



## Lower bounds on the runtime of crossover-based algorithms via decoupling and family graphs

Sutton, Andrew M.; Witt, Carsten

*Published in:*

Proceedings of the 2019 Genetic and Evolutionary Computation Conference

*Link to article, DOI:*

[10.1145/3321707.3321848](https://doi.org/10.1145/3321707.3321848)

*Publication date:*

2019

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Sutton, A. M., & Witt, C. (2019). Lower bounds on the runtime of crossover-based algorithms via decoupling and family graphs. In *Proceedings of the 2019 Genetic and Evolutionary Computation Conference* (pp. 1515-1522). Association for Computing Machinery. <https://doi.org/10.1145/3321707.3321848>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Lower Bounds on the Runtime of Crossover-Based Algorithms via Decoupling and Family Graphs

Andrew M. Sutton

University of Minnesota Duluth  
Duluth, MN, USA  
amsutton@d.umn.edu

Carsten Witt

Technical University of Denmark  
Kgs. Lyngby, Denmark  
cawi@dtu.dk

## ABSTRACT

The runtime analysis of evolutionary algorithms using crossover as search operator has recently produced remarkable results indicating benefits and drawbacks of crossover and illustrating its working principles. Virtually all these results are restricted to upper bounds on the running time of the crossover-based algorithms. This work addresses this lack of lower bounds and rigorously bounds the optimization time of simple algorithms using uniform crossover on the search space  $\{0, 1\}^n$  from below via two novel techniques called decoupling and family graphs. First, a simple steady-state crossover-based evolutionary algorithm without selection pressure is analyzed and shown that after  $O(\mu \log \mu)$  generations, bit positions are sampled almost independently with marginal probabilities corresponding to the fraction of one-bits at the corresponding position in the initial population. Afterwards, a crossover-based algorithm using tournament selection is analyzed by a novel generalization of the family tree technique originally introduced for mutation-only EAs. Using these so-called family graphs, almost tight lower bounds on the optimization time on the ONEMAX benchmark function are shown.

## CCS CONCEPTS

• Theory of computation → Theory of randomized search heuristics;

## KEYWORDS

randomised search heuristics, crossover, decoupling, runtime analysis

## ACM Reference Format:

Andrew M. Sutton and Carsten Witt. 2019. Lower Bounds on the Runtime of Crossover-Based Algorithms via Decoupling and Family Graphs. In *Genetic and Evolutionary Computation Conference (GECCO '19), July 13–17, 2019, Prague, Czech Republic*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3321707.3321848>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

GECCO '19, July 13–17, 2019, Prague, Czech Republic

© 2019 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

ACM ISBN 978-1-4503-6111-8/19/07...\$15.00

<https://doi.org/10.1145/3321707.3321848>

## 1 INTRODUCTION

Understanding the influence of crossover on an evolving population of individuals is an important problem from the perspective of both evolutionary computation and population genetics. In this paper, we present two novel techniques that might be useful in proving lower bounds on the running time of processes that use crossover. The first technique, called *decoupling*, uses drift analysis to study the approach of the process toward so-called *linkage equilibrium*: the state in which the statistical dependence of alleles on specific individuals has become negligible. The second technique, called *family graphs*, adapts ancestral traces into a graph structure to compute lower bounds. For decoupling, we provide a proof of concept for its utilization. In the case of family graphs, we are able to show almost tight lower bounds on a steady-state genetic algorithm implementing probabilistic binary tournament selection.

Our results address an intensively researched subject. The study of crossover in the theory community has developed over time with a series of works that aimed to find environments in which crossover provides provable performance gains [5, 6, 11, 13, 15, 18, 19, 22, 25, 27, 29]. Crossover has also been proved to protect against pathological mutation rates. On the class of monotone functions, Lengler [17] recently established precise values for the critical mutation rates on a number of mutation-based algorithms. Above these critical rates, mutation alone cannot efficiently optimize all monotone functions. However, crossover and mutation together can efficiently optimize all monotone functions, regardless of mutation strength. Crossover and mutation can also work together to create sufficient diversity to optimize jump functions [3].

Most runtime results on crossover are in the form of upper bounds. A notable exception is the work by Oliveto and Witt [20], who proved exponential lower bounds on a generational genetic algorithm (GA) with uniform crossover, mutation and fitness-proportional selection on ONEMAX for populations smaller than  $n^{1/4}$ . The analysis is based on a technical drift analysis and shows that selective pressure is low enough so that genetic drift occurs to move the frequency of at least one allele down to zero with high probability. Our contribution here is different since it aims at more general tools for the proof of lower bounds in GAs. Another exception are lower bounds on the runtime of crossover-based algorithms for the all-pairs shortest-path problems [8]; again our contribution is different from this analysis due to its generality.

The dynamics of crossover has also been studied in the context of nonlinear dynamical systems, specifically so-called *quadratic dynamical systems* in which a probability distribution evolves over time governed by random pairwise interactions [24]. This

framework, which expresses several important models in the natural sciences (such as Boltzmann's ideal gas law and the Hardy-Weinberg model of population genetics), can be specialized to so-called *crossover systems* in which populations evolve by random mating and crossover [23]. An important question from the perspective of population genetics is how quickly such a system approaches linkage equilibrium, i.e., when the statistical influence between alleles is sufficiently decoupled. Geiringer [9] showed that in the absence of selective forces, every infinite population trajectory converges to a stationary population that corresponds to linkage equilibrium.

An important question for such systems is the rate of convergence to the equilibrium distribution, a general technique for tight bounds on the rate of convergence in the infinite population case is given by Rabani, Rabinovich and Sinclair [23]. This result was refined to finite populations by Ollivier [21] who proved that a *generational* finite population model (in which the population at time  $t - 1$  entirely replaces the population at time  $t$ ) mixes to its equilibrium distribution in time  $O(\log n)$  for sufficiently large populations ( $\mu \gg n^2$ ). In the context of these results, to our knowledge, our decoupling time argument yields the first tight bound on the linkage equilibrium convergence rate in a steady-state process, i.e., in which individuals can survive for multiple generations.

The concrete algorithm we study in this paper is a generalization of the StSt  $\binom{\mu}{2}$  GA that was recently introduced as a simplified model of the estimation-of-distribution (EDA) algorithm cGA [32]. Using our novel techniques, we prove runtime bounds that reveal that the StSt  $\binom{\mu}{2}$  GA has a runtime of  $\tilde{O}(\sqrt{n}\mu)$  on ONEMAX, which is identical up to polylogarithmic factors to the runtime of the cGA.<sup>1</sup> This is interesting because it links the behavior of the EDA to a GA that employs standard uniform crossover and highlights the importance of crossover and binary tournament selection on this behavior.

*Prior work regarding family trees and graphs.* It is natural to represent the history of individuals in population-based algorithms using tree or more general graph structures. To study decoupling in generational GAs using infinite populations, such graphs appear, e.g., in the work [23] mentioned above. From the perspective of proving lower bounds on the runtime of evolutionary algorithms, analyses of family trees can be traced back to [30], where simple steady-state EAs using fitness-proportional selection were considered. However, a more thorough introduction to the proof technique is given in [31] with respect to the  $(\mu+1)$  EA. Family trees were also used to prove lower bounds on the runtime of evolution strategies [10] and immune algorithms [12] and were used in general theoretical frameworks for the analysis of mutation-selection balance in EAs [14, 16]. Since then, the family tree technique has been extended to cover  $(\mu+\lambda)$  EAs and memetic algorithms [26], and it was very recently used to prove tight bounds on the runtime of  $(\mu+\lambda)$  EAs [1].

This paper is structured as follows. In Section 2, we introduce the GAs considered and state major tools for the analysis. Section 3 is concerned with proving upper bounds on the decoupling time and illustrates how knowledge of the decoupling time can be used to bound the runtime from below. Section 4 shows upper bounds on

the optimization time of the considered GA on ONEMAX, depending on population size, number of bits and selection probability  $\epsilon$ . Afterwards, Section 5 proves general lower bounds using the family graph technique, depending on population size and number of bits. We finish with some conclusions.

## 2 PRELIMINARIES

We consider the StSt  $\binom{\mu}{2}$  GA with selection parameter  $\epsilon$  (StSt  $\binom{\mu}{2}$  GA $_\epsilon$  for short) illustrated in Algorithm 1. The algorithm maintains a population  $P_t$ ,  $t \geq 0$  of  $\mu$  length- $n$  binary strings, initially chosen uniformly at random. In each iteration, the algorithm chooses two strings  $x$  and  $y$  from the population without replacement, and produces offspring  $x'$  and  $y'$  by using uniform crossover. In particular, for each  $i \in \{1, \dots, n\}$ , independently and with probability  $1/2$ ,  $x'_i = y_i$  and  $y'_i = x_i$ , otherwise  $x'_i = x_i$  and  $y'_i = y_i$ .

After the offspring are produced, a probabilistic tournament selection is applied. This means that with probability  $\epsilon$ , the fitter of the  $x'$  and  $y'$  replaces the other. Finally both  $(x', y')$  replace  $(x, y)$  to create the new population. In this way, setting  $\epsilon := 1$  recovers the StSt  $\binom{\mu}{2}$  GA introduced in [32], and setting  $\epsilon := 0$  results in a steady-state variant of the quadratic dynamical system studied by Rabani et al. [23] and Ollivier [21].

We assume the optimization of a function  $f: \{0, 1\}^n \rightarrow \mathbb{R}$ , and define the running time of Algorithm 1 as the smallest  $t$  such that the optimum of  $f$  is in the population  $P_t$ . It will be convenient to define the term *frequency* of a bit as the fraction of ones at a bit position in the population. Formally, the frequency is given by  $p_{i,t} := |\{x \in P_t \mid x_i = 1\}|/\mu$ .

---

### Algorithm 1: StSt $\binom{\mu}{2}$ GA $_\epsilon$

---

```

 $t \leftarrow 0$ ;  $P_0$  consists of  $\mu$  individuals chosen u. a. r.;
while termination criterion not met do
    Choose  $x$  and  $y$  from  $P_t$  uniformly at random without
    replacement;
    Apply uniform crossover to  $x$  and  $y$ , resulting in the two
    offspring  $x'$  and  $y'$ ;
    with probability  $\epsilon$  do
        if  $f(x') < f(y')$  then swap  $x'$  and  $y'$ ;
         $y' \leftarrow x'$ ;
    Create  $P_{t+1}$  from  $P_t$  by replacing  $x$  with  $x'$  and  $y$  with
     $y'$ ;
     $t \leftarrow t + 1$ ;
    
```

---

In Section 4, we will use the following interesting converse of Jensen's inequality.

**LEMMA 1** (INEQ. (2) IN [2]). *Let  $I$  be an interval in  $\mathbb{R}$ ,  $f: I \rightarrow \mathbb{R}$  be a convex function on  $I$ ,  $[m, M] \subseteq I$ ,  $-\infty < m < M < +\infty$ ,  $x = (x_1, \dots, x_n)$  any  $n$ -tuple in  $I^n$  and  $\mathbf{p} = (p_1, \dots, p_n)$  a non-negative  $n$ -tuple such that  $P_n = \sum_{i=1}^n p_i > 0$ . Then*

$$\frac{1}{P_n} \sum_{i=1}^n p_i f(x_i) \leq \frac{M - \bar{x}}{M - m} f(m) + \frac{\bar{x} - m}{M - m} f(M),$$

where  $\bar{x} = (1/P_n) \sum_{i=1}^n p_i x_i$ .

<sup>1</sup>The asymptotic notation  $\tilde{O}(f(n))$  hides polylog( $n$ ) factors.

As a main tool for the analysis of the stochastic processes we encounter, we will use the following drift theorem, which for simplicity is stated for discrete state spaces only.

**THEOREM 2 (MULTIPLICATIVE DRIFT [4, 7]).** *Let  $X^{(t)}$ ,  $t \geq 0$ , be a stochastic process ranging over a discrete set  $S \cup \{0\}$ , where  $S = [x_{\min}, x_{\max}] \subseteq \mathbb{R}^+$  is a positive interval. Furthermore, let  $T$  denote the first hitting time  $t$  where  $X^{(t)} = 0$ . Suppose there exists a real number  $\delta > 0$  such that for all  $t < T$  and  $x \in S$  where  $\Pr(X^{(t)} = x) > 0$ ,*

$$\mathbb{E}\left(X^{(t)} - X^{(t+1)} \mid X^{(t)} = x\right) \geq \delta x.$$

Then, for all  $x_0 \in S$  where  $\Pr(X^{(t)} = x_0) > 0$ ,

$$\mathbb{E}\left(T \mid X^{(0)} = x_0\right) \leq \frac{\ln(x_0/x_{\min}) + 1}{\delta}$$

and, for any  $r > 0$ ,

$$\Pr(T \geq (\ln(x_0/x_{\min}) + r)/\delta) \leq e^{-r}.$$

The following lemma dealing with the number of one-bits of offspring and bounds on the increase of this number will be crucial for our analyses. It follows by a simple reformulation of the stochastic process generating the offspring via uniform crossover, and standard Chernoff bounds. Related statements are well-known in the literature, e. g., in Lemma 13 of [20]. We write  $\text{Bin}(n, p)$  to denote random variables following the binomial distribution with parameters  $n$  and  $p$ .

**LEMMA 3.** *Let  $x, y \in \{0, 1\}^n$  and let  $a := |\{i \mid x_i = y_i = 1\}|$  denote the number of common one-bits of  $x$  and  $y$  and  $b := H(x, y)$  their Hamming distance. Then for the offspring  $x'$  and  $y'$  produced by uniform crossover on  $x$  and  $y$  it holds that  $|x'| = a + \text{Bin}(b, 1/2)$  and (not independently)  $|y'| = a + \text{Bin}(b, 1/2)$ . In particular,  $\max\{|x'|, |y'|\} - \max\{|x|, |y|\} \leq n^{1/2} \ln n$  with probability  $1 - n^{-\omega(1)}$ . Furthermore,  $\max\{|x'|, |y'|\} - \max\{|x|, |y|\} \leq n^{1/2+\delta}$  with probability  $1 - 2^{-\Omega(n^{2\delta})}$  for  $0 < \delta < 1$ .*

### 3 BOUNDS ON THE DECOUPLING TIME

We first investigate the decoupling time of the  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$  with  $\epsilon = 0$ . In this setting, selection is absent and we want to determine how quickly the bits become independent. We consider an arbitrary but fixed bit  $b^*$  and investigate its ancestral line. If nothing else is mentioned, all considerations are with respect to this bit.

For  $i \in \{1, \dots, \mu\}$  and  $t \in \mathbb{N}$ , let  $D_t^{(i)}$  be the probability distribution on  $\{1, \dots, \mu\}$  where  $\Pr(D_t^{(i)} = j)$  is the probability that bit  $b^*$  in individual  $i$  at time  $t$  has descended from the initial individual  $j \in \{1, \dots, \mu\}$ . We use  $D_t^{(i)}(j)$  as a shorthand for  $\Pr(D_t^{(i)} = j)$ . Note that for all  $i \in \{1, \dots, \mu\}$  we have  $D_0^{(i)}(i) = 1$  and  $D_0^{(i)}(j) = 0$  for  $j \neq i$ . Clearly, since it is a distribution,  $\sum_{j=1}^{\mu} D_t^{(i)}(j) = 1$  for all  $i \in \{1, \dots, \mu\}$  and all  $t \in \mathbb{N}$ .

We will also need the following fact, which is in a sense dual:

**LEMMA 4.** *For all  $j \in \{1, \dots, \mu\}$  and all  $t \geq 0$ ,  $\sum_{i=1}^{\mu} D_t^{(i)}(j) = 1$ .*

**PROOF.** Note that  $D_0^{(i)}(j) = \delta_{ij}$ , and we proceed by induction on  $t$ . If crossover occurs between two individuals  $i$  and  $\ell$  at time  $t$ , the probability mass of  $i$  and  $\ell$  is evenly distributed to the two offspring of  $i$  and  $\ell$  at time  $t + 1$ . Thus  $D_t^{(i)}(j) + D_t^{(\ell)}(j) = D_{t+1}^{(i)}(j) + D_{t+1}^{(\ell)}(j)$ .

Since all distributions not involved in crossover do not change, the claim follows.  $\square$

So we can imagine  $D_t^{(i)}(j)$  as a doubly stochastic matrix. The aim is to show that for every  $i$ ,  $D_t^{(i)}$  converges to a uniform distribution on  $\{1, \dots, \mu\}$ . We will also show that this holds independently for all bits.

We collect some important properties of the behavior of the  $D_t^{(i)}$  over time:

**LEMMA 5.** *If  $i$  is not involved in crossover at time  $t$ , then  $D_{t+1}^{(i)} = D_t^{(i)}$ . Otherwise,  $D_{t+1}^{(i)}(j) = D_{t+1}^{(\ell)}(j) = (D_t^{(i)}(j) + D_t^{(\ell)}(j))/2$  for some  $\ell$  being uniform on  $\{1, \dots, \mu\} \setminus \{i\}$ .*

**PROOF.** Clearly if an individual  $i$  is not involved in crossover in generation  $t$ , its ancestral line does not change from  $t$  to  $t + 1$ . On the other hand, if it is selected for crossover, an individual  $\ell$  uniform on the remaining population is selected as a mate. The ancestral line at  $i$  in  $t + 1$  is either inherited from parent  $i$  or parent  $\ell$  with equal probability.  $\square$

We consider the potential function  $\Phi_t = \sum_{i=1}^{\mu} \sum_{j=1}^{\mu} |D_t^{(i)}(j) - 1/\mu|$  and show that it converges to 0. In particular,  $\Phi_t = 0$  exactly if  $\Pr(D_t^{(i)} = j)$  is uniform for all individuals  $i$  and ancestors  $j$ . We now show that, for the selection-free  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$  (i. e.,  $\epsilon = 0$ ), the drift of the potential function can be characterized exactly.

**LEMMA 6.**  $\mathbb{E}(\Phi_t - \Phi_{t+1} \mid \Phi_t) = \Phi_t / (\mu - 1)$ .

**PROOF.** For  $i \in \{1, 2, \dots, \mu\}$ , individual  $i$  is chosen as a parent for crossover in time  $t$  with probability  $1/\mu + (1 - 1/\mu)(1/(\mu - 1)) = 2/\mu$ . By Lemma 5,

$$\begin{aligned} \mathbb{E}\left(D_{t+1}^{(i)}(j) \mid D_t^{(i)}(j)\right) &= \left(1 - \frac{2}{\mu}\right) D_t^{(i)}(j) \\ &\quad + \frac{2}{\mu} \left(\frac{D_t^{(i)}(j)}{2} + \left(\sum_{\ell \neq i} \frac{D_t^{(\ell)}(j)}{2} \cdot \frac{1}{\mu - 1}\right)\right) \\ &= \left(1 - \frac{1}{\mu - 1}\right) \left(D_t^{(i)}(j) - \frac{1}{\mu}\right) + \frac{1}{\mu}, \end{aligned}$$

since, by Lemma 4,  $\sum_{\ell \neq i} D_t^{(\ell)}(j) = 1 - D_t^{(i)}(j)$ . Examining the above equation, we can see that  $D_t^{(i)}(j) > 1/\mu \implies \mathbb{E}\left(D_{t+1}^{(i)}(j)\right) > 1/\mu$ . Similarly,  $D_t^{(i)}(j) < 1/\mu \implies \mathbb{E}\left(D_{t+1}^{(i)}(j)\right) < 1/\mu$ . Thus, taking the absolute deviation from  $1/\mu$ , we have

$$\mathbb{E}\left(|D_{t+1}^{(i)}(j) - 1/\mu| \mid D_t^{(i)}(j)\right) = \left(1 - \frac{1}{\mu - 1}\right) |D_t^{(i)}(j) - 1/\mu|.$$

The claim follows from the definition of  $\Phi_t$ .  $\square$

We can now apply multiplicative drift to analyze the time until the potential has been reduced. Recall that we have  $n$  matrices  $D_t^{(i)}(j)$  altogether by looking at the  $n$  values for the bit  $b^*$  we fixed above and accordingly  $n$  different potentials  $\Phi_t$ .

**LEMMA 7.** *Let  $c, s \geq 0$ . Then after  $t^* := (\mu - 1)((c + 1) \ln(\mu) + \ln 2 + s)$  steps, with probability at least  $1 - ne^{-s}$  it holds that  $\Phi_{t^*} \leq 1/\mu^c$  with respect to the potential functions for all bits.*

PROOF. We apply the multiplicative drift theorem (Theorem 2) on the result from Lemma 6. We have  $\Phi_t \leq 2\mu$  since each row of the matrix  $D_t^{(i)}(j)$  has at most  $\mu$  entries below  $1/\mu$ , contributing at most 1 to  $\Phi_t$ , and each row sum is at most 1 by definition, again contributing at most 1 to  $\Phi_t$ . Hence, using  $\delta = 1/(\mu - 1)$ ,  $X^{(0)} = 2\mu$  and  $x_{\min} = 1/\mu^c$  in the multiplicative drift theorem, we obtain an expected time of at most

$$(\mu - 1)(\ln(2\mu/x_{\min}) + 1) = (\mu - 1)(c \ln(\mu) + \ln(\mu) + \ln 2 + 1)$$

to reduce the potential to at most  $x_{\min}$ . Moreover, by the tail bounds for multiplicative drift we obtain a time bound

$$t^* = (\mu - 1)(c + 1) \ln(\mu) + \ln 2 + s$$

that holds with probability at least  $1 - e^{-s}$ , i. e.,  $\Pr(\Phi_{t^*} > x_{\min}) \leq e^{-s}$ . The lemma now follows by a union bound over the  $n$  bits.  $\square$

Recall that the parent distribution  $D_t^{(i)}$  of individual  $i$  at time  $t$  was with respect to a specific bit  $b^*$ . We now consider different bits and make this reference explicit by writing  $D_{t,b^*}^{(i)}$  where appropriate.

LEMMA 8. *Let  $i \in \{1, \dots, \mu\}$  be an arbitrary individual. For any two bits  $b^*, b^{**} \in \{1, \dots, n\}$ , the distributions  $D_{t,b^*}^{(i)}$  and  $D_{t,b^{**}}^{(i)}$  are pairwise independent.*

PROOF. The claim holds at time 0 since the corresponding distributions are atomic. When creating the population at time  $t + 1$  and individuals  $i, j$  is chosen as parent, exactly the distributions  $D_{t+1}^{(i)}$  and  $D_{t+1}^{(j)}$  are changed with respect to time  $t$ . Since for the two bits  $b^*$  and  $b^{**}$  the decision to inherit the bit from  $i$  or  $j$  is made independently, the independence of  $D_{t+1,b^*}^{(i)}$  and  $D_{t+1,b^{**}}^{(i)}$  (analogously for  $j$ ) follows.  $\square$

We now put everything together to obtain our main theorem.

THEOREM 9. *Let  $p_1, \dots, p_n$  be the frequencies of one-bits at the  $n$  bits of the initial population of the StSt  $\binom{\mu}{2}$  GA $_\epsilon$  with  $\epsilon = 0$ , and consider any point of time  $t \geq (\mu - 1)(2 \ln(\mu) + \ln(n) + \ln 2 + s)$  for some  $s \geq \ln n$ . Then for any individual  $x$  in the population at time  $t$  and any  $y \in \{0, 1\}^n$  it holds that*

$$\Pr(x = y) \geq \left(1 - \frac{1}{n}\right)^n \left(\prod_{i|y_i=0} (1 - p_i)\right) \left(\prod_{i|y_i=1} p_i\right) - ne^{-s}.$$

and

$$\Pr(x = y) \leq \left(1 + \frac{1}{n}\right)^n \left(\prod_{i|y_i=0} (1 - p_i)\right) \left(\prod_{i|y_i=1} p_i\right) + ne^{-s}.$$

PROOF. We apply Lemma 7 setting  $c = 1 + \ln(n)/\ln(\mu)$  and assume that the error of probability at most  $ne^{-s}$ , which appears as an error term in the bounds (appropriately added or subtracted), does not occur. According to Lemma 7, after  $t$  steps we have  $\Phi_t \leq \mu^{-c} = 1/(\mu n)$ . By definition of the potential function, it follows that  $\min_{i,j} \{D_t^{(i)}(j)\} \geq \frac{1}{\mu} - \frac{1}{\mu n}$ . Let  $k \in \{1, \dots, n\}$  be an arbitrary bit position, and let  $S \subseteq \{1, \dots, \mu\}$  be the set of individuals with a 1-bit

in position  $k$  in the initial population. The probability that the bit in position  $k$  equals 1 at time  $t$  is thus

$$\sum_{i=1}^{\mu} \sum_{j \in S} D_t^{(i)}(j) \geq \mu p_k \left(\frac{1}{\mu} - \frac{1}{\mu n}\right) \geq p_k \left(1 - \frac{1}{n}\right).$$

Symmetrically, the probability is at most  $p_k(1 + 1/n)$ . Analogously, the probability that bit in position  $k$  is set to 0 lies in the interval  $[(1 - p_k)(1 - 1/n), (1 - p_k)(1 + 1/n)]$ . Multiplying this for all bit positions and using the independence of the bitwise distributions (Lemma 8), the two statements follow.  $\square$

### 3.1 Use of Decoupling in Lower Bounds on Runtime

After the decoupling time, we know the sampling distribution of the selection-free StSt  $\binom{\mu}{2}$  GA $_\epsilon$  quite precisely. If the initial frequencies correspond to a sampling distribution with a high variance, e. g.,  $p_i \in [c, 1 - c]$  for a positive constant  $c \leq 1/2$  for all  $i$ , then sampling any specific string after the decoupling time is very unlikely. Thus, if the optimum is not found before the decoupling time, then it has a very low chance of being found in a reasonable amount of time.

In the following theorem, we show a proof of concept for this approach. We initialize the StSt  $\binom{\mu}{2}$  GA $_\epsilon$  deterministically with a population of extremely high diversity: there are  $n/2$  pairs of individuals that have the maximal possible Hamming distance  $n$ , which is the best case for crossover to make progress towards a target (here the all-ones string). Despite this beneficial initialization, decoupling occurs too quickly for the algorithm to benefit from it.

THEOREM 10. *Assume  $\mu = O(n^{1/2-\delta})$  for some  $0 < \delta < 1/2$  and consider the following deterministic initialization:  $\mu/2$  individuals are the bit string  $(10)^{n/2}$  and the other  $\mu/2$  are  $(01)^{n/2}$ . Then the expected optimization time of the StSt  $\binom{\mu}{2}$  GA $_\epsilon$  with  $\epsilon = 0$  to create  $x^* := 1^n$  is at least  $\Omega(e^{n^{\delta/2}})$  for any positive constant  $k > 0$ .*

PROOF. Let  $t := (\mu - 1)(2 \ln \mu + 2 \ln n + \ln 2 + n^{\delta/2})$  by choosing  $s := n^{\delta/2} + \ln n$  in the bound from Theorem 9. By the theorem, for any search point  $y$  the probability of sampling it after time  $t$  is at most

$$(1 + 1/n)^n 2^{-n} + ne^{-s} = O(e^{-n^{\delta/2}}).$$

Hence, if  $x^*$  is not produced before time  $t$  then the expected number of additional steps necessary to produce it is at least  $\Omega(e^{n^{\delta/2}})$ .

By Lemma 3, a single crossover operation does not add more than  $n^{1/2+\delta/3}$  one-bits to the best parent with probability  $1 - 2^{-\Omega(n^{2\delta/3})}$ . We pessimistically assume that the maximum number of ones in the population is always extended by  $n^{1/2+\delta/3}$  in every generation, and note that taking a union bound over the first  $t$  generations ensures this holds with probability superpolynomially close to one. Under these conditions, after  $t = O(n^{1/2-\delta/2})$  generations the overall progress in the number of ones is at most  $o(n)$  and so  $x^*$  is not hit by any individual before the decoupling time. The law of total probability bounds the expected optimization time from below.  $\square$

## 4 UPPER BOUNDS FOR (WEAK) SELECTION

In this section, we consider the case  $\epsilon > 0$  and analyze the runtime behavior of the StSt  $\binom{\mu}{2}$  GA $_\epsilon$  on ONEMAX. The case of  $\epsilon = 1$  was considered before by Witt [32], who showed an upper runtime

bound  $O(\mu\sqrt{n}\log n)$  that holds w. h. p. if  $\mu \geq c\sqrt{n}\ln n$  for a sufficiently large constant  $c > 0$ . The bound is interesting since it is, up to logarithmic factors, identical to a runtime bound for the estimation of distribution algorithm cGA [28] that can be modeled as a GA using steady-state binary tournament selection and gene-pool crossover. We generalize the result from [32] to arbitrary selection probabilities here, before we prove almost tight lower bounds in Section 5.

**THEOREM 11.** *Let  $\mu \geq c(1/\epsilon)\sqrt{n}\ln n$  for some sufficiently large constant  $c > 0$  and  $\mu = n^{O(1)}$ . Then the runtime of the StSt  $\binom{\mu}{2}$  GA $_{\epsilon}$  on ONEMAX is  $O((\mu/\epsilon)\sqrt{n}\log n)$  w. h. p..*

To prove this theorem, we take an approach that is inspired by the analysis in [32]. First of all, we bound the probability of a frequency reaching the irrecoverable state 0 due to genetic drift during the suggested time. Afterwards a drift analysis of a potential function based on the sum of frequencies is conducted.

We start with the following helper lemma that bounds the effect of genetic drift on single frequencies. It can be proven analogously and in large tracks identically to Lemma 3.7 in [32].

**LEMMA 12.** *Consider the StSt  $\binom{\mu}{2}$  GA $_{\epsilon}$  on ONEMAX. Choose an arbitrary but fixed bit and let  $X_t$ ,  $t \geq 0$ , denote its frequency value at iteration  $t$ . Then  $Y_t = \mu X_t$  stochastically dominates a walk on  $\{0, 1, \dots, \mu-1, \mu\}$  with transition probabilities  $\Pr(Y_{t+1} = i+1 \mid Y_t = i) = (i/\mu)(\mu-i)/(\mu-1)$ ,  $\Pr(Y_{t+1} = i-1 \mid Y_t = i) = (i/\mu)(\mu-i)/(\mu-1)$  for  $i \in \{1, \dots, \mu-1\}$  as well as  $\Pr(Y_{t+1} = i \mid Y_t = i) = 1 - 2(i/\mu)(\mu-i)/(\mu-1)$  for  $i \in \{0, \dots, \mu\}$ . The random walks w. r. t. different bits are not necessarily independent.*

*The probability that the  $X_t$ -value becomes 1/4 or less within  $s$  steps is at most  $e^{-\mu^2/(48s)} + e^{-\mu/32}$ .*

**PROOF.** We denote the index of the considered bit by  $k$ . If there are  $i$  individuals having a one at position  $k$ , then uniform selection will select an individual with a one at bit  $k$  with probability  $i/\mu$  as the first parent. Since the two individuals used for tournament selection (before applying crossover) are chosen independently without replacement, the probability that bit  $k$  coincides in the tournament equals  $1 - 2(i/\mu)(\mu-i)/(\mu-1)$ . Since crossover will not have any effect in this case, this already proves the claimed probability of the event  $Y_{t+1} = Y_t$ .

The event that the first chosen individual has a one at bit  $k$  and the second individual a zero has probability  $p^* := (i/\mu)(\mu-i)/(\mu-1)$ , as does the opposite event of a zero in the first and a one in the second individual. Note that with probability 1/2 the entries at position  $k$  are swapped by crossover. Still, after crossover the two offspring satisfy these two cases each with the same probability  $p^*$ . The crucial observation is that due to crossover, the value of the two selected individuals are swapped with probability 1/2, and swapping occurs independently of all other bit values.

Hence, if the ranking of the two offspring does not depend on bit  $k$  then the selected offspring will have a one with probability exactly  $p^*$ . If it depends on bit  $k$  then the offspring with a one at bit  $k$  has a higher probability of being selected due to the structure of ONEMAX, so  $\Pr(Y_{t+1} = i+1 \mid Y_t = i) \geq p^*$  in any case. This proves the first statement of the lemma.

For the second statement, we note that with probability at least  $1 - e^{-\mu/32}$  according to Chernoff bounds, the frequency of bit  $k$  is

at least 3/8, which we assume to happen. Considering the random walk analyzed above, the expected number of decreasing steps within  $s$  steps is at most  $s/2$ . To reach 1/4 from frequency at least 3/8, it is necessary to have at least  $\mu/8$  more decreasing than increasing steps within  $s$  steps. By the Hoeffding bound with random variables  $Y_i \in \{-1, 1\}$ , the probability of this is at most  $e^{-2(\mu/8)^2/(2s)} = e^{-\mu^2/(48s)}$ . The lemma follows by adding up the failure probabilities.  $\square$

Having established this tool, we can prove our theorem.

**PROOF OF THEOREM 11.** We consider a phase consisting of  $t := c'(1/\epsilon)\mu\sqrt{n}\ln n$  iterations for a sufficiently large constant  $c' > 0$ . Applying the second statement of Lemma 12, we obtain that no frequency drops below 1/4 within  $t$  steps with high probability if  $c$  is chosen as a sufficiently large constant.

The remainder of the proof carries out a drift analysis on the potential  $\Phi_t = \sum_{i=1}^n (1 - p_{i,t})$ , which was already frequently used in the literature to analyze cGA and related algorithms. The overall approach is inspired by the proof of Theorem 3.9 in [32].

We will show a drift  $E(\Phi_t - \Phi_{t+1} \mid \Phi_t) = \Omega((1/\epsilon)V_t/(\mu\sqrt{n}))$ , where  $V_t := \sum_{i=1}^n p_i(1 - p_{i,t})$ . Note that the latter quantity would be the sampling variance if bit  $i$  was sampled independently with probability  $p_i$ , however, bit values are not sampled independently.

An iteration of StSt  $\binom{\mu}{2}$  GA $_{\epsilon}$  draws two individuals  $x$  and  $y$  uniformly at random without replacement, where w. l. o. g.  $|x| \leq |y|$ . Each of these has an expected ONEMAX-value of  $\sum_{i=1}^n p_{i,t}$  using linearity of expectation. Also by linearity of expectation, the expected value of the Hamming distance  $D$  of  $x$  and  $y$  is

$$E(D) := \sum_{i=1}^n 2p_i(1 - p_i) \frac{\mu}{\mu-1} \geq \sum_{i=1}^n 2p_i(1 - p_i),$$

which we will use below. The number of common one-bits of  $x$  and  $y$  is denoted by  $a$  hereinafter, which means that  $x$  and  $y$  together have a total number of  $2a + D$  one-bits and average of  $a + D/2$  one-bits.

Crossing  $x$  and  $y$  over creates the two offspring  $x'$  and  $y'$ . The bits where  $x$  and  $y$  coincide also coincide in the offspring, so selection will not change their frequencies. The bits where they differ contain together  $D$  ones and contribute  $D/\mu$  to the  $\Phi_t$ -value. As in Lemma 3, we consider an experiment where  $D$  independent trials with success probability 1/2 are executed on the bits different in  $x$  and  $y$ , identifying a success with crossover putting the one-bit from the considered position into offspring  $x'$ . By the properties of the binomial distribution, the number of successes is  $D/2 + \Omega(\sqrt{D})$  with constant probability. Hence, the probability that  $x'$  has  $\Omega(\sqrt{D})$  more ones than the average  $a + D/2$  is  $\Omega(1)$ . If this happens,  $x'$  is the fitter individual.

To determine the drift of  $\Phi_t$ , we are interested in the expected value of  $\sqrt{D}$ , knowing  $E(D)$ . To this end, we use the converse of Jensen's inequality stated in Lemma 1 with respect to a probability distribution  $(p_1, \dots, p_n)$  on the outcomes of a discrete random variable  $X$  with support  $[0, M]$ . Hence, for convex  $f$  we have

$$E(f(X)) \leq \frac{M - E(X)}{M} f(0) + E(X) \frac{f(M)}{M}.$$

Using this with  $f(x) = -\sqrt{x}$  and  $M = n$ , we obtain  $E(\sqrt{D}) \geq \frac{E(D)}{\sqrt{n}}$ . Altogether, we obtain an expected surplus of one-bits in the fitter

individual over the average  $a + D/2$  that is bounded from below by

$$\Omega\left(\mathbb{E}(\sqrt{D})\right) = \Omega\left(\sum_{i=1}^n 2p_i(1-p_i)/\sqrt{n}\right) = \Omega\left(\sum_{i=1}^n (2/3)(1-p_i)/\sqrt{n}\right), \tag{1}$$

using that  $p_{i,t} \geq 1/3$ .

We now investigate the expected number of one-bits at the considered  $D$  positions in the offspring of crossover. With probability  $1 - \epsilon$ , fitness of  $x'$  and  $y'$  is disregarded to that the expected total number of one-bits in the offspring  $x'$  and  $y'$  equals the previous total number of one-bits in the parents. Hence, if fitness is disregarded, the expected change of  $\Phi$ -value is 0.

Let us now consider the case that fitness is regarded and the better offspring replaces the parents. Every one-bit in the better offspring contributes  $2/\mu$  to  $\Phi_{t+1}$  since it will be copied into two slots. Hence, replacing both  $x$  and  $y$  by the better offspring, which has an expected number of  $a + D/2 + \Omega(\mathbb{E}(\sqrt{D}))$  one-bits, decreases the  $\Phi_t$ -value in expectation by  $\Omega(\mathbb{E}(\sqrt{D}))/\mu$ . Using (1), the expected decrease is at least

$$\frac{c_2}{\mu\sqrt{n}} \sum_{i=1}^n (1-p_i) = \frac{c_2}{\mu\sqrt{n}} \Phi_t$$

for some constant  $c_2 > 0$ . Using the law of total probability, we obtain

$$\mathbb{E}(\Phi_t - \Phi_{t+1} \mid \Phi_t) = (1 - \epsilon) \cdot 0 + \epsilon(c_2/(\mu\sqrt{n}))\Phi_t = \epsilon(c_2/(\mu\sqrt{n}))\Phi_t$$

Using the multiplicative drift theorem (Theorem 2), the expected number of iterations until the potential is reduced to at most  $x_{\min} := (n - 1)/\mu$  is at most

$$\frac{\ln(x_0/x_{\min}) + 1}{\epsilon c_2/(\mu\sqrt{n})} = O((1/\epsilon)\mu\sqrt{n} \log \mu) = O((1/\epsilon)\mu\sqrt{n} \log n),$$

and, using tail bounds for multiplicative drift,  $O((1/\epsilon)\mu\sqrt{n} \log n)$  holds with high probability. The proof is completed by observing that when the potential is at most  $(n - 1)/\mu$ , there must be at most  $n - 1$  zeros distributed throughout the entire population. The presence of the all-ones string in the population now follows by application of the pigeonhole principle.  $\square$

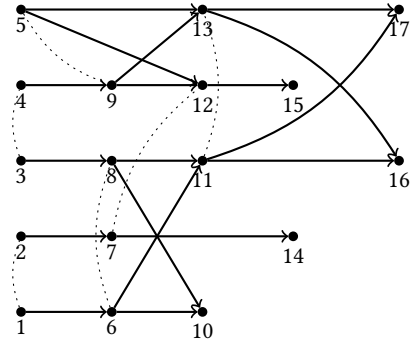
## 5 LOWER BOUNDS VIA FAMILY GRAPHS

In [32], only upper bounds on the runtime of the  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$  with  $\epsilon = 1$  were obtained, and it was stated as an open problem to prove lower bounds. We address this challenge by a novel proof technique called the family graphs.

As mentioned above, it is natural to represent the history of individuals in population-based algorithms using tree or more general graph structures. The  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$  creates an ancestor structure that shares similarities with the so-called family trees of steady-state EAs like the  $(\mu+1)$  EA [31]. Crossover can easily create cycles in this graph structure (if edge directions are ignored) such that we from now on speak of family graphs. Formally, a family graph of the  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$  is a stochastic process evolving an inductively growing, directed, acyclic, labeled graph structure with nodes representing individuals created during the run of the algorithm along with their time of creation:

At time 0, i. e., directly after initialization, the graph consists of  $\mu$  isolated vertices representing the  $\mu$  members of the initial population, labeled with the individuals from the initial population along with their indices:  $(1, x^{(1)}), \dots, (\mu, x^{(\mu)})$ . The graph at time  $t$  is generated from the one at time  $t-1$  by adding two vertices corresponding to the children produced by crossover at time  $t-1$ , more precisely, labeled with the two pairs  $(\mu+2t-1, x^{(\mu+2t-1)})$  and  $(\mu+2t, x^{(\mu+2t)})$  according to the children  $x^{(\mu+2t-1)}, x^{(\mu+2t)}$  generated in the last crossover operation (not indicating which child will possibly be replaced by its sibling). These new children vertices are connected to the two vertices  $(a, x^{(a)}), (b, x^{(b)})$  representing their parents in the crossover operation via directed edges pointing towards the children; more precisely the two edges  $((a, x^{(a)}), (\mu+2t-1, x^{(\mu+2t-1)}))$ , and  $((b, x^{(b)}), (\mu+2t, x^{(\mu+2t)}))$  are inserted.

Finally, if the generation is selective (replaces the less fit children with its sibling) another two edges between parents and children are introduced, more precisely the edges  $((a, x^{(a)}), (\mu+2t, x^{(\mu+2t)}))$  and  $((b, x^{(b)}), (\mu+2t-1, x^{(\mu+2t-1)}))$  to indicate that the offspring may be swapped in the operation. Hence, the gadget connecting parents and offspring of crossover operations will contain two directed edges in non-selective steps and four in selective. Note also that the two parents  $x^{(a)}$  and  $x^{(b)}$  will no longer be present in the population at time  $t$ ; however, the graph structure is only growing with respect to the subset inclusion. See Figure 1 for an illustration.



**Figure 1:** Example of a family graph of the  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$  with  $\mu = 5$  after 6 generations. We omit the individuals  $x^{(t)}$  labeling the nodes for clarity. Dotted lines are not part of the graph but illustrate parents of crossover operations. There are three selective crossover operations.

We will exploit that our graph structure (more precisely, the nodes and edges with only time stamps but no individuals as labels) neither depends on the underlying fitness function nor the actual individuals created during a run of the  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$ . This is made formal in the following lemma.

**LEMMA 13.** *For  $t \geq 0$ , the probability distribution on family graphs induced by running the  $\text{StSt} \binom{\mu}{2} \text{GA}_\epsilon$  for  $t$  generations is independent of the search points  $x^{(1)}, \dots, x^{(\mu+2t)}$  labeling the graph.*

**PROOF.** The selection step choosing two individuals for crossover is neither dependent on fitness nor on the bit strings forming the

parents. The generation gives rise to two new vertices that are both connected to both parents via two or four directed edges. The structure of these edges is only dependent on the event that the crossover is selective, which is independent of fitness and individuals.  $\square$

Inspired by the proof bounding the depth of family trees of the  $(\mu+1)$  EA [31], we now bound the diameter of the family graph of the  $\text{StSt} \binom{\mu}{2}$   $\text{GA}_\epsilon$  at time  $t$ . This diameter corresponds to the maximum number of crossover operations that may lead from an initial population via a sequence of crossover operations involving this individual, its offspring, and possibly other individuals as parents, to an individual from the population at time  $t$ .

**LEMMA 14.** *Let  $D(t)$  denote the diameter of the family graph of the  $\text{StSt} \binom{\mu}{2}$   $\text{GA}_\epsilon$  at time  $t$ . Then for all  $t \geq 0$ ,  $\Pr(D(t) \geq 12t/\mu) \leq \mu 2^{-\Omega(t/\mu)}$ .*

**PROOF.** Let  $\ell := \lceil 12t/\mu \rceil$ . We look into the probability that a specific path indexed by numbers  $i_0 < i_1 < \dots < i_\ell$  is created, where  $i_0 \leq \mu$  and  $i_1 > \mu$ . The probability that such a path is created is bounded from above by  $(2/\mu)^\ell$  since the probability of choosing the vertex indexed  $i_{t-1}$  during generation  $\lceil (i_t - \mu)/2 \rceil$  (i. e., during the generation creating the  $i_t$ -indexed individual) is at most  $2/\mu$ . More precisely, the probability is 0 if the individual is no longer present in the population and it is  $1/\mu + (1 - 1/\mu)(1/(\mu - 1)) = 2/\mu$  otherwise since binary tournament selection chooses two individuals uniformly without replacement.

Fixing  $i_0$ , the number of different paths  $i_1, \dots, i_\ell$  is bounded from above by  $\binom{2t}{\ell}$ . Hence, the probability that the path  $i_0, \dots, i_\ell$  emerges is at most

$$\binom{2t}{\ell} \left(\frac{2}{\mu}\right)^\ell \leq \binom{\ell\mu/6}{\ell} \left(\frac{2}{\mu}\right)^\ell \leq \left(\frac{e\ell\mu}{6\ell}\right)^\ell \left(\frac{2}{\mu}\right)^\ell = \left(\frac{e}{3}\right)^\ell = 2^{-\Omega(t/\mu)}$$

The lemma now follows via a union bound over all  $\mu$  initial individuals.  $\square$

The bound from Lemma 14 will be a crucial tool in lower bounding the optimization time of the  $\text{StSt} \binom{\mu}{2}$   $\text{GA}_\epsilon$ . The following general theorem covers any function with a unique optimum. Compared to Theorem 11, there is a gap between upper and lower bounds of order  $(\log^3 n)/\epsilon$ , i. e., only of polylogarithmic size in the case  $\epsilon = \Theta(1)$ .

**THEOREM 15.** *Let  $f$  be a function with a unique optimum and let  $\mu = n^{O(1)}$ . Then with probability  $1 - n^{-\Omega(\log n)}$ , the optimization time of the  $\text{StSt} \binom{\mu}{2}$   $\text{GA}$  with arbitrary  $\epsilon$  is bounded from below by  $\Omega(\mu\sqrt{n}/\log n)$ .*

**PROOF.** Without loss of generality, since the algorithm is unbiased, we assume the all-ones string to be the optimum. Moreover, with overwhelming probability  $1 - \mu 2^{-\Omega(n)} = 1 - 2^{-\Omega(n)}$ , each of the  $\mu$  initial individuals has Hamming distance more than  $n/4$  from the optimum, which we now assume to happen. Otherwise, the optimization time is estimated from below by 0.

We consider a phase of  $t := \mu\sqrt{n}/(48 \ln n)$  generations. By Lemma 14, the family graph has a diameter bounded by  $\ell := \sqrt{n}/(4 \ln n)$  with a probability of at least  $1 - 2^{-\Omega(\sqrt{n}/\log n)}$ . We assume this to happen as well.

Next we investigate the family graph in levels. Let level  $i$  comprise all vertices  $v$  such that the *longest* path from an initial individual to  $v$  is  $i$ . Note that the number of levels is bounded from above by  $\sqrt{n}/(4 \ln n)$  according to our assumption. The crucial claim is that the maximum number of one-bits at level  $i$  is by at most  $\sqrt{n} \ln n$  bigger than the maximum number of one-bits at level  $i - 1$ , with probability  $1 - n^{-\Omega(\log n)}$ .

To prove the claim, we consider a vertex  $v$  from level  $i$  resulting from a selective crossover and its two parents  $a$  and  $b$ . These parents are both on level at most  $i - 1$ . We use Lemma 3) and observe for  $v$  that the crossover of  $a$  and  $b$  leading into level  $i$  increases the number of one-bits by at most  $\sqrt{n} \ln n$  with probability  $1 - n^{-\Omega(n)}$ . This applies regardless of the crossover being selective or not. Hence, using a union bound over all  $\mu = n^{O(1)}$  vertices on level  $i$ , the number of one-bits increases by at most  $(\ln n)\sqrt{n}$  compared to the previous level with a failure probability at most  $n^{O(1)-\Omega(\log n)} = n^{-\Omega(\log n)}$ .

The theorem now follows by realizing that the accumulated increase of number of one-bits over all at most  $\sqrt{n}/(4 \ln n)$  levels is at most

$$\frac{\sqrt{n}}{4 \ln n} (\ln n) \sqrt{n} \leq \frac{n}{4}$$

with a probability of at least  $1 - n^{-\Omega(\log n)} - 2^{-\Omega(\sqrt{n}/\log n)} - 2^{-\Omega(n)} = 1 - n^{-\Omega(\log n)}$ .  $\square$

The argumentation in the proof of Theorem 15 takes a somewhat different approach compared to classical analyses of the  $(\mu+1)$  EA using family trees. Let us take the lower bound for the  $(\mu+1)$  EA on  $\text{ONEMAX}$  as an example. Our analysis of  $\text{StSt} \binom{\mu}{2}$   $\text{GA}_\epsilon$  has in common with this analysis that the event of family trees (graphs) being very deep is proved unlikely. Let us assume the depth (diameter) of the graph is bounded by  $D$ . Afterwards, the classical and our approach are different. The analysis of  $(\mu+1)$  EA considers a phase of  $t := c\mu n$  steps for some constant  $c > 0$  and argues that every possible path created at time points  $t_1, \dots, t_D$ , ignoring whether selected or not by the algorithm, is unlikely to create the optimum of  $\text{ONEMAX}$ , more precisely has probability  $2^{-\Omega(n)}$ . Taking a union bound over all  $\binom{t}{D}$ , i. e., exponentially many choices of these points of time, the probability of observing a path of length at most  $D$  leading to the optimum is still obtained as exponentially small if all constants are chosen carefully. Hence, at the heart of the classical analysis there is a selection-less process.

In contrast to this, our analysis in Theorem 15 does not consider all paths that can emerge in a phase of asymptotically  $\mu\sqrt{n}/(\log n)$  steps. Instead, we consider the actually emerging paths of which there are only polynomially many. Due to selection, some of these paths will make a larger than expected progress. We bound this progress level-wise by considering the largest order statistic of only polynomially many random variables following the distribution  $|\text{Bin}(n, 1/2) - n/2|$ , each of which stochastically dominates the progress of a single crossover operation over its best parent.

## CONCLUSIONS

We have introduced two novel techniques for proving lower bounds on the runtime of crossover-based algorithms, addressing a lack of such techniques in the literature. The first technique, decoupling,



shows bounds on the time for a steady-state GA to converge to linkage equilibrium in a selection-free setting; combining this with an analysis of the possible progress before this so-called decoupling time, we have obtained lower bounds on the expected runtime. We then turned to a setting with weak selection and generalized the family tree technique, originally proposed for the  $(\mu+1)$  EA without crossover, to our algorithm. We have obtained a general lower bound for functions with a unique optimum depending on the number of bits  $n$ , the population size  $\mu$  and the selection probability  $\epsilon$ . Moreover, to assess the quality of the lower bounds, upper bounds on the ONEMAX function have been shown. These bounds are almost tight for  $\epsilon = \Theta(1)$ . It is an open problem to close the gaps between upper and lower bounds.

A special case of our algorithm of study StSt  $\binom{\mu}{2}$  GA $_{\epsilon}$ , using  $\epsilon = 1$  and simply called StSt  $\binom{\mu}{2}$  GA, was introduced in [32] and motivated as a simplified model of the estimation-of-distribution algorithm cGA, with the implicit gene-pool crossover of cGA replaced by uniform crossover. Our analysis reveals on ONEMAX that this StSt  $\binom{\mu}{2}$  GA has the same runtime behavior  $\tilde{\Theta}(\sqrt{n}\mu)$  (i. e.,  $\Theta(\sqrt{n}\mu)$  up to polylogarithmic factors) as the cGA, which illustrates that the combination of binary tournament and crossover are indeed the crucial parameters responsible for this runtime behavior. We consider this result as a step towards a unified framework for the analysis of EDAs and classical crossover-based GAs.

We are optimistic that our techniques will be useful in analyzing more advanced GAs from a lower-bound perspective, including settings where both mutation and crossover are used.

## REFERENCES

- [1] Denis Antipov, Benjamin Doerr, Jiefeng Fang, and Tangi Hetet. 2018. Runtime analysis for the  $(\mu+\lambda)$  EA optimizing OneMax. In *Proc. of GECCO '18*. ACM Press, 1459–1466.
- [2] M. Klaričić Bakula, J. Pečarić, and J. Perić. 2012. On the converse Jensen inequality. *Applied Mathematics and Computation* 218 (2012), 6566–6575.
- [3] Duc-Cuong Dang, Tobias Friedrich, Timo Kötzing, Martin S. Krejca, Per Kristian Lehre, Pietro S. Oliveto, Dirk Sudholt, and Andrew M. Sutton. 2018. Escaping Local Optima Using Crossover With Emergent Diversity. *IEEE Transactions on Evolutionary Computation* 22 (2018), 484–497.
- [4] Benjamin Doerr and Leslie Ann Goldberg. 2013. Adaptive drift analysis. *Algorithmica* 65 (2013), 224–250.
- [5] Benjamin Doerr, Edda Happ, and Christian Klein. 2012. Crossover can provably be useful in evolutionary computation. *Theoretical Computer Science* 425 (2012), 17–33.
- [6] Benjamin Doerr, Daniel Johannsen, and Timo Kötzing. 2010. More Effective Crossover Operators for the All-Pairs Shortest Path Problem. In *Proc. of PPSN '10*. Springer, 184–193.
- [7] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. 2012. Multiplicative Drift Analysis. *Algorithmica* 64 (2012), 673–697.
- [8] Benjamin Doerr and Madeleine Theile. 2009. Improved analysis methods for crossover-based algorithms. In *Proc. of GECCO '09*. ACM Press, 247–254.
- [9] Hilda Geiringer. 1944. On the probability theory of linkage in Mendelian heredity. *The Annals of Mathematical Statistics* 15 (1944), 25–57.
- [10] Jens Jägersküpfer and Carsten Witt. 2005. Rigorous runtime analysis of a  $(\mu+1)$  ES for the Sphere function. In *Proc. of GECCO '05*. ACM Press, 849–856.
- [11] Thomas Jansen and Ingo Wegener. 2005. Real royal road functions – where crossover provably is essential. *Discrete Applied Mathematics* 149 (2005), 111–125.
- [12] Thomas Jansen and Christine Zarges. 2011. On benefits and drawbacks of aging strategies for randomized search heuristics. *Theoretical Computer Science* 412 (2011), 543–559.
- [13] Timo Kötzing, Dirk Sudholt, and Madeleine Theile. 2011. How Crossover Helps in Pseudo-Boolean Optimization. In *Proc. of GECCO '11*. ACM Press, 989–996.
- [14] Per Kristian Lehre. 2010. Negative drift in populations. In *Proc. of PPSN '10*. Springer, 244–253.
- [15] Per Kristian Lehre and Xin Yao. 2011. Crossover can be constructive when computing unique input-output sequences. *Soft Computing* 15 (2011), 1675–1687.
- [16] Per Kristian Lehre and Xin Yao. 2012. On the Impact of Mutation-Selection Balance on the Runtime of Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation* 16 (2012), 225–241.
- [17] Johannes Lengler. 2018. A General Dichotomy of Evolutionary Algorithms on Monotone Functions. In *Proc. of PPSN '18*. Springer, 3–15. Extended version: <http://arxiv.org/abs/1803.09227>.
- [18] Frank Neumann, Pietro S. Oliveto, Günter Rudolph, and Dirk Sudholt. 2011. On the Effectiveness of Crossover for Migration in Parallel Evolutionary Algorithms. In *Proc. of GECCO '11*. ACM Press, 1587–1594.
- [19] Pietro S. Oliveto, Jun He, and Xin Yao. 2008. Analysis of Population-Based Evolutionary Algorithms for the Vertex Cover Problem. In *Proc. of CEC '08*. IEEE Press, 1563–1570.
- [20] Pietro S. Oliveto and Carsten Witt. 2015. Improved time complexity analysis of the Simple Genetic Algorithm. *Theoretical Computer Science* 605 (2015), 21–41.
- [21] Yann Ollivier. 2003. Rate of convergence of crossover operators. *Random Struct. Algorithms* 23 (2003), 58–72.
- [22] Chao Qian, Yu Yang, and Zhi-Hua Zhou. 2013. An analysis on recombination in multi-objective evolutionary optimization. *Artificial Intelligence* 204 (2013), 99–119.
- [23] Yuval Rabani, Yuri Rabinovich, and Alistair Sinclair. 1998. A computational view of population genetics. *Random Structures and Algorithms* 12 (1998), 313–334.
- [24] Yuri Rabinovich, Alistair Sinclair, and Avi Wigderson. 1992. Quadratic Dynamical Systems (Preliminary Version). In *Proc. of FOCS '92*. IEEE Computer Society, 304–313.
- [25] Dirk Sudholt. 2005. Crossover is Provably Essential for the Ising Model on Trees. In *Proc. of GECCO '05*. ACM Press, 1161–1167.
- [26] Dirk Sudholt. 2009. The impact of parametrization in memetic evolutionary algorithms. *Theoretical Computer Science* 410 (2009), 2511–2528.
- [27] Dirk Sudholt. 2017. How Crossover Speeds up Building Block Assembly in Genetic Algorithms. *Evolutionary Computation* 25 (2017), 237–274.
- [28] Dirk Sudholt and Carsten Witt. 2016. Update Strength in EDAs and ACO: How to Avoid Genetic Drift. In *Proc. of GECCO '16*. ACM Press, 61–68.
- [29] Richard A. Watson and Thomas Jansen. 2007. A Building-Block Royal Road Where Crossover Is Provably Essential. In *Proc. of GECCO '07*. ACM Press, 1452–1459.
- [30] Carsten Witt. 2003. Population size vs. runtime of a simple EA. In *Proc. of CEC '03*. IEEE Press, 1996–2003.
- [31] Carsten Witt. 2006. Runtime Analysis of the  $(\mu+1)$  EA on Simple Pseudo-Boolean Functions. *Evolutionary Computation* 14 (2006), 65–86.
- [32] Carsten Witt. 2018. Domino convergence: why one should hill-climb on linear functions. In *Proc. of GECCO '18*. ACM Press, 1539–1546.