parent population P consisting of μ randomly chosen individuals;
2. $C := \emptyset$.
3. While $|C| < \mu$ do
   - **Fitness proportional selection:** Select two individuals x' and x'' from P according to fitness-proportional selection without replacement.
   - **Uniform crossover:** Create an offspring x by setting each bit $x_i = x'_i$ with probability 1/2 and $x_i = x''_i$ otherwise, for $1 \le i \le n$.
   - **Standard Bit Mutation:** Flip each bit x_i of x with probability $1/n$, for $1 \le i \le n$.
   - $C := C \cup \{x\}$.
4. Set $P := C$ and go to 2.

Figure 1: The Simple GA

call the algorithm **fitness-Proportional Recombinative SGA** and refer to it as **PR-SGA**. If no mutation is used and uniform selection is used instead of a proportional one, then we will call the algorithm **Uniform Recombinative SGA** and refer to it as **UR-SGA**.

The whole analysis depends on the fact that we bound three quantities: The first two are h (the best OneMax-value of a population), and ℓ (the worst OneMax-value). Obviously, if $h < n$ then the optimum has not been found. The third quantity, denoted by s, is defined as the number of bit positions that are not *converged*, which means that both bit values are taken by individuals in the population. The number of non-converged positions is a simple measure of diversity; for simplicity s is also called *the diversity* hereinafter. We study this measure since crossover does not have any effect on bit positions that are converged. Good bounds on the diversity will be obtained under the condition that the selection operator chooses individuals uniformly or almost uniformly.

Fitness-proportional selection is close to uniform if h/ℓ is close to 1. We will formalize this by limiting the so-called *bandwidth* of the population, where we distinguish between the *additive bandwidth* $h - \ell$ and the *multiplicative bandwidth* h/ℓ. One simple, but absolutely crucial observation is that $h - \ell \le s$, i. e., the additive bandwidth cannot be larger than the number of non-converged positions. From a bound on the diversity, hence a bound on the additive bandwidth, we can also easily get a bound on the multiplicative bandwidth, which in this way is proved to stay close to 1. The idea of bounding the bandwidth is not new. It has been previously used to derive runtime results for elitist EAs [6, 21] and stochastic selection EAs without crossover [12].

Our main result is that the **SGA** for OneMax has exponential optimization time w. o. p. if $\mu \le n^{1/8-\varepsilon}$ for some arbitrarily small $\varepsilon > 0$. The high-level proof strategy is as follows: We first show that additive and also multiplicative bandwidths are bounded in a certain way for a certain number of generations after initialization w. o. p. This will allow us to obtain non-trivial bounds on the diversity, which in turn limit the effect of crossover. Then (assuming a bound on the diversity) a drift analysis w. r. t. the best OneMax-value h will be performed, which shows on the one hand that the optimum is not reached in exponential time w. o. p.; similarly ℓ will be bounded from below. But then the limit on the additive and multiplicative bandwidth and hence the bound on the diversity will persist for an exponential number of generations w. o. p., which closes the proof.

The rest of the paper is structured as follows. In Section 2 we prove that the genetic drift leads to very small diversity, i. e., many converged bit positions. Along the way we prove that also the **UR-SGA** and the **PR-SGA** are inefficient on OneMax. In Section 3.2 we present the drift analysis with respect to the best fitness value, which leads to the main result of the paper. Due to space limitations, several proofs had to be omitted from this conference paper.

2. ANALYSIS OF DIVERSITY

The analysis of the **UR-SGA** reveals the main proof ideas for controlling the diversity of the population, i. e., the number of bit positions that are not converged. After, the effects of proportional selection and mutation will be discussed.

2.1 Uniform selection

Let X be a random variable representing the number of individuals in the population P of size μ having a one-bit in an arbitrary (but fixed) position i, and X' the value of the random variable after one generation of the **UR-SGA**. In the following lemma we show that the conditional random variable $Z_k := (X' \mid X = k)$ is binomially distributed, with parameters $B(\mu, p = k/\mu)$.

Lemma 1. $Z_k := (X' \mid X = k) \sim B(\mu, p = k/\mu)$.

Proof. The current population has k individuals with a one-bit at position i, and $\mu - k$ individuals with a zero-bit in that position. We consider an individual x that is created through one crossover step and show that the probability it has a one-bit at position i is k/μ. Then the lemma will follow because one generation of the **UR-SGA** creates a new population by performing μ independent crossover steps each with probability k/μ.

With probability $\frac{k}{\mu} \cdot \frac{k-1}{\mu-1}$ two individuals with a one-bit in position i are selected for crossover. Then x will have a one-bit in position i after crossover with probability 1. On the other hand, with probabilities $\frac{k}{\mu} \cdot \frac{\mu-k}{\mu-1}$ and $\frac{\mu-k}{\mu} \cdot \frac{k}{\mu-1}$ one of the parents will have a one-bit in position i while the other will have a zero-bit. In these cases x will obtain a one-bit after the crossover step with probability 1/2. Overall the total probability is:

$$p = 2\frac{1}{2}\frac{k(\mu-k)}{\mu(\mu-1)} + \frac{k(k-1)}{\mu(\mu-1)} = \frac{k(\mu-k)+k(k-1)}{\mu(\mu-1)} = \frac{k}{\mu}. \quad \square$$

Hence, $E(Z_k) = k$ and $\mathrm{Var}(Z_k) = \mu \cdot p(1-p) = k(1-k/\mu)$. From Lemma 1 it follows that the Markov process $\{X_t\}_{t\ge 0}$ described by the X-values over time behaves like a martingale, i. e., in expectation the number of individuals in the population with a one-bit in an arbitrary position remains the same from one generation of the **UR-SGA** to another. This implies that there is no drift (i. e., $E(X' - X | X = k) = 0$). Nevertheless, random fluctuations (measured by the variance of the random variable) will drive the process to one of its absorbing states 0 and μ in expected polynomial time. In order to capture the movement of the process through a drift analysis, we map X_t with the potential function $Y_t := (X_t - \frac{\mu}{2})^2$ (a well-known approach, cf. [13, Lemma 7]). In the following lemma we show that the drift of Y_t equals the variance of X_t.

Lemma 2. *Let $\{X_t\}_{t\geq 0}$, be a Markov process that is a martingale (i. e., $E(X_t \mid X_{t-1}) = X_{t-1}$) on state space $\{0, 1, \dots, \mu\}$. Then $Y_t := \left(X_t - \frac{\mu}{2}\right)^2$ is a Markov process with drift $E(Y_t - Y_{t-1} \mid X_{t-1} = k) = \mathrm{Var}(X_t \mid X_{t-1} = k)$.*

Proof. We evaluate the drift:

$$\begin{aligned}
&E(Y_t - Y_{t-1} \mid X_{t-1} = k) = \\
&= E\left(\left(X_t - \frac{\mu}{2}\right)^2 - \left(X_{t-1} - \frac{\mu}{2}\right)^2 \mid X_{t-1} = k\right)\\
&= E\left(\left(X_t - \frac{\mu}{2}\right)^2 - \left(k - \frac{\mu}{2}\right)^2 \mid X_{t-1} = k\right)\\
&= E\left(X_t^2 - \mu \cdot X_t + \frac{\mu^2}{4} - k^2 + \mu \cdot k - \frac{\mu^2}{4} \mid X_{t-1} = k\right)\\
&= E(X_t^2 \mid X_{t-1} = k) - \mu \cdot E(X_t \mid X_{t-1} = k) - k^2 + \mu \cdot k\\
&= E(X_t^2 \mid X_{t-1} = k) - \mu \cdot k - k^2 + \mu \cdot k\\
&= E(X_t^2 \mid X_{t-1} = k) - (E(X_t \mid X_{t-1} = k))^2\\
&= \mathrm{Var}(X_t \mid X_{t-1} = k). \quad \square
\end{aligned}$$

A straightforward drift analysis using Lemma 2 bounds the first hitting time for either 0 or μ from above by $\frac{\mu^2}{4\delta}$ where $\delta = \min_k \mathrm{Var}(X_t \mid X_{t-1} = k)$. This implies that by considering that the variance is minimal for $k \in \{1, \mu - 1\}$ (hence, $\mathrm{Var}(X_t \mid X_{t-1} = k) \geq 1 - 1/\mu$) we get an expected time (number of generations) $\frac{\mu^2}{4(1-1/\mu)} = O(\mu^2)$ to obtain a population in the **UR-SGA** where an arbitrary bit position i has converged, i. e., either all the individuals have a one-bit or they all have a zero-bit at position i. In Lemma 4, we look more carefully into the drift process to derive a tighter bound on such first hitting time. The following lemma will be useful for the purpose.

Lemma 3. $\mathrm{Var}(X_t \mid X_{t-1}) \geq \frac{\mu}{4} - \frac{\sqrt{Y_{t-1}}}{2}$.

Proof. We consider the potential function Y_t mapping X_t and note that $Y_{t-1} := (X_{t-1} - \mu/2)^2$ implies $X_{t-1} = \mu/2 \pm \sqrt{Y_{t-1}}$. Since the two solutions are symmetric, we examine the one with the negative sign (i. e., $X_{t-1} = \mu/2 - \sqrt{Y_{t-1}} \leq \mu/2$). We get

$$\begin{aligned}
&\mathrm{Var}(X_t \mid X_{t-1}) = \mu \cdot \frac{X_{t-1}}{\mu} \cdot \left(1 - \frac{X_{t-1}}{\mu}\right)\\
&\geq X_{t-1} \cdot \left(1 - \frac{\mu/2}{\mu}\right) = \left(\frac{\mu}{2} - \sqrt{Y_{t-1}}\right) \cdot \frac{1}{2} = \frac{\mu}{4} - \frac{\sqrt{Y_{t-1}}}{2}.
\end{aligned}$$

$\square$

Lemma 4. *The expected number of generations for an arbitrary bit position i in the population of the* **UR-SGA** *to converge is $O(\mu \log \mu)$.*

Proof. We apply the variable drift theorem [3] w. r. t. the process $Z_t := \mu^2/4 - Y_t$ with the aim to estimate the expected first hitting time $E(T)$ for state 0. Using Lemma 3, the drift is bounded by

$$\begin{aligned}
&E(Z_t - Z_{t-1} \mid Z_{t-1}) = E(Y_t - Y_{t-1} \mid Y_{t-1})\\
&= \mathrm{Var}(X_t \mid X_{t-1}) \geq \frac{\mu}{4} - \frac{\sqrt{Y_{t-1}}}{2} = \frac{\mu}{4} - \frac{\sqrt{\mu^2/4 - Z_{t-1}}}{2},
\end{aligned}$$

which is monotone increasing in Z_{t-1}. Note that the smallest positive value for Z_{t-1} equals $\mu^2/4 - (1 - \mu/2)^2 = \mu - 1$. The variable drift theorem yields

$$E(T) \leq \frac{\mu - 1}{1/2} + \int_{\mu-1}^{\mu^2/4} \frac{1}{\mu/4 - \sqrt{\mu^2/4 - z}/2}\,\mathrm{d}z = O(\mu \log \mu),$$

where a closed formula for the integral was obtained from a computer algebra system. $\square$

Before we can prove the final result of this subsection, we need a bound on the maximum progress achieved by crossover in one generation. This is given in the following lemma (using a similar reasoning as in [8]), which also takes mutation into account for future needs. Throughout the paper, events are proved to occur w. o. p., which means probability $1 - 2^{-\Omega(n^\delta)}$ for some constant $\delta > 0$.

Lemma 5. *With probability at least $1 - e^{-\Omega(n^{2\varepsilon})}$ the maximum progress achieved by the crossover operator in one step is $n^{1/2+\varepsilon}$. With probability at least $1 - e^{-\Omega(n^{2\varepsilon})}$ the maximum progress achieved by a crossover and mutation step is $2n^{1/2+\varepsilon}$. The maximum progress per generation is bounded in the same way if $\mu = poly(n)$.*

The main idea behind the following theorem is that w. o. p. all the bits have converged before the optimum has been found.

Theorem 1. *Let $\mu \leq n^{1/2-\varepsilon}$. With overwhelming probability the* **UR-SGA** *requires infinite time to optimize* OneMax.

Proof. By Lemma 4 the expected number of generations for an arbitrary bit position to converge is at most $cn^{1/2-\varepsilon} \log n$ for some constant c. By applying a union bound and Markov's inequality in repeated phases, we get a probability of $2^{-\Omega(n^{\varepsilon/2})}$ that more than $cn^{1/2-\varepsilon/2} \log n$ generations are required to have *all* positions converged. Due to symmetry, the probability that all zero-bits are obtained in a position (rather than one-bits) is 1/2. This implies a probability that this does not happen for any of the n bits is 2^{-n}. Hence, w. o. p. all the individuals of the population will have a zero-bit in some position j after $cn^{1/2-\varepsilon/2} \log n$ generations. If this happens, the optimum will never be found because the crossover operator cannot generate a one-bit in position j.

We complete the proof by showing that at least $n^{1/2-\varepsilon/4}/3$ generations are required w. o. p. to find the optimum. By Lemma 5 the crossover operator gains at most $n^{1/2+\varepsilon/4}$ one-bits in each generation w. o. p. (re-choosing the ε in the lemma appropriately). Since by Chernoff bounds each individual has at least $n/3$ zero-bits w. o. p. after initialisation, at least $(n/3)/n^{1/2+\varepsilon/4} = (1/3)n^{1/2-\varepsilon/4}$ generations are required to visit the optimum. By summing up the failure probabilities the theorem follows. $\square$

2.2 Proportional Selection + Mutation

In this subsection we investigate how much the X_t defined w. r. t. the **SGA** differs from the binomial distribution of the **UR-SGA** due to proportional selection (i. e., next two lemmas) and its effects, combined with mutation, on the drift of Y_t (i. e., Lemmas 8 and 9). In particular, no significant difference can be observed if $\mu \leq n^{1/6-\varepsilon}$. We conclude the subsection by showing that after expected $O(\mu n^{\varepsilon/8})$ generations all the bit positions have converged at least once (i. e., Lemma 10) and use the result to prove infinite runtime of the **PR-SGA** w. o. p.

LEMMA 6. *Let h^t and ℓ^t be respectively the best and the worst* ONEMAX*-value in the population of the* **SGA** *at generation t. Let $\mu \leq n^{1/6-\varepsilon}$ with $\varepsilon > 0$ an arbitrarily small constant. Then $h/\ell \leq 1 + \frac{1}{17\mu^2}$ up to generation $t = n^{1/6}$ w. o. p.*

The bound on the multiplicative bandwidth h/ℓ derived above is crucial for the following analyses. Hereinafter, we call populations satisfying $h/\ell \leq 1 + 1/(17\mu^2)$ *compact*. For compact populations, fitness-proportional selection is so close to uniform that the analyses from the previous subsection basically carry over. The straightforward proofs have been omitted.

LEMMA 7. *For a compact population the probabilities of the* **SGA** *selecting x_h and x_ℓ as first parent or as second parent are bounded as follows:*

$$P(x'_h) \leq \frac{1}{\mu} + \frac{1}{17\mu^3} \qquad P(x''_h) \leq \frac{1}{\mu - 1} + \frac{1}{17\mu^2(\mu - 1)}$$

and

$$P(x'_\ell) \geq \frac{1}{\mu} - \frac{1}{17\mu^3} \qquad P(x''_l) \geq \frac{1}{\mu - 1} - \frac{1}{17\mu^2(\mu - 1)}.$$

LEMMA 8. *Let $\mu \leq n^{1/6-\varepsilon}$ and the population be compact. Then for the* **SGA** *it holds that,*

$$k - \frac{1}{7\mu} \leq E(X_t \mid X_{t-1} = k) \leq k + \frac{1}{7\mu}.$$

LEMMA 9 (LEMMA 2'). *Let $\{X_t\}_{t\geq 0}$, be a process defined on state space $\{0, 1, \ldots, \mu\}$ such that $k - 1/(7\mu) \leq E(X_t \mid X_{t-1} = k) \leq k + 1/(7\mu)$. Then $Y_t := \left(X_t - \frac{\mu}{2}\right)^2$ is a drift process with*

$$E(Y_t - Y_{t-1} \mid X_{t-1} = k) \geq \mathrm{Var}(X_t \mid X_{t-1} = k) - \frac{4}{9}.$$

One thing to note is that the process X_t induced by the **SGA** does not necessarily have absorbing states. More precisely, a converged bit position might be turned into non-converged as the result of a mutation.

LEMMA 10 (LEMMA 4'). *The expected number of generations for an arbitrary bit position i in the population of the* **SGA** *with $\mu \leq n^{1/6-\varepsilon}$ to converge is $O(\mu \log \mu)$. With probability at least $1 - 2^{-\Omega(n^{\varepsilon/8}/\log n)}$, after $O(\mu n^{\varepsilon/8})$ generations all bit positions have been converged at least once.*

The proof of the first statement of Lemma 10 is essentially the same as that of Lemma 4 where we apply the variable drift theorem again. The second statement follows by standard arguments, i. e., applying Markov's inequality iteratively in repeated phases to show that w. o. p. a given bit has converged at least once in the considered time phase and a union bound at the end to show that the same holds for all bits.

A byproduct of Lemma 10 is that if no mutation is used, then the algorithm will be stuck once all the bits have converged. Hence, before we present the final results on the bandwidth of the **SGA** in the next section, we use the previous lemma combined with Lemma 5 to show infinite runtime for **PR-SGA** w. o. p.

THEOREM 2. *Let $\mu \leq cn^{1/6-\varepsilon}$. With overwhelming probability the* **PR-SGA** *requires infinite time to optimize* ONEMAX.

PROOF. The proof follows the same idea and calculations of Theorem 1. W. o. p., at least $n^{1/2-\varepsilon}/3$ generations are required to find the optimum. However, again with overwhelming probability $1 - 2^{-\Omega(n^{\varepsilon/8})}$, after $\mu n^{\varepsilon/8} = O(n^{1/6})$ generations all the individuals in the population all have either a one-bit or a zero-bit in each position (using Lemma 10, which also applies if the mutation operator is removed).

The difference compared to Theorem 1 is that fitness-proportional selection is biased towards one-bits. Hence there is not a probability of 1/2 that all the individuals have a zero-bit at position i rather than a one-bit. We proceed by contradiction to prove that there is at least one position i where all the individuals have a zero-bit at this point of time. Assume to the contrary that after $O(n^{1/6})$ generations there is no position i where all the individuals in the population have a zero-bit. Then all the individuals have a one-bit in every position. We conclude that the algorithm has found the solution in $O(n^{1/6})$ generations contradicting the assumption that at least $n^{1/2-\varepsilon}/3$ generations are required with overwhelming probability. □

3. ANALYSIS OF THE SGA

The lower bounds proved for the **UR-SGA** and **PR-SGA** in theorems 1 and 2, respectively, rely on the fact that all bit positions will converge in few generations w. o. p. and that converged bits will stay converged forever since there is no mutation. With respect to the **SGA**, we will prove that almost all positions will be converged at any time w. o. p.; in other words, the diversity s is bounded. Throughout this section, we assume that $\mu \leq n^{1/8-\varepsilon}$ for some constant $\varepsilon > 0$.

3.1 Low Diversity and Bandwidth

To bound the diversity, we consider time phases consisting of $\mu n^{\varepsilon/8}$ generations. In the first phase, diversity will collapse, and this will be maintained for the following phases.

LEMMA 11. *Consider the* **SGA** *at some generation t, where $t \geq \mu n^{\varepsilon/8}$ and $t \leq 2^{n^{\varepsilon/10}}$. If all populations up to this generation are compact, then $s = O(\mu^2 n^{\varepsilon/8})$ at generation t with probability $1 - 2^{-\Omega(n^{\varepsilon/9})}$.*

PROOF. By Lemma 10, in the first $\mu n^{\varepsilon/8}$ generations, all bits have converged at least once with probability $1 - 2^{-\Omega(n^{\varepsilon/9})}$. Now we consider an upper bound on the number of bits that have left the converged state by the end of the phase.

We define an indicator random variable $X_{i,j,k}$ for the event that the converged state of bit k is left when creating the j-th individual in the i-th generation of the phase, where $1 \leq i \leq \mu n^{\varepsilon/8}$, $1 \leq j \leq \mu$ and $1 \leq k \leq n$.

To leave the converged state, the bit position must be flipped at least once. Since each bit flips with probability $1/n$, we get $P(X_{i,j,k} = 1) = 1/n$, and the expected value of the sum S of the $X_{i,j,k}$ is

$$E(S) = \sum_i \sum_j \sum_k P(X_{i,j,k}) = \frac{\mu \cdot T \cdot n}{n} = \mu \cdot T = \mu^2 n^{\varepsilon/8}$$

Obviously, S is an upper bound on the number of positions that leave the converged state. By Chernoff bounds

$E(S) \le 2\mu^2 n^{\varepsilon/8}$ with probability $1 - 2^{-\Omega(n^{\varepsilon/8})}$, which, together with the fact that all bits converge at least once in the phase proves the statement for generation $\mu n^{\varepsilon/8}$. For later generations, the statement follows by considering additional phases of length $\mu n^{\varepsilon/8}$. The total failure probability in at most $2^{n^{\varepsilon/10}}$ generations is still $2^{-\Omega(n^{\varepsilon/9})}$. □

In the following, we need to assume further properties of the current population, namely a sharper bound on the multiplicative bandwidth and not too extreme values of ℓ and h. In the end, we will conclude that all the properties are satisfied for an exponential number of generations w. o. p.

Definition 1. A population of the **SGA** is called *very compact* iff $h/\ell \le 1 + 1/(20\mu^2)$. It is called *centered* if $\ell \ge n/10$ and $h \le 9n/10$.

Lemma 12. *Assume a very compact population and consider a phase of length $\mu n^{\varepsilon/8}$ during which all populations are centered. Then during the phase the population is at least compact and after the phase the population is very compact with probability $1 - 2^{-\Omega(n^{\varepsilon/9})}$.*

Proof. If the multiplicative bandwidth is at most $1 + 1/(20\mu^2)$ then by simple calculations, the additive bandwidth is at most $c_\ell n/(20\mu^2)$, where $c_\ell n$ denotes the OneMax-value of the worst individual. In a centered population, $c_\ell \ge 1/10$. By Lemma 5, each generation increases the additive bandwidth by at most $n^{1/2+\varepsilon/8}$ w. o. p. In the proof of Lemma 11, we consider phases of length $\mu n^{\varepsilon/8}$. Hence, such a phase increases the additive bandwidth by at most $\mu n^{1/2+\varepsilon/4}$. Hence, the additive bandwidth throughout the phase is at most

$$\frac{c_\ell n}{20\mu^2} + \mu n^{1/2+\varepsilon/4} \;\le\; \frac{c_\ell n}{20\mu^2} + \frac{\mu n}{\mu^4 n^{5.25\varepsilon}} \;\le\; \frac{c_\ell n}{19\mu^2},$$

where the first inequality follows from our upper bound $\mu \le n^{1/8-\varepsilon}$ and the second one holds if n is large enough. Again by simple calculations, the multiplicative bandwidth during the phase is at most

$$1 + \frac{c_\ell n}{19\mu^2}\frac{1}{c_\ell n} \le 1 + \frac{1}{17\mu^2},$$

i. e., the populations are at least compact throughout the phase.

For compact populations, we can apply Lemma 11, which tells us that after every phase of length $\mu n^{\varepsilon/8}$ we have $s = O(\mu^2 n^{\varepsilon/8})$ with probability $1 - 2^{-\Omega(n^{\varepsilon/9})}$. We have already noted that $h - \ell \le s$. Hence, after the phase it holds w. o. p. that the multiplicative bandwidth is bounded from above by

$$1 + \frac{h-\ell}{\ell} \le 1 + \frac{O(\mu^2 n^{\varepsilon/4})}{c_\ell n} \le 1 + \frac{O(\mu^2)}{c_\ell \mu^8 n^{7.25\varepsilon}} \le 1 + \frac{1}{20\mu^2}$$

if n large enough. □

Moreover, as it will be shown in Section 3.2, the minimum OneMax-value of the population remains at least $c_\ell n$ for an exponential number of generations w. o. p. Hence by Lemma 12, the population will remain compact, and by Lemma 11 the diversity will be bounded for an exponential number of generations w. o. p. if our assumption on the minimum OneMax-value ℓ is valid. This will be dealt with in the next subsection. Note also that a compact population could still contain an optimal search point. Also this will be proved to be unlikely in the next subsection.

3.2 Drift of Best and Worst Fitness Values

The aim is to bound h and ℓ in a drift analysis using a so-called potential function. Similarly as in [12], the potential of an individual x is defined by $g(x) := e^{\kappa \text{OneMax}(x)}$ for some $\kappa := \kappa(n)$ to be chosen later, and $g(X) := \sum_{i=1}^{\mu} g(x_i)$ for every population $X := \{x_1, \dots, x_\mu\}$ (note that populations are multisets). Let us consider a current population at generation t and the process of creating the next population at generation $t+1$ (dropping the time indices unless there is risk of confusion). This process consists of μ consecutive operations choosing two parent individuals, crossing them over and mutating the result. Let P_i and Q_i be the two random parent individuals in the i-th operation (at generation t), $1 \le i \le \mu$, and let K_i be the random offspring. The next lemma notes an important observation on the OneMax-value of the offspring.

Hereinafter, $\Delta^{(\mathrm{m})}(j)$ denotes the random change in OneMax-value when applying standard bit mutation to an individual with j one-bits, $B(a,b)$ denotes a random variable following the binomial distribution with parameters a and b, and $H(\cdot,\cdot)$ denotes the Hamming distance.

Lemma 13. *It holds that*

$$|K_i| = \frac{|P_i| + |Q_i| + 2C(P_i,Q_i)}{2} + {} + \Delta^{(\mathrm{m})}(|P_i|/2 + |Q_i|/2 + C(P_i,Q_i)),$$

where $C(P_i,Q_i) \sim B(H(P_i,Q_i),1/2) - H(P_i,Q_i)/2$. Moreover,

$$|K_i| = \frac{|P_i| + 2C(P_i,Q_i) + 2\Delta^*(|P_i| + C(P_i,Q_i))}{2} + \frac{|Q_i| + 2C(P_i,Q_i) + 2\Delta^*(|Q_i| + C(P_i,Q_i))}{2},$$

where $\Delta^*(j) := B(n/2 - j/2, 1/n) - B(j/2, 1/n)$ is the random increase in one-bits given that each bit in a string of length $n/2$ with $j/2$ one-bits is flipped with probability $1/n$, i. e., half the standard mutation probability.

Proof. By definition, the crossover part of the i-th operation leads to $|P_i \cap Q_i| + B(H(P_i,Q_i),1/2)$ one-bits before mutation. Moreover $|P_i \cup Q_i| = |P_i \cap Q_i| + H(P_i,Q_i)$, which means that $(1/2)(|P_i| + |Q_i|) = |P_i \cap Q_i| + H(P_i,Q_i)/2$. Therefore, an individual with $(1/2)(|P_i| + |Q_i|) + C(P_i,Q_i)$ one-bits is subjected to mutation, which is the first statement of the lemma. The increase in one-bits due to mutation is a random variable with distribution

$$\begin{aligned}
&B(n - (|P_i|/2 + |Q_i|/2 + C(P_i,Q_i)), 1/n) \\
&\qquad - B(|P_i|/2 + |Q_i|/2 + C(P_i,Q_i), 1/n) \\
&= B(n/2 - |P_i|/2 - C(P_i,Q_i)/2, 1/n) \\
&\qquad + B(n/2 - |Q_i|/2 - C(P_i,Q_i)/2, 1/n) \\
&\qquad - \Big(B(|P_i|/2 + C(P_i,Q_i)/2, 1/n) \\
&\qquad + B(|Q_i|/2 + C(P_i,Q_i)/2, 1/n)\Big),
\end{aligned}$$

where the equality follows from the fact that if $X_1 \sim B(n_1,p)$ and $X_2 \sim B(n_2,p)$ then $X_1 + X_2 \sim B(n_1+n_2,p)$. The second statement follows now by regrouping terms. □

Due to linearity of expectation and $E(C(P_i,Q_i)) = 0$, we have $E(\Delta^*(|P_i| + C(P_i,Q_i))) = 1/2 - |P_i|/n$, and analogously for Q_i. This results in $E(|K_i|) = (|P_i|/2 + (1/2 -$

$|P_i|/n)) + (|Q_i|/2 + (1/2 - |Q_i|/n))$. In other words, the random K_i depends on the random P_i and Q_i, whereas $E(|K_i|)$ only depends on $|P_i|$ and $|Q_i|$, each of which has weight 1/2. When looking at $E(|K_i|)$, we see that one operation is "split" into two analogous terms, whose values are determined by $|P_i|$ and $|Q_i|$, respectively.

However, the random potential of the offspring is given by $e^{\kappa|K_i|}$, and we have to bound the expectation $E(e^{\kappa|K_i|})$. We will see below that $E(e^{\kappa|K_i|})$ is not too different from $e^{E(|K_i|)}$ if κ is small enough. Actually, we will also be able to "split" an operation into two terms that "basically" only depend on P_i and Q_i, respectively.

LEMMA 14.

$$E(e^{\kappa|K_i|}) \leq \frac{E(e^{\kappa(|P_i|+2C(P_i,Q_i)+2\Delta^*(|P_i|+C(P_i,Q_i)))})}{2} + \frac{E(e^{\kappa(|Q_i|+2C(P_i,Q_i)+2\Delta^*(|Q_i|+C(P_i,Q_i)))})}{2}$$

PROOF. The statement follows from Lemma 13 since the geometric mean is at most the arithmetic mean, i. e., $e^{a/2} \cdot e^{b/2} \leq e^a/2 + e^b/2$ for arbitrary a and b. □

Both terms on the right-hand side have in common that they depend on the random $C(P_i, Q_i)$. The influence of $C(P_i, Q_i)$ will be neglegible for sufficiently small κ, as the following lemma shows.

LEMMA 15. *Let $s = H(P_i, Q_i) \geq 1$. If $|P_i| \geq (1+c)(n/2)$ for some arbitrarily small constant $c > 0$ and $s \leq (c/4)n$ then choosing $\kappa \leq \frac{c}{20s}$ yields*

$$E(e^{\kappa(2C(P_i,Q_i)+2\Delta^*(|P_i|+C(P_i,Q_i)))}) \leq 1 - \Omega(\kappa).$$

If the assumption on $|P_i|$ is dropped and c is small enough then

$$E(e^{\kappa(2C(P_i,Q_i)+2\Delta^*(|P_i|+C(P_i,Q_i)))}) \leq 1 + O(\kappa).$$

PROOF. $\Psi(P_i, Q_i) := e^{\kappa(2C(P_i,Q_i)+2\Delta^*(|P_i|+C(P_i,Q_i)))}$, the random variable considered here, is dependent on the combined effect of crossover and mutation. Note that Δ^* is decreasing in its argument and that we have $C(P_i, Q_i) \geq -s$. In the following, we work with the upper bound

$$\Psi(P_i, Q_i) \leq e^{2\kappa C(P_i,Q_i)} \cdot e^{2\kappa\Delta^*(|P_i|-s)}$$

and assume that $|P_i| \geq (1 + c)(n/2)$.

We concentrate first on the first expectation. By definition, $C(P_i, Q_i) \sim B(s, 1/2) - s/2$. Using the moment-generation of the binomial distribution, we obtain

$$E(e^{2\kappa C(P_i,Q_i)}) = E(e^{2\kappa(-s/2)} \cdot e^{2\kappa B(s,1/2)}) = e^{-\kappa s} \cdot \left(\frac{1}{2} + \frac{1}{2}e^{2\kappa}\right)^s.$$

Assuming that $\kappa \leq 1/2$, we use the inequality $e^x \leq 1 + x + x^2$ for $x \leq 1$ and obtain

$$E(e^{2\kappa C(P_i,Q_i)}) \leq e^{-\kappa s} \cdot \left(1 + \kappa + 2\kappa^2\right)^s \leq e^{-\kappa s} e^{(\kappa+2\kappa^2)s} = e^{2\kappa^2 s},$$

which for $\kappa = c/(20s)$ (a choice that will turn out useful later) gives us the upper bound

$$E(e^{2\kappa C(P_i,Q_i)}) \leq e^{2c^2/(400s)} \leq 1 + \frac{c^2}{100s}$$

using $e^x \leq 1 + 2x$ for $x \leq 1$ and assuming $2c^2/(400s) \leq 1$.

Next we deal with the effect of mutation, more precisely we bound the expected value of $e^{2\kappa\Delta^*(|P_i|-s)}$. Following the proof of the simplified drift theorem [17], we first bound the plain drift $E(\Delta^*(|P_i| - s))$ and then its moment-generating function. Recall that $\Delta^*(j) = B(n/2-j/2, 1/n) - B(j/2, 1/n)$ is the random increase in OneMax-value when a bit string of length $n/2$, containing $j/2$ ones, is subject to standard bit mutation with probability $1/n$. Hence, we have

$$E(\Delta^*(|P_i| - s)) = \frac{1}{2} - \frac{|P_i| - s}{n} \leq \frac{-c}{4},$$

where the last inequality follows by the assumptions made in the lemma.

Moreover, we know that the number of flipping bits follows an exponential decay, more precisely

$$\text{Prob}(\Delta^*(i) = z) \leq \frac{n/2}{|z|}\left(\frac{1}{n}\right)^{|z|} \leq \frac{1}{|z|!} \leq e^{-|z|+2}$$

for any i and any $z \in \mathbb{Z}$. If $\lambda(n) = 1/2$, this implies

$$E(e^{\lambda(n)\Delta^*(i)}) = \sum_{z\in\mathbb{Z}} e^{\lambda(n)z}\,\text{Prob}(\Delta^*(i) = z)$$
$$\leq \sum_{z>0} e^{\lambda(n)z}e^{-|z|+2} + \sum_{z\leq 0} e^{-|z|+2} \leq \sum_{z\geq 1} e^{2-z/2} + \frac{e^2}{1-e^{-1}}$$
$$= \frac{e^{3/2}}{1-e^{-1/2}} + \frac{e^2}{1-e^{-1}} < 24.$$

Now, expanding the moment-generating function, we get for any $\kappa = \kappa(n) \leq \lambda(n) = 1/2$ that

$$E(e^{\kappa\Delta^*(i)}) = 1 + \kappa E(\Delta^*(i)) + \sum_{z=2}^{\infty} \frac{E((\kappa\Delta^*(i))^z)}{z!}$$
$$\leq 1 + \kappa E(\Delta^*(i)) + \kappa^2(n) \sum_{z=0}^{\infty} \frac{\frac{E((\lambda(n)\Delta^*(i))^z)}{z!}}{\lambda^2(n)}$$
$$\leq 1 + \kappa E(\Delta^*(i)) + \frac{\kappa^2(n)}{1/4} E(e^{\lambda(n)\Delta^*(i)})$$
$$\leq 1 + \kappa E(\Delta^*(i)) + 96\kappa^2(n)$$

If $2\kappa \leq -E(\Delta^*(i))/192$, a simple upper bound is obtained from this as follows:

$$E(e^{2\kappa\Delta^*(i)}) \leq 1 + 2\kappa E(\Delta^*(i)) - \frac{1}{2}(2\kappa)E(\Delta^*(i)) \leq 1 + \kappa E(\Delta^*(i)).$$

Since $E(\Delta^*(i)) \leq -c/4$ and thus $-E(\Delta^*(i))/192 \geq c/768$, the choice $\kappa := c/(20s)$ from above obviously satisfies the condition (if n is not too small and $s = \omega(1)$) and we get

$$E(e^{2\kappa\Delta^*(i)}) \leq 1 - \frac{c^2}{80s}.$$

Altogether, the random variable under consideration has been bounded according to

$$\Psi(P_i, Q_i) \leq \left(1 + \frac{c^2}{100s}\right)\left(1 - \frac{c^2}{80s}\right) \leq 1 - \frac{c^2}{400s},$$

which is $1 - \Omega(\kappa)$.

If the assumption on $|P_i|$ is dropped, then we work with the trivial bound $E(\Delta^*(i)) \leq 1$. Recalling that $E(e^{\kappa\Delta^*(i)}) \leq 1 + \kappa E(\Delta^*(i)) + 96\kappa^2(n)$, we get $E(e^{\kappa\Delta^*(i)}) \leq 1 + 192c/(20s)$

if c/s is small enough and $s \geq 1$. Then (for small enough c)

$$\Psi(P_i, Q_i) \leq \left(1 + \frac{c^2}{100s}\right)\left(1 + \frac{192c}{20s}\right)^2 \leq 1 + \frac{20c}{s},$$

which is $1 + O(\kappa)$. □

Our aim is to bound $E(g(X_{t+1}) - g(X_t) \mid X_t)$. If we know the expected value of the random number of times S_i that individual $x_i := x_i^{(t)}$ is chosen as first or second parent in the μ operations, then the inequalities from Lemma 13 and 14 allow us to bound the population drift in the following way:

$$E(g(X_{t+1}) \mid X_t)) \leq \sum_{i=1}^{\mu} E(S_i) \cdot \frac{E(e^{\kappa(|x_i| + 2C(x_i, x_i') + 2\Delta^*(|x_i| + C(x_i, x_i')))})}{2}. \quad (1)$$

Here x_i' denotes the random other parent in the operation choosing x_i. The following simple lemma bounds $E(S_i)$.

Lemma 16. *$E(S_i) \leq 2h/\ell$.*

The straightforward proof has been omitted. More effort is needed to bound the second expectation in (1). If $g(X_t)$ is large, then the following lemma will help us to obtain a negative drift. Hereinafter, a sufficiently small constant $c > 0$ and the choice $\kappa := c/(20s)$ are fixed.

Lemma 17. *Suppose that $s \leq n^{1/4-\varepsilon}$ and $\mu = poly(n)$. If $g(X) \geq e^{\kappa(1+2c)n/2}$, then there is a non-empty set $X^* \subset X$ of individuals $x \in X$ satisfying $|x| \geq (1+c)n/2$. Moreover, $g(X) = (1 + 2^{-\Omega(n^{3/4})}) \sum_{x \in X^*} g(x)$.*

Proof. Asssume $X^* = \emptyset$. Then $g(X) \leq \mu e^{\kappa(1+c)n/2} = e^{\kappa(1+c)n/2 + \ln \mu}$. Since $\kappa = c/(20s) = \Omega(n^{-1/4})$ and $\ln \mu = O(\log n)$ by our assumption, we arrive at the contradiction $\kappa(1+c)n/2 + \ln \mu \leq \kappa(1+1.5c)n/2$ if n is not too small. The second claim follows since $\kappa(1+1.5c)n/2 = \kappa(1+2c)n/2 - \Omega(n^{3/4+\varepsilon})$. □

From now on, also $s \leq n^{1/4-\varepsilon}$ is assumed (which follows from Lemma 11 for $\mu \leq n^{1/8-\varepsilon}$). The next lemma states a multiplicative drift of the potential away from large values.

Lemma 18. *If $g(X_t) \geq e^{\kappa(1+2c)n/2}$, then*

$$E(g(X_{t+1}) \mid X_t) \leq (1 - \Omega(\kappa)) \cdot g(X_t).$$

Proof. According to Lemma 17, there is a subset $X^* \subset X_t$ such that

$$g(X_t) = \sum_{x \in X^*} g(x) + 2^{-\Omega(n^{3/4})} \sum_{x \in X_t} g(x).$$

We have already argued that

$$E(g(X_{t+1}) \mid X_t)) \leq \sum_{i=1}^{\mu} E(S_i) \cdot \frac{E(e^{\kappa(|x_i| + 2C(x_i, x_i') + 2\Delta^*(|x_i| + C(x_i, x_i')))})}{2}.$$

By Lemma 16, $E(S_i) \leq 2h/\ell \leq 2(\ell + s)/\ell$. Since a centered population is assumed ($\ell \geq n/10$), we get $E(S_i) = 2 + O(s/n) = 2 + O(n^{-3/4})$. Hence, for those i such that $x_i \in X^*$ we get from Lemma 15 that

$$E(S_i) \cdot \frac{E(e^{\kappa(|x_i| + 2C(x_i, x_i') + 2\Delta^*(|x_i| + C(x_i, x_i')))})}{2}$$
$$\leq (1 + O(n^{-3/4}))(1 - \Omega(\kappa))e^{|x_i|} = (1 - \Omega(\kappa))e^{|x_i|}$$

using $\kappa = \Omega(n^{-1/4})$. For the $x_i \notin X^*$ we know by Lemma 15 that

$$E(S_i) \cdot \frac{E(e^{\kappa(|x_i| + 2C(x_i, x_i') + 2\Delta^*(|x_i| + C(x_i, x_i')))})}{2}$$
$$\leq (1 + O(n^{-3/4}))(1 + O(\kappa))e^{|x_i|} = (1 + O(\kappa))e^{|x_i|}.$$

Altogether,

$$E(g(X_{t+1}) \mid X_t))$$
$$\leq (1 - \Omega(\kappa)) \left(\sum_{x \in X^*} g(x)\right) + 2^{-\Omega(n^{3/4})}(1 + O(\kappa)) \sum_{x \notin X^*} g(X)$$
$$= \left(1 - \Omega(\kappa) + 2^{-\Omega(n^{3/4})}\right) g(X_t) = (1 - \Omega(\kappa)) g(X_t). \quad \square$$

We are almost ready to apply the following simplified negative-drift theorem, whose proof follows the lines of [14].

Theorem 3 (Simplified Drift with Scaling). *Let X_t, $t \geq 0$, be the random variables describing a stochastic process over a finite state space $S \subseteq \mathbb{R}$ and denote $\Delta_t(i) := (X_{t+1} - X_t \mid X_t = i)$ for $i \in S$ and $t \geq 0$. Suppose there exist an interval $[a, b]$ in the state space, and, possibly depending on $\ell := b - a$, a bound $\varepsilon(\ell) > 0$ and a scaling factor $r(\ell)$ such that for all $t \geq 0$ the following three conditions hold:*

1. *$E(\Delta_t(i)) \geq \varepsilon(\ell)$ for $a < i < b$,*
2. *$\text{Prob}(\Delta_t(i) \leq -j \cdot r(\ell)) \leq e^{-j+1}$ for $i > a$ and $j \in \mathbb{N}$,*
3. *$r(\ell) \leq \min\{\ell, \sqrt{\varepsilon(\ell)\ell/(1352 \log(\ell/\varepsilon(\ell)))}\}$.*

For the first hitting time $T^* := \min\{t \geq 0 \colon X_t \leq a \mid X_0 \geq b\}$ it then holds

$$\text{Prob}(T^* \leq e^{\varepsilon(\ell)\ell/(1352 r^2(\ell))}) = O(e^{-\varepsilon(\ell)\ell/(1352 r^2(\ell))}).$$

Lemma 18 states a multiplicative drift but Theorem 3 is for an additive setting. Hence, as in [12], we switch over to the potential function $g'(X) := \ln g(X)$. Since ln is concave, Jensen's inequality yields

$$E(g'(X_{t+1}) \mid X_t) = E(\ln(g(X_{t+1})) \mid X_t)$$
$$\leq \ln(E(g(X_{t+1}) \mid X_t).$$

Hence, if $g'(X_t) \geq \kappa(1+2c)n/2$ then by Lemma 18

$$E(g'(X_{t+1}) \mid X_t) \leq \ln((1 - \Omega(\kappa))g(X_t))$$
$$= \ln(1 - \Omega(\kappa)) + \ln(g(X_t)) = -\Omega(\kappa) + g'(X_t),$$

which establishes the additive drift $E(g'(X_{t+1}) - g'(X_t) \mid X_t) = -\Omega(\kappa)$. Using the potential function $g''(X) := \kappa(1 + 4c)n/2 - g'(X)$, where the negation is necessary to fit the perspective of Theorem 3, and the drift interval $a := 0, b := \kappa(1+4c)n/2 - \kappa(1+2c)n/2 = \kappa cn$, we obtain a drift of $\Omega(\kappa)$ for all $g''(X)$ such that $a \leq g''(X) \leq b$. The first condition of Theorem 3 has been established. Moreover, by Chernoff bounds $g''(X_0) \geq b$ at initialization of the **SGA** with probability $1 - 2^{-\Omega(n^{1/2})}$, and using Lemma 5, we still have $g''(X_{\mu n^{\varepsilon/8}}) \geq b$, i. e., when diversity has collapsed for the first time and the drift analysis is started in reality.

For the second condition, we set $r(\ell) := \max\{s, n^{1/4-\varepsilon}\} = O(n^{1/4-\varepsilon})$. Crossover can change the potential by at most $s \leq r(\ell)$. To change the potential by $j \cdot r(\ell)$, at least

$(j-1)r(\ell) \geq r(\ell)$ bits have to flip in at least one of the $\mu = poly(n)$ mutations that happen in a generation. This probability is easily bounded by e^{-j+1}. This verifies the second condition. Altogether, the parameters of the drift theorem satisfy $\ell = b - a = \Omega(n^{3/4+2\varepsilon})$, $\varepsilon(\ell) = \Omega(n^{-1/4+2\varepsilon})$. Since $\varepsilon(\ell)\ell/r^2(\ell) = \Omega(n^{6\varepsilon})$, the time to pass the drift interval is $2^{\Omega(n^{6\varepsilon})}$ w. o. p. If $g(X) \leq e^{\kappa(1+4c)n/2}$ by definition no individual from X can have more than $(1+4c)n/2$ one-bits, in particular the optimum is not reached.

A symmetrical argument can be applied to the minimum OneMax-value of the individuals in the population. Putting everything together gives us the following lemma, which says that the population is centered for an exponential number of generations, assuming that the diversity is bounded.

Lemma 19. *Assuming $s \leq n^{1/4-\varepsilon}$ the whole time, all populations up to generation $2^{c' n^{6\varepsilon}}$, for some constant $c' > 0$, satisfy $\ell \geq (1-c)n/2$ and $h \leq (1+c)n/2$ with probability $1 - 2^{-\Omega(n^{6\varepsilon})}$, where $c > 0$ is an arbitrarily small constant.*

But since these bounds on h and ℓ are enough to make the analysis from the previous subsection work, we also have $s \leq n^{1/4-2\varepsilon}$ for an exponential number of generations w. o. p., altogether a compact and centered population for an exponential number of generations. In each generation there is a probability of only $2^{-\Omega(n^{\varepsilon/9})}$ that one of our assumptions (compact and centered population) is not satisfied. Since the sum of the failure probabilities within $2^{n^{\varepsilon/10}}$ generations is still $2^{-\Omega(n^{\varepsilon/9})}$, we have proved the following main result.

Theorem 4. *Let $\mu \leq n^{1/8-\varepsilon}$ for an arbitrarily small constant $\varepsilon > 0$. Then with probability $1 - 2^{-\Omega(n^{\varepsilon/9})}$, the* **SGA** *on* OneMax *does not create individuals with more than $(1+c)n/2$ or less than $(1-c)n/2$ one-bits, where $c > 0$ is an arbitrarily small constant, within the first $2^{n^{\varepsilon/10}}$ generations. In particular it does not reach the optimum then.*

4. CONCLUSIONS

A runtime analysis of the **SGA** for OneMax has been presented. It has been proved that the algorithm cannot optimize the function as long as the population is not larger than $\mu \leq n^{1/8-\varepsilon}$. Preliminary experimental results do not show significant difference in the performance of the **SGA** with increased population size compared to the mutation only fitness-proportional EA previously analysed in [12] with arbitrary population sizes. Hence, it remains an open problem to extend the results presented herein to the **SGA** with larger population sizes. In the same paper it was proved that appropriate scaling mechanisms turn the mutation-only EA into an efficient algorithm. The same results would carry on for the **SGA** in a straightforward manner by performing similar analyses.

5. REFERENCES

[1] A. Auger and B. Doerr eds. *Theory of Randomized Seach Heuristics–Foundations and Recent Developments*. World Scientific, 2010.

[2] B. Doerr, E. Happ, and C. Klein. Crossover can provably be useful in evolutionary computation. In *Proc. of GECCO '08*, pages 539–546, 2008.

[3] B. Doerr, M. Fouz, and C. Witt. Sharp bounds by probability-generating functions and variable drift. In *Proc. of GECCO '11*, pages 2083–2090, 2011.

[4] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and machine learning*. Addison-Wesley, 1989.

[5] E. Happ, D. Johannsen, C. Klein, and F. Neumann. Rigorous Analyses of Fitness-Proportional Selection for Optimizing Linear Functions. In *Proc. of GECCO '08*, pages 953–960, 2008.

[6] J. Jägersküpper and C. Witt. Rigorous runtime analysis of a (μ+1) ES for the Sphere function. In *Proc. of GECCO '05*, pages 849–856, 2005.

[7] T. Jansen and I. Wegener. Real royal road functions: where crossover provably is essential. *Discrete Applied Mathematics*, 149(1-3):111–125, 2005.

[8] T. Kötzing, D. Sudholt, and M. Theile. How crossover helps in pseudo-boolean optimization. In *Proc. of GECCO '11*, pages 989–996, 2011.

[9] P. K. Lehre. Negative drift in populations. In *Proc. of PPSN '10*, pages 244–253, 2010.

[10] P. K. Lehre. Fitness-levels for non-elitist populations. In *Proc. of GECCO '11*, pages 2075–2082, 2011.

[11] F. Neumann, P. S. Oliveto, G. Rudolph, and D. Sudholt. On the effectiveness of crossover for migration in parallel evolutionary algorithms. In *Proc. of GECCO '11*, pages 1587–1594, 2011.

[12] F. Neumann, P. S. Oliveto, and C. Witt. Theoretical analysis of fitness-proportional selection: Landscapes and efficiency. In *Proc. of GECCO '09*, pages 835–842, 2009.

[13] F. Neumann, D. Sudholt, and C. Witt. Analysis of different MMAS ACO algorithms on unimodal functions and plateaus. *Swarm Intelligence*, 3(1):35–68, 2009.

[14] F. Neumann, D. Sudholt, and C. Witt. A few ants are enough: ACO with iteration-best update. In *Proc. of GECCO '10*, pages 63–70, 2010.

[15] F. Neumann and C. Witt. *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Springer, 2010.

[16] P. S. Oliveto, J. He, and X. Yao. Analysis of population-based evolutionary algorithms for the vertex cover problem. In *Proc. of CEC '08*, pages 1563–1570, 2008.

[17] P. S. Oliveto and C. Witt. Simplified drift analysis for proving lower bounds in evolutionary computation. *Algorithmica*, 59(3):369–386, 2011.

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

[19] M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press, 1998.

[20] R. A. Watson and T. Jansen. A building-block royal road where crossover is provably essential. In *Proc. of GECCO '07*, pages 1452–1459, 2007.

[21] C. Zarges. On the utility of the population size for inversely fitness proportional mutation rates. In *Proc. of FOGA '09*, pages 39–46, 2009.