



This is a repository copy of *Tight bounds on the expected runtime of a standard steady state genetic algorithm*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/182523/>

Version: Published Version

Article:

Oliveto, P.S., Sudholt, D. and Witt, C. (2022) Tight bounds on the expected runtime of a standard steady state genetic algorithm. *Algorithmica*, 84 (6). pp. 1603-1658. ISSN 0178-4617

<https://doi.org/10.1007/s00453-021-00893-w>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Tight Bounds on the Expected Runtime of a Standard Steady State Genetic Algorithm

Pietro S. Oliveto¹ · Dirk Sudholt^{1,2} · Carsten Witt³ 

Received: 16 November 2020 / Accepted: 8 November 2021
© The Author(s) 2021

Abstract

Recent progress in the runtime analysis of evolutionary algorithms (EAs) has allowed the derivation of upper bounds on the expected runtime of standard steady-state genetic algorithms (GAs). These upper bounds have shown speed-ups of the GAs using crossover and mutation over the same algorithms that only use mutation operators (i.e., steady-state EAs) both for standard unimodal (i.e., ONEMAX) and multimodal (i.e., JUMP) benchmark functions. The bounds suggest that populations are beneficial to the GA as well as higher mutation rates than the default $1/n$ rate. However, making rigorous claims was not possible because matching lower bounds were not available. Proving lower bounds on crossover-based EAs is a notoriously difficult task as it is hard to capture the progress that a diverse population can make. We use a potential function approach to prove a tight lower bound on the expected runtime of the $(2+1)$ GA for ONEMAX for all mutation rates c/n with $c < 1.422$. This provides the last piece of the puzzle that completes the proof that larger population sizes improve the performance of the standard steady-state GA for ONEMAX for various mutation rates, and it proves that the optimal mutation rate for the $(2+1)$ GA on ONEMAX is $(\sqrt{97} - 5)/(4n) \approx 1.2122/n$.

Keywords Evolutionary algorithms · Runtime analysis · Crossover · Lower bounds

✉ Pietro S. Oliveto
P.Oliveto@sheffield.ac.uk
Carsten Witt
cawi@dtu.dk

¹ University of Sheffield, Sheffield S1 4DP, UK

² University of Passau, Passau, Germany

³ DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark

1 Introduction

The runtime analysis of randomized search heuristics like evolutionary algorithms (EAs), simulated annealing, ant colony optimization and estimation-of-distribution algorithms is a young and active subfield in algorithm research that has produced remarkable results in the last 20 years [2,11,15,23]. Its main goal is to understand the working principles of the algorithms in different scenarios by deriving runtime bounds depending on problem characteristics, choice of algorithms and parameters. This line of research started with simple evolutionary algorithms using mutation only. Still today, the role of crossover, also called recombination, is less well understood than the one of mutation. In fact, explaining when recombination and mutation based genetic algorithms (GAs)¹ perform better than more traditional general purpose search heuristics that use mutation alone is regarded as one of the fundamental problems in evolution-inspired computation.

Traditionally proofs showing that crossover is a useful operator relied either on excessively low crossover rates [16,18] or on some diversity-enforcing mechanism to make recombination effective by increasing the probability that members of the population are different [8,10,20,22,27]. However, it was never shown whether this enforced diversity was necessary or whether it was an additional requirement for the proofs to hold. Recently some results have appeared proving the superiority of standard steady state GAs² over mutation-only algorithms, without the need of any additional diversity enforcing mechanisms. Dang et al. [7] proved that for sufficiently large population sizes, the $(\mu+1)$ GA is at least a linear factor faster than the best algorithm using only standard bit mutation for the JUMP benchmark function. Hence, they showed that crossover may help algorithms to escape more quickly from local optima. Sutton [28] even proved that for the NP-hard Closest String problem from computational biology, the $(\mu+1)$ GA with sufficiently large population size and restarts is a fixed parameter tractable (FPT) algorithm while if only standard bit mutation is used (i.e., $(\mu+1)$ EA) it is not.

Strikingly, recombination has also been proven to be useful on unimodal functions. Lengler [21] has shown that there exist monotone functions for which the $(\mu+1)$ EA with not too low standard bit mutation rate c/n (i.e., $c > 2.13$) requires exponential runtime with high probability while the $(\mu+1)$ GA with sufficiently large population sizes can solve them in $O(n \log n)$ expected runtime for arbitrary mutation rates i.e., $\Theta(1)/n$. Analyses have revealed that the $(\mu+1)$ GA is faster than the $(\mu+1)$ EA using any standard bit mutation rate and population size, even on unimodal functions where the latter is particularly efficient i.e., ONEMAX [5,6]. Furthermore, if the fitness of offspring that are identical to their parents is not unnecessarily re-evaluated, then the algorithm is faster than any unary unbiased black box algorithm for the problem [19], albeit slower than if the diversity is enforced [3,27]. To prove these results, precise

¹ The term “evolutionary algorithms” is commonly used as an umbrella term for evolutionary algorithms with and without crossover. We use the term “Genetic Algorithm (GA)” to emphasize the use of crossover in an evolutionary algorithm.

² Steady state GAs are those that replace at most a proper subset of the population in each generation (usually one new individual is created). Typically they use standard bit mutation which flips each bit independently with probability c/n .

analyses up to the leading constants are required since for ONEMAX the algorithms have the same asymptotic expected runtime $O(n \log n)$ for moderate population sizes.

An important insight from these analyses is that if diversity is enforced as in Sudholt's work [27], then inevitably there are no advantages of using population sizes greater than $\mu = 2$ for ONEMAX. On the other hand, the analysis of Corus and Oliveto [6] provides upper bounds that decrease with the population size (up to some sub-logarithmic limit). For large enough population sizes the best derived upper bound is roughly $1.64n \ln n$ while, for $\mu = 2$, Corus and Oliveto only provide a larger upper bound of $\frac{4e^c n \ln n}{c(c+4)} + O(n)$ [5]. Due to a mistake in one probability calculation this turns out to actually be $\frac{9e^c n \ln n}{c(2c+9)} + O(n)$.

Indeed, *all* the positive results summarised above regarding the plain $(\mu+1)$ GA required sufficiently large population sizes. While the comparative statements with the mutation-based algorithms were possible because of the availability of lower bounds on their expected runtime, rigorously showing whether the suggested population sizes are actually necessary requires lower bounds on the expected runtime of the $(\mu+1)$ GA. Proving lower bounds for GAs with crossover is a notoriously hard task. The only available analysis concerning a standard GA is the proof that the simple genetic algorithm (SGA [12]) cannot solve ONEMAX in polynomial time with overwhelming probability due to the ineffectiveness of the fitness proportional selection operator [25,26]. There have been recent attempts to generalize proof methods like the family tree technique to crossover-based algorithms; however, these only apply in a specific setting without mutation.

Providing lower bounds on the expected runtime of the $(\mu+1)$ GA for ONEMAX has turned out to be surprisingly difficult. Sudholt simplified the analysis by considering a "greedy" $(2+1)$ GA that always selects amongst the fittest individuals in the population and is sped-up by automatically achieving the best possible crossover operation between different parents [27]. A less greedy $(2+1)$ GA was considered by Corus and Oliveto where individuals are only immediately crossed over optimally if the Hamming distance between the parents is larger than 2 [5]. These simplified algorithms allow the analysis to ignore the improvements which may occur in standard GAs when one parent is crossed over with another one of different fitness. However, it was never proven that the algorithms are indeed faster than the standard $(2+1)$ GA, hence that the bounds are also valid for the latter algorithm. In this paper we provide a lower bound for the $(2+1)$ GA with no simplifications that matches its upper bound up to the leading constant, hence providing a rigorous proof that larger populations are beneficial to the GA for ONEMAX. The preciseness of the results also allows us to derive that the value of $c \in (0, 1.422]$ that yields the optimal mutation rate c/n is approximately $c = 1.21221445$.

A major difficulty in proving rigorous lower bounds for populations with crossover is to find a way to aggregate the state of the algorithm such that it accurately captures the current distance from the optimum, but also the potential improvements of the crossover operator. These advancements could be very big if the parents have a large Hamming distance, and our aim is to show that this rarely happens. We solve the aggregation problem for the $(2+1)$ GA by defining a potential function that captures the current fitness and opportunities for easy improvements through crossover. By

showing bounds on the expected increase in the potential, we are able to quantify how the distance to the optimum decreases in one generation. The challenge lies in proving this for every possible population, from those with identical individuals to those with a good amount of diversity. Once the potential is appropriately bounded, we can use standard drift analysis arguments to bound the expected time from below.

1.1 Main Contributions

The expected optimisation time of the (2+1) GA is bounded from above as follows.

Theorem 1 *The expected optimisation time of the (2+1) GA with mutation rate c/n , $c > 0$ a constant, on ONEMAX is at most*

$$\frac{9e^c}{c(2c + 9)} \cdot n \ln(n) + O(n).$$

For $c = 1$ this is $\frac{9}{11} \cdot en \ln(n) + O(n) \approx 2.224n \ln(n) + O(n)$. The upper bound follows from applying the analytical framework in [5] with a corrected transition probability for p_r , using the value $1/(4e)$ instead of $5/(24e)$ (we shall give more details in Sect. 3). It can also be proven with mild adaptations of the proof of [27, Theorem 4]. We provide a self-contained proof in Sect. 3.

Our main contribution is the following lower bound that matches the upper bound proven in Theorem 1 up to small-order terms.

Theorem 2 *The expected optimisation time of the (2+1) GA with mutation rate c/n , and $0 < c \leq 1.422$ a constant, on ONEMAX is at least*

$$\frac{9e^c}{c(2c + 9)} \cdot n \ln(n) - O(n \log \log n).$$

Since the bounds from Theorems 1 and 2 have the same leading constant $\frac{9e^c}{c(2c+9)}$, which is minimised for

$$c = \frac{\sqrt{97} - 5}{4} \approx 1.21221445,$$

we identify this as the optimal mutation rate for the (2+1) GA (up to small-order terms) within the range of rates covered by Theorem 2.

Theorem 3 *Amongst all mutation rates c/n with $c \in (0, 1.422]$, the choice $c = \frac{\sqrt{97}-5}{4}$ is the optimal mutation rate of the (2+1) GA on ONEMAX, up to small-order terms. Then the expected optimisation time is $\approx 2.18417n \ln(n) + O(n)$.*

The best identified mutation rate for the (2+1) GA is lower than the one minimising the upper bound for larger population sizes $\mu \geq 5$ (it is at least $1.425/n$ and increases with μ) always providing upper bounds below $1.7n \ln n$ and decreasing with μ [6]. This implies that the (2+1) GA with mutation rate $\frac{\sqrt{97}-5}{4n}$ is at least 28% slower than any $(\mu+1)$ GA with $\mu \geq 5$ and appropriate mutation rates.

Algorithm 1 (2+1) GA

```

Initialize  $P = \{x^1, x^2\}$  by selecting two search points from  $\{0, 1\}^n$  independently and uniformly at
random (u. a. r.).
for  $t \leftarrow 1, 2, \dots$  do
  Select  $y^1$  and  $y^2$  from  $P$  u.a.r. with replacement.
  Create  $z$  by applying uniform crossover to  $y^1$  and  $y^2$ .
  Flip each bit in  $z$  independently with probability  $c/n$ .
  Remove the worst of  $x^1, x^2$  and  $z$ , breaking ties u. a. r.
end for

```

Structure of the paper. Section 2 formally defines the (2+1) GA and lists important tools for the analysis, including the drift theorem used for our main result. Section 3 presents the above-mentioned, corrected upper bound from Theorem 1 and its self-contained proof. In Sect. 4, we introduce the potential function that captures the state of the (2+1) GA and is crucial for the drift analysis proving the lower bound in Theorem 2. As determining the drift requires a careful case analysis, we give a roadmap of this analysis in Sect. 5, followed by the technical Sects. 6–8 that analyze the drift in different scenarios. In Sect. 9 we then put all pieces of our analysis together to complete the proof of Theorem 2. In addition, Sect. 10 gives empirical results and results of regression analyses. The latter confirm that our theoretical results are remarkably precise and provide further insight into small-order terms. We finish with some conclusions. To streamline the presentation of the most important cases in the drift analysis, some less insightful proofs have been moved to an appendix.

This paper extends a previous conference paper [24] where most proofs had to be omitted because of space constraints. The experimental results have not been published before either.

2 Preliminaries

The (2+1) GA is defined in Algorithm 1. The algorithm initialises the population with two randomly chosen individuals. At each generation it selects two random parents with replacement to be mated via uniform crossover. The operator assigns each bit to the offspring by selecting the corresponding bit from one parent with probability $1/2$ and from the other with the same probability. Standard bit mutation is then applied to the offspring by flipping each of its bits independently with probability c/n . Finally, the worst individual amongst the parents and the offspring is removed to select the new population. Ties are broken uniformly at random.

We will analyse the expected runtime of the algorithm to optimise the function $f(x) = \text{OneMax}(x) = \sum_{i=1}^n x_i$, which counts the number of 1-bits in a bitstring. Formally, the runtime is defined as the number of fitness function evaluations, which equals the smallest t such that the current population contains an optimum, up to an additive term of at most 2 stemming from evaluating the initial population.

Throughout this article, we will use state-of-the-art analysis techniques for randomized search heuristics, including drift analysis and concentration inequalities. In the

following, we list three important tools, starting with the drift theorem that is crucial for the lower bound on the runtime.

Theorem 4 (Multiplicative drift, lower bound, e. g., [29])

Let $X_t, t \geq 0$, be a stochastic process, adapted to a filtration \mathcal{F}_t , over a state space $S \subseteq \mathbb{R}^{\geq x_{\min}}$, where $x_{\min} > 0$. Assume that X_t is non-increasing, i. e., $X_{t+1} \leq X_t$ for all $t \geq 0$. Let T be the smallest $t \geq 0$ such that $X_t \leq x_{\min}$. If there exist positive real numbers $\beta, \delta > 0$ such that for all $t < T$ it holds that

1. $E(X_t - X_{t+1} \mid \mathcal{F}_t) \leq \delta X_t$
2. $P(X_t - X_{t+1} \geq \beta X_t) \leq \beta \delta / \ln(X_t)$

then

$$E(T \mid \mathcal{F}_0) \geq \frac{1 - \beta \ln(X_0/x_{\min})}{1 + \beta} \frac{1}{\delta}$$

The following Chernoff bound goes back to [4] and is formulated in the style of [9, Theorem 1.10.1].

Theorem 5 Let X_1, \dots, X_n be independent random variables taking values in $[0, 1]$. Let $X := \sum_{i=1}^n X_i$. Let $\delta \geq 0$. Then

$$P(X \geq (1 + \delta)E(X)) \leq e^{-\min\{\delta^2, \delta\}E(X)/3}.$$

The following lemma bounds a sum by an integral.

Lemma 6 For any integrable function $f : \mathbb{R} \rightarrow \mathbb{R}$, the following statements hold.

1. If f is non-increasing in $[a, b]$, $\sum_{i=a}^b f(i) \leq f(a) + \int_a^b f(i) \, di$.
2. If f is non-decreasing in $[a, b]$, $\sum_{i=a}^b f(i) \leq f(b) + \int_a^b f(i) \, di$.
3. If there is an $\alpha \in [a, b]$ such that f is non-increasing in $[a, \alpha]$ and non-decreasing in $[\alpha, b]$ then $\sum_{i=a}^b f(i) \leq f(a) + f(b) + \int_a^b f(i) \, di$.

Proof The first statement follows from splitting off the term $f(a)$ and using $\int_{j-1}^j f(i) \, di \geq f(j)$ for all $j \in [a+1, b]$, which follows from f being non-increasing. This implies $\sum_{i=a+1}^b f(i) \leq \int_a^b f(i) \, di$. The second statement is proved similarly. For the last statement, we note $\sum_{i=a}^b f(i) \leq \sum_{i=a}^{\lfloor \alpha \rfloor} f(i) + \sum_{i=\lceil \alpha \rceil}^b f(i)$. Then we apply the first statement to $\sum_{i=a}^{\lfloor \alpha \rfloor} f(i)$ and the second statement to $\sum_{i=\lceil \alpha \rceil}^b f(i)$, yielding a bound of $f(a) + f(b) + \int_a^{\lfloor \alpha \rfloor} f(i) \, di + \int_{\lceil \alpha \rceil}^b f(i) \, di \leq \sum_{i=a}^b f(i) \leq f(a) + f(b) + \int_a^b f(i) \, di$. \square

Finally, we frequently need the following estimate of the largest binomial coefficient, which can be found in [9, Inequality 1.4.18].

Lemma 7 For all $n \in \mathbb{N}$ and $k \in \{1, \dots, n\}$ it holds that

$$\binom{n}{k} \leq \binom{n}{\lceil n/2 \rceil} \leq 2^n \sqrt{\frac{2}{\pi n}}.$$

3 Upper Bounds for the (2+1) GA

In this section we aim to convince the reader in two different ways that the upper bound for the (2+1) GA is as stated in Theorem 1, with a leading constant of $\frac{9e^c}{c(2c+9)}$ instead of the leading constant $\frac{4e^c}{c(c+4)}$ claimed in [5].

Sudholt [27] provided an upper bound for $(\mu+\lambda)$ GAs with a diversity mechanism in the tie-breaking rule for the replacement selection. The analysis is based on a fitness-level argument and a simple Markov chain analysis made on each fitness level. Corus and Oliveto [5] analysed the standard $(\mu+1)$ GA, for which the choice $\mu = 2$ yields the (2+1) GA as defined in Algorithm 1. They observed that the $(\mu+1)$ GA can lose diversity after creating it, which required a more complex Markov chain analysis of each fitness level. We first explain their framework and argue why the leading constant is $\frac{9e^c}{c(2c+9)}$. In addition, we give a self-contained bound based on the analysis in [27] with an ad-hoc Markov chain framework similar to the one used in [5], simplified for the fixed value of $\mu = 2$ (and $\lambda = 1$). This shows that both previous works [5,27], with appropriate modifications, yield the same upper bound for the (2+1) GA.

3.1 The Framework by Corus and Oliveto

Corus and Oliveto coupled the standard artificial fitness levels method [15] with a Markov chain framework to bound the expected runtime of the $(\mu+1)$ GA for ONEMAX and any population size μ [5]. More precisely, they divided the search space into the canonical $n + 1$ fitness levels, each containing all search points with i 1-bits, and assumed that the algorithm is in level L_i if all the individuals of the population have exactly i 1-bits. Since crossover may only speed up the optimisation process if diversity is present in the population, a Markov chain was used on each fitness level to distinguish between populations with and without diversity. The Markov chain at fitness level i consists of two transient states $S_{1,i}$ and $S_{2,i}$ (i.e., resp. with/without diversity) and an absorbing state $S_{3,i}$, that is reached when a solution with better fitness is identified. For each level the analysis was performed by pessimistically assuming that initially on each level the population has no diversity (i.e., all individuals are identical). For this assumption to hold, and a valid upper bound on the expected runtime achieved, once the absorbing state is reached for level L_i , the expected time for the improved individual to take over the population has to be taken into account before the analysis of the absorption time of the Markov chain for level L_{i+1} may be carried out (i.e., the population is initially in state $S_{1,i+1}$ for each level L_i). When the algorithm is on the current level L_i with no diversity ($S_{1,i}$) only two changes of states may occur, both due to mutation: either the absorption state $S_{3,i}$ is reached by increasing the number of 1-bits in an individual, or a state with some diversity is reached by switching the order of 1-bits and 0-bits in an offspring, hence reaching state $S_{2,i}$. From this state, $S_{3,i}$ may be reached more quickly than from state $S_{1,i}$ if diverse parents are selected for reproduction and at least an extra 1-bit obtained in the offspring via crossover with constant probability if the diversity is not lost before (eg., by selecting two identical parents, creating a copy by not flipping any bits, and removing the diverse individual in the selection for replacement step). Overall, the proof strategy requires summing up

the bounds on the absorption times of the $n + 1$ Markov chains and subsequent takeover times over the $n + 1$ levels according to the artificial fitness levels methodology.

The described technique allows to calculate the transition probabilities of the Markov chain, hence to provide upper bounds on the expected runtime of the $(\mu+1)$ GA, as a function of any population size μ . In particular, it provides bounds on the expected runtime of the $(\mu+1)$ GA up to the leading constant for population sizes of $\mu = o(\log n / \log \log n)$ that are smaller than those of any steady-state EA using only standard bit mutation (i.e., no crossover) (i.e., for larger population sizes the take over time becomes asymptotically larger than the $O(n \log n)$ expected runtime of the (1+1) EA). An important insight from the Markov chain analysis is that the probability of losing diversity in state $S_{2,i}$ is higher for population size $\mu = 2$ than for larger populations: in the former case the diversity may be completely lost in every generation by either of the two individuals taking over, which is not the case for larger populations. However, in the analysis of this transition probability for the special case $\mu = 2$, the bound provided by the authors of [5] is by a factor of 2 smaller than the correct one due to a mistake in the calculation of the probability that different parents are selected for reproduction and the variation operators produce an offspring identical to either parent (and the differing parent is removed by selection of replacement). This miscalculation led to an upper bound on the expected runtime of the (2+1) GA of $\frac{4e^c n \ln n}{c(c+4)} + O(n)$ instead of the correct bound of $\frac{9e^c n \ln n}{c(2c+9)} + O(n)$ provided in Theorem 1. In the following subsection we provide a self-contained proof of the result and point out exactly where the miscalculation occurred in [5].

3.2 A Self-Contained Upper Bound Proof for the (2+1) GA

Here we give a self-contained proof of the upper bound for the (2+1) GA, Theorem 1.

Proof of Theorem 1 We use straightforward adaptations of the proof of [27, Theorem 4], using the same notation as in [27]. As in said proof, we distinguish between the following cases that are labelled according to the current best fitness in the population, called i . There are cases $i.1$, $i.2$, and $i.3$ explained in the following. We estimate the expected time spent in all these cases, summed up over all possible values of i , to obtain an upper bound on the total expected optimisation time.

Case i.1: the population contains a search point with fitness i and one search point of lower fitness.

This case is left for good if another search point of fitness at least i is created as then both search points will have fitness at least i . A sufficient condition for this is that the parent with fitness i is selected twice as parent (probability $1/4$) and no 1-bit is flipped during mutation (probability at least $(1 - c/n)^{n-1} = \Omega(1)$). Hence the expected waiting time in case $i.1$ is $O(1)$.

The other two cases are:

Case i.2: the population contains two copies of the same individual, with fitness i .

Case i.3: the population contains two different individuals, both having fitness i .

The algorithm can switch between the two cases as diversity can be gained or lost.

We first bound the expected time spent in Case $i.3$. This case can be left for good if the two different search point are chosen as parents (probability $1/2$) if crossover

generates a surplus of 1-bits (estimated from below by $1/4$ in [27, page 427]) and if mutation does not flip any bits (probability at least $(1 - c/n)^n$). Together, we obtain a probability of at least $(1 - c/n)^n/8$. Hence, the expected time spent in Case $i.3$ is at most $8/(1 - c/n)^n = O(1)$.

Let $T_{i,2}$ be the random time spent in Case $i.2$. We leave this case for good if mutation only flips a single 0-bit and no 1-bit. This event has probability at least $p^+ := (n - i) \cdot c/n \cdot (1 - c/n)^{n-1}$. We further make a transition to Case $i.3$ if mutation flips exactly one 0-bit and one 1-bit, leading to a different search point with fitness i , and then choosing one of the identical parents for removal (probability $2/3$). The probability of said events is at least $p^{2 \rightarrow 3} := i(n - i) \cdot (c/n)^2 \cdot (1 - c/n)^{n-2} \cdot 2/3$.

From Case $i.3$ the algorithm can move back to Case $i.2$, so that we may have to consider multiple visits to Case $i.2$. A necessary condition for going back to Case $i.2$ is that the created offspring is identical to one of the search points in the population, and the other search point is being selected for removal (probability $1/3$). The probability for creating an identical offspring is maximised when the two search points have Hamming distance 2. Either the same parent is selected twice (probability $1/2$) and mutation does not flip any bit (probability $(1 - c/n)^n$), or the same parent is selected twice and mutation creates the other parent (probability $O(1/n^2)$), or different parents are selected (probability $1/2$), crossover and mutation set the two differing bits identical to one of the parents (probability $1/2$ i.e., this is the probability that was wrongly estimated to be $1/4$ in [5] leading to an upper bound of $5/(24e)$ for going back to Case $i.2$ rather than the correct $1/(4e)$ bound which we derive now) and mutation does not flip any of the common bits (probability $(1 - c/n)^{n-2}$). Together, the probability of going back to Case $i.2$ is at most

$$\frac{1}{3} \cdot \left(\frac{(1 - c/n)^n}{2} + O(n^{-2}) + \frac{(1 - c/n)^{n-2}}{4} \right) = \frac{(1 - c/n)^n}{4} + O(n^{-1}).$$

Hence the conditional probability that, when Case $i.3$ is left towards Case $i.2$ or higher fitness, it is left towards Case $i.2$ is at most

$$p^{3 \rightarrow 2} := \frac{(1 - c/n)^n/4 + O(n^{-1})}{(1 - c/n)^n/4 + O(n^{-1}) + (1 - c/n)^n/8} = \frac{2}{3} + O(n^{-1}).$$

With these pessimistic estimations for transition probabilities, we obtain the following recurrence:

$$T_{i,2} \leq 1 + (1 - p^+ - p^{2 \rightarrow 3})T_{i,2} + p^{2 \rightarrow 3}p^{3 \rightarrow 2}T_{i,2}$$

which is equivalent to

$$T_{i,2} \leq \frac{1}{p^+ + p^{2 \rightarrow 3}(1 - p^{3 \rightarrow 2})}$$

and plugging in the above values yields the following upper bound for $T_{i,2}$:

$$\frac{1}{(n-i) \cdot c/n \cdot (1 - c/n)^{n-1} + i(n-i) \cdot (c/n)^2 \cdot (1 - c/n)^{n-2} \cdot 2/3 \cdot (1 - \frac{2}{3} - O(n^{-1}))} \leq \frac{1}{(n-i) \cdot c/n + i(n-i) \cdot (c/n)^2 \cdot 2/9} \cdot (e^c + O(n^{-1})).$$

Summing up these upper bounds for all i yields a total time bound of

$$\begin{aligned} &\leq (e^c + O(n^{-1})) \sum_{i=0}^{n-1} \frac{1}{(n-i) \cdot c/n + i(n-i) \cdot (c/n)^2 \cdot 2/9} \\ &= (e^c + O(n^{-1})) \frac{n}{c} \sum_{i=0}^{n-1} \frac{1}{(n-i)(1 + i \cdot c/n \cdot 2/9)}. \end{aligned}$$

A closer inspection of the function $f(i) := \frac{1}{(n-i)(1+i \cdot c/n \cdot 2/9)}$ reveals that it is non-decreasing in $[0, n - 1]$ if $c \leq 9/2$. For $c > 9/2$ it is non-increasing in $[0, n(2c - 9)/(4c)]$ and non-decreasing in $[n(2c - 9)/(4c), n - 1]$. In either case, we obtain an upper bound from Lemma 6:

$$\begin{aligned} &\sum_{i=0}^{n-1} \frac{1}{(n-i) + i(n-i) \cdot c/n \cdot 2/9} \\ &\leq \frac{1}{n} + \frac{1}{1 + (n-1) \cdot c/n \cdot 2/9} + \int_{i=0}^{n-1} \frac{1}{(n-i)(1 + i \cdot c/n \cdot 2/9)} \, di. \end{aligned}$$

The first two summands are in $O(1)$. The integral simplifies as follows, using e.g. equation 3.3.20 in [1]:

$$\begin{aligned} \int_{i=0}^{n-1} \frac{1}{(n-i)(1 + i \cdot c/n \cdot 2/9)} \, di &= \left[\frac{9}{2c+9} \cdot \ln \left(\frac{1 + i \cdot c/n \cdot 2/9}{n-i} \right) \right]_0^{n-1} \\ &= \frac{9}{2c+9} \left(\ln \left(\frac{1 + (n-1) \cdot c/n \cdot 2/9}{1} \right) - \ln \left(\frac{1}{n} \right) \right) \\ &= \frac{9}{2c+9} \cdot \ln(n + (n-1) \cdot c \cdot 2/9). \end{aligned}$$

Plugging everything together, the total time bound is

$$\begin{aligned} &(e^c + O(n^{-1})) \frac{n}{c} \left(O(1) + \frac{9}{2c+9} \cdot \ln(n + (n-1) \cdot c \cdot 2/9) \right) \\ &= \frac{9e^c}{c(2c+9)} \cdot n \ln(n + (n-1) \cdot c \cdot 2/9) + O(n) \end{aligned}$$

$$\leq \frac{9e^c}{c(2c + 9)} \cdot n \ln(n) + O(n).$$

Noting that the times in all Cases $i.1$ and $i.3$, summed up for all i , are $O(n)$ proves the claim. \square

4 A Potential Function Approach

We now turn to the main contribution of this paper, the tight lower bound for the (2+1) GA. Before going into detail, this section describes the main idea behind our approach and clarifies some further notation and fundamental observations that will be used in the remainder.

We write a population $\{x^1, x^2\}$ in order of monotonically decreasing fitness, that is, $f(x^1) \geq f(x^2)$. Let n_{11} be the number of bit positions where both parents have ones and likewise for n_{00} and the number of zeros. Let n_{10} be the number of positions where x^1 has a 1 and x^2 has a 0 and likewise for n_{01} . Then we have $f(x^1) = n_{11} + n_{10}$ and $f(x^2) = n_{11} + n_{01}$. Since by assumption, $f(x^1) \geq f(x^2)$, we have $n_{10} \geq n_{01}$ and $n_{10} = n_{01}$ is equivalent to the two individuals having equal fitness. In case $n_{10} = 0$, both individuals are identical. Such a population is called *monomorphic* in population genetics, and we use this term here.

Note that the (2+1) GA is an *unbiased* algorithm in the sense of Lehre and Witt [19]. In brief, this means that the algorithm treats all bit positions and all bit values symmetrically when generating new search points. Owing to this symmetry of bit positions and the fact that the fitness function is symmetrical itself, i. e., it only depends on the number but not the positions of the one-bits, it suffices to know n_{11} , n_{10} and n_{01} to fully characterise the state of the algorithm. Note that n_{00} can be derived as $n - n_{11} - n_{10} - n_{01}$.

The following lemma characterises probabilities of setting a bit to 1 in the offspring after a crossover of two different parents and a mutation of the result.

Lemma 8 *Consider a crossover of two parents x, y followed by mutation with mutation rate p_m , resulting in an offspring z . For all i ,*

$$P(z_i = 1) = \begin{cases} 1 - p_m & \text{if } x_i = y_i = 1 \\ 1/2 & \text{if } x_i \neq y_i \\ p_m & \text{if } x_i = y_i = 0. \end{cases}$$

Proof If $x_i = y_i$ then crossover will create an offspring with the same bit value. The statement for $x_i \neq y_i$ holds because of symmetry, or using the following, alternative argument. The offspring has a 1 if crossover creates a 1 and mutation does not flip bit i , or if crossover creates a 0 and mutation does flip bit i . The probability of the former event is $1/2 \cdot (1 - p_m)$ and the probability of the latter event is $1/2 \cdot p_m$. Together, this gives $1/2$. \square

Note that differing bits $x_i \neq y_i$ are set to 1 with probability $1/2$, irrespective of the mutation rate. Hence, when two parents are selected, we only need to consider the

effect of mutation on the bits where the parents agree. We frequently and tacitly use this fact.

Our lower bound applies when only considering populations where the number of zeros in the fitter parent is at most $n/\text{polylog}(n)$ and at least $\text{polylog}(n)$. This implies that all probabilities that involve flipping a 0 to 1 are polylogarithmically small.

The main tool for our lower bound is going to be drift analysis, applied to a potential function that captures the current state and potential easy fitness improvements.

Definition 1 For a population P with values $n_{11}, n_{10}, n_{01}, n_{00}$ we define the potential of P as

$$\varphi(P) = n_{11} + n_{10} + \frac{n_{01}}{3}.$$

The intuition is that $n_{11} + n_{10}$ describes the current best fitness in the population. The term $n_{01}/3$ adds potential to the best fitness as the population has the potential to exploit the diversity given by the n_{01} 1-bits that only exist in the less fit individual during a successful crossover operation.

The choice of the factor $1/3$ is motivated as follows. We know from previous work [5,27] that the most helpful populations for improvements are those where two search points have the same number of ones and Hamming distance 2, that is, $n_{10} = n_{01} = 1$. (Larger Hamming distances have the potential for larger fitness improvements, but such populations are rarely reached when the number of zeros becomes reasonably small.)

Assume the current state has $n_{10} = n_{01} = 1$, corresponding to a potential of $n_{11} + n_{10} + 1/3$. The most likely transitions (and, when only $O(n/\text{polylog}(n))$ zeros are left, the only transitions with probability $\Omega(1)$) are (1) collapsing the population to copies of one parent (and potential $n_{11} + n_{10}$) and (2) creating a surplus of one 1-bit by crossover and not flipping anything else (potential $n_{11} + n_{10} + 1$). The probability of the former event is roughly³ $(1 - p_m)^n/4$, which is the probability of selecting the same parent twice, not flipping any bits and then selecting the other population member for removal plus the probability of selecting different parents, creating one parent by crossover and not flipping any bits. The probability of the latter event is roughly $(1 - p_m)^n/8$, which is the probability of selecting different parents, setting both differing bits to 1 and not flipping any bits in the subsequent mutation.

Comparing these terms, the conditional probability of an improvement via crossover is roughly $1/3$. In case a monomorphic population is reached, the potential reduces by $1/3$ and this happens with conditional probability $1 - 1/3$. In the latter event, the potential increases by $1 - 1/3$ and this happens with conditional probability $1/3$. The net effect of these transitions in the expected change of the potential is $(1 - 1/3)1/3 - 1/3(1 - 1/3) = 0$. So the potential balances out the effects of “volatile” states left quickly.

Obviously, our analysis still needs to account for other, less likely transitions. For populations with $n_{10} = n_{01} > 1$ the conditional transition probabilities change as the probability of creating one of the parents by crossover depends on the Hamming

³ In this informal discussion we ignore events of smaller probability (e.g. picking the same parent twice and creating the other population member by a lucky mutation).

distance $n_{10} + n_{01}$ between parents. For $n_{10} = n_{01} > 1$ the likely progress in a successful crossover may be smaller than $n_{01}/3$. Hence the term $+n_{01}/3$ in Definition 1 is a precise estimate for the likely progress when $n_{01} = 1$ and for larger n_{01} it is an overestimation.

It suffices to restrict our considerations to moderate values of n_{10} and n_{01} . The reason is that the (2+1) GA always has a constant probability of creating a monomorphic population in one generation, regardless of the current population. This means that large values of n_{10} and n_{01} are very unlikely.

Lemma 9 *Let $t \geq \log^2 n$ and $t = n^{O(1)}$. With probability $1 - n^{-\Omega(\log n)}$, all populations within the time interval $[\log^2 n, t]$ have Hamming distance at most $\log^2 n$ between their two individuals.*

Proof We call a generation that creates a population of two identical individuals a *monomorphic* generation. The crucial idea is to show that monomorphic generations are very frequent so that large Hamming distances are unlikely to occur.

The probability of a monomorphic generation happening is at least $(1/4)(1 - 1/n)^n(1/3) = \Omega(1)$ since it is sufficient to select a fittest parent twice, to clone it and to remove the other parent (which has probability at least $1/3$). For a number $t \geq 0$ of generations after a monomorphic one, let D_t denote the maximum number of bits in which the two parents *ever* have differed during these t generations. The crucial idea is that only mutations can increase this D -value. The total number of bits flipped in t generations is the sum of tn Poisson trials with success probability c/n each. Hence, within t generations following a monomorphic one, the D -value is bounded from above by $2ct$ with probability $1 - 2^{-\Omega(t)}$ according to Chernoff bounds (Theorem 5), and clearly the Hamming distance is no larger than the D -value. We set $t := (\log^2 n)/(2c)$ to bound the D -value by $\log^2 n$.

The proof is completed by noting that the probability of not observing a monomorphic generation within $\log^2 n$ generations is $(1 - \Omega(1))^{\log^2 n} = n^{-\Omega(\log n)}$. Together with the failure bound $2^{-\Omega(t)}$, which is $n^{-\Omega(\log n)}$ for $t = (\log^2 n)/(2c)$, and a union bound, this means that in any polynomial number of generations following the first monomorphic one the Hamming distance never exceeds $\log^2 n$ with probability $1 - n^{O(1)}n^{-\Omega(\log n)} = 1 - n^{-\Omega(\log n)}$. □

5 Roadmap for the Analysis

We give a deliberately informal, high-level view of our analysis, where $\Delta := \varphi(P_{t+1}) - \varphi(P_t)$ denotes the change in potential in one generation. By the law of total probability, Δ can be split up according to the number of zeros flipped by mutation:

$$\begin{aligned} E(\Delta) &= E(\Delta \mid \text{no zeros flip}) \cdot P(\text{no zeros flip}) \\ &\quad + E(\Delta \mid \text{one zero flips}) \cdot P(\text{one zero flips}) \\ &\quad + E(\Delta \mid \text{at least 2 zeros flip}) \cdot P(\text{at least 2 zeros flip}). \end{aligned}$$

The case that no zeros flip does not increase the potential, hence we aim to bound the first line from above by 0. The second line captures the most important case: one zero flips and subsequent progress is made. The second line will be bounded by the dominant term in our claimed lower bound. The third line involves the probability of flipping at least two zeros. If the number of zeros is small, this is unlikely and thus the third line only contributes a small order term.

The above high-level view is not particularly accurate. Firstly, the above estimations need to account for error terms. Secondly, the notion of “ i zeros flip” used above is not well-defined. This is because the number of zeros that can flip during mutation depends on the parent selection. The same parents may be selected twice, and then the number of zeros depends on the fitness of the parent. If two different parents are used, we only consider mutations of bits that agree in both parents, as per Lemma 8.

Hence, we need to distinguish between different events from the parent selection. To formalise this, let P_{11} , P_{22} , and P_{12} denote the events that parent selection chooses the first parent twice, the second parent twice and both parents, respectively. We further denote by F_{00} the number of flipping bits amongst the n_{00} bits and likewise for F_{11} and n_{11} bits. We use asterisks to indicate the union of sets: F_{0*} is the number of flipping bits among $n_{01} + n_{00}$ bits and F_{*0} is the number of flipping bits among $n_{10} + n_{00}$ bits. Variables F_{1*} and F_{*1} are defined analogously. Armed with this notation, we express the third line rigorously with a combination of events.

Lemma 10 *For all populations with $n_{11} \geq n - n/\log^3 n$ and $n_{10} + n_{01} \leq \log^2 n$,*

$$\begin{aligned} & E(\Delta \mid P_{11}, F_{0*} \geq 2)P(F_{0*} \geq 2) \\ & + E(\Delta \mid P_{22}, F_{*0} \geq 2)P(F_{*0} \geq 2) \\ & + E(\Delta \mid P_{12}, F_{00} \geq 2)P(F_{00} \geq 2) = O(n_{0*}/(n \log n)) \end{aligned}$$

Proof We give one common way of bounding the three lines from the statement. For all $F \in \{F_{0*}, F_{*0}, F_{00}\}$, the drift can only increase by at most $n_{01} + F$ as every flipping 0-bit can only increase the potential by at most 1 and crossover can increase the potential on all n_{01} by at most 1. Also note that $P(F_{*0} \geq 2)$ has the largest probability amongst all variables F (as the underlying number of zeros is maximal for F_{*0}). Thus, all 3 lines are bounded as

$$\begin{aligned} & E(\Delta \mid F \geq 2)P(F_{*0} \geq 2) \\ & = \sum_{i=2}^{\infty} E(\Delta \mid F = i)P(F_{*0} = i) \leq \sum_{i=2}^{\infty} (n_{01} + i) \cdot P(F_{*0} = i) \\ & \leq \sum_{i=2}^{\infty} (n_{01} + i) \left(\frac{cn_{0*}}{n}\right)^i = \left(\frac{cn_{0*}}{n}\right)^2 \sum_{i=0}^{\infty} (n_{01} + i + 2) \left(\frac{cn_{0*}}{n}\right)^i \end{aligned}$$

Since $n_{0*} \leq n - n_{11} \leq n/\log^3 n$, we get $\left(\frac{cn_{0*}}{n}\right)^2 = O(n_{0*}/(n \log^3 n))$. The sum is bounded by $n_{01} + 2 + \sum_{i=0}^{\infty} i \left(\frac{cn_{0*}}{n}\right)^i$, which is $n_{01} + O(1) = O(\log^2 n)$. Together, this implies the claim. \square

The cases of no zeros flipping and one zero flipping are more difficult to handle. Corresponding drift estimates will be derived in the following sections. More precisely, we derive closed formulas for the drift of the potential function depending on the choice of parents in the following Sect. 6 and then distinguish between generations that flip no zeros (Sect. 7) and one zero (Sect. 8). As mentioned in the introduction, the drift bounds obtained in these major technical sections are finally put together in Sect. 9.

6 Closed Formulas for the Potential Drift

Before proceeding to bound the drift of the potential, we derive closed bounds for the potential under the condition P_{12} that different parents are being selected. These bounds hold for arbitrary given values of F_{00} , F_{11} , the numbers of flipping common zeros and ones, respectively. These closed formulas will then be used in the subsequent sections to bound the drift for the cases $F_{00} = 0$ and $F_{00} = 1$, respectively.

For the derivation of closed formulas it is important to distinguish between two cases: the parents having equal fitness, $n_{10} = n_{01}$, or the parents having unequal fitness, $n_{10} > n_{01}$. This is because, depending on the fitness of the offspring, different cases may occur and the probabilities used during replacement selection may differ. For instance, when both parents have an equal fitness and the offspring happens to have the same fitness as well, there is a 3-way tie that is resolved uniform randomly by selection. When both parents have different fitness, 3-way ties are not possible (there can only be ties between the offspring and one parent). However, the offspring might be strictly worse than the better parent and strictly better than the worse parent. Such a case cannot happen when $n_{10} = n_{01}$. It therefore makes sense to split the analysis into these two cases and to derive separate closed bounds on the potential drift.

6.1 A Closed Formula for Selecting Different Parents of Equal Fitness ($n_{10} = n_{01}$)

In the following lemma we shall first derive a closed formula for the potential drift when two different parents are chosen that have the same fitness, $n_{10} = n_{01}$. The lemma defines a threshold ℓ that reflects the number of bits crossover needs to set to 1 to achieve the same fitness as x^1 and x^2 .

Lemma 11 *Let $S \sim \text{Bin}(n_{10} + n_{01}, 1/2)$ and $\ell := n_{10} - F_{00} + F_{11}$. Then for all populations with $n_{10} = n_{01}$,*

$$\begin{aligned} & E(\Delta \mid P_{12}, F_{00}, F_{11}) \\ &= P(S \geq \ell) \cdot \left(\frac{5}{6} \cdot E(S \mid S \geq \ell) + F_{00} - \frac{2F_{11}}{3} - n_{10} \right) + P(S = \ell) \cdot \frac{n_{10} - F_{00} - F_{11}}{18}. \end{aligned}$$

Proof Consider a step where both parents are selected and F_{00} , F_{11} are known. The number of bits set to 1 by crossover is given by $S \sim \text{Bin}(n_{10} + n_{01}, 1/2)$. By the law

of total probability,

$$E(\Delta \mid P_{12}, F_{00}, F_{11}) = \sum_{s \in \mathbb{Z}} E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) \cdot P(S = s).$$

Since $n_{10} = n_{01}$, the fitness of both parents is $n_{11} + n_{10}$ and the fitness of the offspring is $n_{11} + s + F_{00} - F_{11}$. If $s < \ell$, the latter is less than $n_{11} + n_{10}$ and the offspring will be rejected. If $s > \ell$, the offspring is fitter than the parent, one of the parents will be chosen uniformly at random for removal. If the offspring is as fit as both parents, the offspring is removed with probability $1/3$. Thus,

$$\begin{aligned} E(\Delta \mid P_{12}, F_{00}, F_{11}) &= \sum_{s \in \mathbb{Z}} E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) \cdot P(S = s) \\ &= \sum_{s > \ell} E(\Delta \mid S = s) \cdot P(S = s) + \frac{2}{3} \cdot E(\Delta \mid S = \ell) \cdot P(S = \ell) \\ &= \sum_{s \geq \ell} E(\Delta \mid S = s) \cdot P(S = s) - \frac{1}{3} \cdot E(\Delta \mid S = \ell) \cdot P(S = \ell). \end{aligned}$$

Now we estimate $E(\Delta \mid S = s)P(S = s)$ for $s \geq \ell$. Let S_{10} be the number of bits among the n_{10} bits that are set to 1 in the offspring and define S_{01} analogously for the n_{01} bits. Note that $S := S_{10} + S_{01}$ where $S_{10} \sim \text{Bin}(n_{10}, 1/2)$ and $S_{01} \sim \text{Bin}(n_{01}, 1/2)$ according to Lemma 8.

Given $S_{10} = s_{10}$ and $S_{01} = s_{01}$, the potential difference Δ is derived as follows. Among the n_{11} bits, F_{11} bits flip to 0, reducing their contribution from 1 to $1/3$ each, leading to a contribution of $-2F_{11}/3$ to the potential difference. All the n_{10} bits contribute 1 to the potential $\varphi(P_t)$. In P_{t+1} , s_{10} bits contribute 1 and the remaining $n_{10} - s_{10}$ bits contribute $1/3$ each. Hence the contribution to the potential difference is $-2(n_{10} - s_{10})/3$. The n_{01} bits contribute $1/3$ each in $\varphi(P_t)$ and in P_{t+1} we have s_{01} bits contributing 1 each and the other $n_{01} - s_{01}$ bits contributing 0. Hence the contribution to the potential difference is $s_{01} - n_{01}/3$. Finally, the contribution of the n_{00} bits to the potential difference is F_{00} .

In the following we denote by $(X \mid Y)$ the conditional variable X conditional on Y , where Y is a sequence of events and/or random variables. Omitting the implicit conditions P_{12}, F_{00}, F_{11} for brevity, we obtain a conditional drift Δ of

$$\begin{aligned} (\Delta \mid S_{10} = s_{10}, S_{01} = s_{01}) &= F_{00} - \frac{2F_{11}}{3} - \frac{2(n_{10} - s_{10})}{3} + s_{01} - \frac{n_{01}}{3} \\ &= F_{00} - \frac{2F_{11}}{3} - n_{10} + s - \frac{1}{3} \cdot s_{10}. \end{aligned} \tag{1}$$

By the law of total probability,

$$E(\Delta \mid S = s) = \sum_{s_{10}=0}^s E(\Delta \mid S = s, S_{10} = s_{10})P(S_{10} = s_{10} \mid S = s)$$

$$\begin{aligned}
 &= \sum_{s_{10}=0}^s \left(F_{00} - \frac{2F_{11}}{3} - n_{10} + s - \frac{1}{3} \cdot s_{10} \right) P(S_{10} = s_{10} \mid S = s) \\
 &= F_{00} - \frac{2F_{11}}{3} - n_{10} + s - \frac{1}{3} \sum_{s_{10}=0}^s s_{10} \cdot P(S_{10} = s_{10} \mid S = s) \\
 &= F_{00} - \frac{2F_{11}}{3} - n_{10} + s - \frac{1}{3} \cdot E(S_{10} \mid S = s).
 \end{aligned}$$

Now, $(S_{10} \mid S = s)$ follows a hypergeometric distribution with parameters n_{10} (number of red balls), $n_{10} + n_{01}$ (number of balls) and s (number of draws). The expectation is thus $s \cdot n_{10} / (n_{10} + n_{01}) = s/2$. Plugging this in yields

$$E(\Delta \mid S = s) = F_{00} - \frac{2F_{11}}{3} - n_{10} + \frac{5}{6} \cdot s. \tag{2}$$

Together, this gives

$$\begin{aligned}
 E(\Delta \mid P_{12}, F_{00}, F_{11}) &= \sum_{s \geq \ell} E(\Delta \mid S = s) \cdot P(S = s) - \frac{1}{3} \cdot E(\Delta \mid S = \ell) \cdot P(S = \ell) \\
 &= \sum_{s \geq \ell} \left(F_{00} - \frac{2F_{11}}{3} - n_{10} + \frac{5}{6} \cdot s \right) \cdot P(S = s) \\
 &\quad - \frac{1}{3} \cdot \left(F_{00} - \frac{2F_{11}}{3} - n_{10} + \frac{5}{6} \cdot \ell \right) \cdot P(S = \ell).
 \end{aligned}$$

Using

$$\begin{aligned}
 \sum_{s \geq \ell} s \cdot P(S = s) &= \sum_{s \geq \ell} s \cdot P(S = s \wedge S \geq \ell) \\
 &= \sum_{s \geq \ell} s \cdot P(S = s \mid S \geq \ell) P(S \geq \ell) = E(S \mid S \geq \ell) P(S \geq \ell),
 \end{aligned}$$

the first terms simplify as

$$\begin{aligned}
 &\sum_{s \geq \ell} \left(F_{00} - \frac{2F_{11}}{3} - n_{10} + \frac{5}{6} \cdot s \right) \cdot P(S = s) \\
 &= \left(F_{00} - \frac{2F_{11}}{3} - n_{10} \right) \sum_{s \geq \ell} P(S = s) + \frac{5}{6} \sum_{s \geq \ell} s \cdot P(S = s) \\
 &= \left(F_{00} - \frac{2F_{11}}{3} - n_{10} \right) P(S \geq \ell) + \frac{5}{6} \cdot E(S \mid S \geq \ell) P(S \geq \ell).
 \end{aligned}$$

The last term simplifies as

$$\begin{aligned}
 & -\frac{1}{3} \cdot \left(F_{00} - \frac{2F_{11}}{3} - n_{10} + \frac{5}{6} \cdot \ell \right) \cdot P(S = \ell) \\
 &= -\frac{1}{3} \cdot \left(F_{00} - \frac{2F_{11}}{3} - n_{10} + \frac{5}{6} \cdot (n_{10} - F_{00} + F_{11}) \right) \cdot P(S = \ell) \\
 &= -\frac{1}{3} \cdot \left(\frac{F_{00}}{6} + \frac{F_{11}}{6} - \frac{n_{10}}{6} \right) \cdot P(S = \ell) \\
 &= -\frac{F_{00} + F_{11} - n_{10}}{18} \cdot P(S = \ell) = \frac{n_{10} - F_{00} - F_{11}}{18} \cdot P(S = \ell).
 \end{aligned}$$

Together, this proves the claim. □

The bound from Lemma 11 depends on the expected surplus $E(S \mid S \geq \ell)$ generated by a crossover on the bits that differ between the two parents, where ℓ reflects the fitness threshold above which offspring are accepted. We use the following formula to simplify such expressions. The proof goes back to work by Gruder [13] that is highlighted in a paper by Johnson [17]. It is given in the appendix for the sake of completion.

Lemma 12 *Let $S \sim \text{Bin}(n, 1/2)$, then for all $\ell \in \mathbb{N}$,*

$$E(S \mid S \geq \ell)P(S \geq \ell) = \frac{\ell}{2} \cdot P(S = \ell) + \frac{n}{2} \cdot P(S \geq \ell).$$

Using Lemma 12, we obtain the following simplified formula.

Lemma 13 *Let $S \sim \text{Bin}(n_{10} + n_{01}, 1/2)$. Then for all populations with $n_{10} = n_{01}$,*

$$\begin{aligned}
 E(\Delta \mid P_{12}, F_{00}, F_{11}) &= P(S \geq n_{10} - F_{00} + F_{11}) \cdot \left(F_{00} - \frac{2F_{11}}{3} - \frac{n_{10}}{6} \right) \\
 &\quad + P(S = n_{10} - F_{00} + F_{11}) \cdot \frac{17n_{10} - 17F_{00} + 13F_{11}}{36}.
 \end{aligned}$$

Proof. Recall $\ell := n_{10} - F_{00} + F_{11}$. By Lemma 11 and Lemma 12,

$$\begin{aligned}
 & E(\Delta \mid P_{12}, F_{00}, F_{11}) \\
 &= P(S \geq \ell) \cdot \left(\frac{5}{6} \cdot E(S \mid S \geq \ell) + F_{00} - \frac{2F_{11}}{3} - n_{10} \right) + P(S = \ell) \cdot \frac{n_{10} - F_{00} - F_{11}}{18} \\
 &= \frac{5}{6} \left(\frac{\ell}{2} \cdot P(S = \ell) + n_{10} \cdot P(S \geq \ell) \right) + P(S \geq \ell) \cdot \left(F_{00} - \frac{2F_{11}}{3} - n_{10} \right) \\
 &\quad + P(S = \ell) \cdot \frac{n_{10} - F_{00} - F_{11}}{18} \\
 &= P(S \geq \ell) \cdot \left(F_{00} - \frac{2F_{11}}{3} - \frac{n_{10}}{6} \right) + P(S = \ell) \cdot \left(\frac{5}{12} \cdot \ell + \frac{n_{10} - F_{00} - F_{11}}{18} \right) \\
 &= P(S \geq \ell) \cdot \left(F_{00} - \frac{2F_{11}}{3} - \frac{n_{10}}{6} \right) + P(S = \ell) \cdot \frac{15\ell + 2n_{10} - 2F_{00} - 2F_{11}}{36}
 \end{aligned}$$

$$= P(S \geq \ell) \cdot \left(F_{00} - \frac{2F_{11}}{3} - \frac{n_{10}}{6} \right) + P(S = \ell) \cdot \frac{17n_{10} - 17F_{00} + 13F_{11}}{36}. \quad \square$$

6.2 A Closed Formula for Selecting Different Parents of Unequal Fitness
($n_{10} > n_{01}$)

For the case of unequal fitness ($n_{10} > n_{01}$), we use similar arguments as before. This scenario has more involved calculations as we need to distinguish different cases: the offspring may be at least as good as the fitter parent, and then the calculations are similar to the equal-fitness scenario. The offspring may also be worse than the fitter parent and better than the worse parent. In this case, the potential is derived from a different formula as the values for n_{10} and n_{01} in the next generation are still determined according to the fitter parent. In case the offspring’s fitness is equal to that of the worse parent, there is a tie and the offspring is only accepted with probability 1/2. The lemma defines two thresholds ℓ_1 and ℓ_2 that reflect the number of bits crossover needs to set to 1 to achieve the fitness of x^1 and x^2 , respectively.

Lemma 14 *Let $S \sim \text{Bin}(n_{10} + n_{01}, 1/2)$, $\ell_1 := n_{10} - F_{00} + F_{11}$ and $\ell_2 := n_{01} - F_{00} + F_{11}$. Then for all populations with $n_{10} > n_{01}$,*

$$\begin{aligned} E(\Delta \mid P_{12}, F_{00}, F_{11}) &= P(S > \ell_1) \cdot \frac{2F_{00} - 2F_{11} - n_{10} + n_{01}}{3} \\ &\quad + P(S = \ell_1) \cdot \frac{F_{00} - F_{11} + n_{01}}{3} \\ &\quad + P(S > \ell_2) \cdot \frac{2F_{00} - n_{01}}{6} \\ &\quad + P(S = \ell_2) \cdot \frac{F_{00}}{6}. \end{aligned}$$

Proof Consider a step where both parents are selected and F_{00}, F_{11} are known. The number of bits set to 1 by crossover is given by $S \sim \text{Bin}(n_{10} + n_{01}, 1/2)$. By the law of total probability,

$$E(\Delta \mid P_{12}, F_{00}, F_{11}) = \sum_{s \in \mathbb{Z}} E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) \cdot P(S = s).$$

Note that the fitness of x^1 is $n_{11} + n_{10}$, that of x^2 is $n_{11} + n_{01}$ and the fitness of the offspring is $n_{11} + s + F_{00} - F_{11}$.

We distinguish the following cases.

1. If $s \geq \ell_1$, the offspring is at least as fit as the fitter parent, and x^2 will be removed. The new values of n_{10}, n_{01} will be determined by the offspring.
2. If $\ell_2 < s < \ell_1$, the offspring is fitter than the worse parent x^2 , and x^2 will be removed. Parent x^1 will remain the fitter parent.
3. If $s = \ell_2$, the offspring is as fit as x^2 and x^2 will be removed with probability 1/2. Then the potential is computed as in the case $\ell_2 < s < \ell_1$.

4. If $s < \ell_2$, the offspring is worse than x^2 and will be rejected. Hence the potential does not change.

Let Δ be the change of potential from the current population to the new population P_{t+1} as described above. Then,

$$\begin{aligned} E(\Delta \mid P_{12}, F_{00}, F_{11}) &= \sum_{s=\ell_1}^{n_{10}+n_{01}} E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) \cdot P(S = s) \\ &\quad + \sum_{s=\ell_2}^{\ell_1-1} E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) \cdot P(S = s) \\ &\quad - \frac{1}{2} \cdot E(\Delta \mid P_{12}, F_{00}, F_{11}, S = \ell_2) \cdot P(S = \ell_2), \end{aligned}$$

where the last line accounts for the fact that with probability $1/2$ the offspring is removed if $s = \ell_2$.

Now we estimate $E(\Delta \mid S = s)P(S = s)$, starting with the case for $s \geq \ell_1$. Let S_{10} be the number of bits among the n_{10} bits that are set to 1 in the offspring and define S_{01} analogously for the n_{01} bits. Note that $S := S_{10} + S_{01}$ where $S_{10} \sim \text{Bin}(n_{10}, 1/2)$ and $S_{01} \sim \text{Bin}(n_{01}, 1/2)$.

Given $S_{10} = s_{10}$ and $S_{01} = s_{01}$, the potential difference Δ is derived as follows. Among the n_{11} bits, F_{11} bits flip to 0, reducing their contribution from 1 to $1/3$ each, leading to a contribution of $-2F_{11}/3$ to the potential difference. All the n_{10} bits contribute 1 to the potential $\varphi(P_t)$. In P_{t+1} , s_{10} bits contribute 1 and the remaining $n_{10} - s_{10}$ bits contribute $1/3$ each. Hence the contribution to the potential difference is $-2(n_{10} - s_{10})/3$. The n_{01} bits contribute $1/3$ each in $\varphi(P_t)$ and in P_{t+1} we have s_{01} bits contributing 1 each and the other $n_{01} - s_{01}$ bits contributing 0. Hence the contribution to the potential difference is $s_{01} - n_{01}/3$. Finally, the contribution of the n_{00} bits to the potential difference is F_{00} . Together,

$$\begin{aligned} (\Delta \mid S_{10} = s_{10}, S_{01} = s_{01}) &= F_{00} - \frac{2F_{11}}{3} - \frac{2(n_{10} - s_{10})}{3} + s_{01} - \frac{n_{01}}{3} \\ &= F_{00} - \frac{2F_{11}}{3} - \frac{2n_{10}}{3} - \frac{n_{01}}{3} + \frac{2}{3} \cdot s_{10} + s_{01} \\ &= F_{00} - \frac{2F_{11}}{3} - \frac{2n_{10} + n_{01}}{3} + s - \frac{1}{3} \cdot s_{10}. \end{aligned} \tag{3}$$

By the law of total probability, for all $s \geq \ell_1$,

$$\begin{aligned} E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) &= \sum_{s_{10}=0}^s E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s, S_{10} = s_{10})P(S_{10} = s_{10} \mid S = s) \\ &= \sum_{s_{10}=0}^s \left(F_{00} - \frac{2F_{11}}{3} - \frac{2n_{10} + n_{01}}{3} + s - \frac{1}{3} \cdot s_{10} \right) P(S_{10} = s_{10} \mid S = s) \end{aligned}$$

$$\begin{aligned}
 &= F_{00} - \frac{2F_{11}}{3} - \frac{2n_{10} + n_{01}}{3} + s - \frac{1}{3} \sum_{s_{10}=0}^s s_{10} \cdot P(S_{10} = s_{10} \mid S = s) \\
 &= F_{00} - \frac{2F_{11}}{3} - \frac{2n_{10} + n_{01}}{3} + s - \frac{1}{3} \cdot E(S_{10} \mid S = s).
 \end{aligned}$$

Now, $(S_{10} \mid S = s)$ follows a hypergeometric distribution with parameters n_{10} (number of red balls), $n_{10} + n_{01}$ (number of balls) and s (number of draws). The expectation is thus $s \cdot n_{10} / (n_{10} + n_{01})$. Plugging this in yields

$$E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) = F_{00} - \frac{2F_{11}}{3} - \frac{2n_{10} + n_{01}}{3} + s \left(1 - \frac{n_{10}}{3n_{10} + 3n_{01}} \right). \tag{4}$$

Now we consider the case $\ell_2 \leq S < \ell_1$. Here x^1 remains the fitter parent and the potential difference Δ is derived as follows. Flipping bits among the n_{11} and n_{10} bits do not change the potential. We have $S_{01} + F_{00}$ bits that each contribute $1/3$ to the potential, compared to n_{01} bits in the previous population. Hence the change in potential is

$$(\Delta \mid P_{12}, F_{00}, F_{11}, S_{01} = s_{01}) = \frac{s_{01} - n_{01} + F_{00}}{3}. \tag{5}$$

By the law of total probability, for all $\ell_2 \leq s < \ell_1$,

$$\begin{aligned}
 &E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s) \\
 &= \sum_{s_{01}=0}^s E(\Delta \mid P_{12}, F_{00}, F_{11}, S = s, S_{01} = s_{01}) P(S_{01} = s_{01} \mid S = s) \\
 &= \frac{E(S_{01} \mid S = s) - n_{01} + F_{00}}{3} \\
 &= s \cdot \frac{n_{01}}{3n_{10} + 3n_{01}} + \frac{-n_{01} + F_{00}}{3}
 \end{aligned}$$

using the same calculations as before.

Plugging everything together,

$$\begin{aligned}
 &E(\Delta \mid P_{12}, F_{00}, F_{11}) \\
 &= \sum_{s=\ell_1}^{n_{10}+n_{01}} \left(F_{00} - \frac{2F_{11}}{3} - \frac{2n_{10} + n_{01}}{3} + s \cdot \frac{2n_{10} + 3n_{01}}{3n_{10} + 3n_{01}} \right) P(S = s) \\
 &\quad + \sum_{s=\ell_2}^{\ell_1-1} \left(s \cdot \frac{n_{01}}{3n_{10} + 3n_{01}} + \frac{-n_{01} + F_{00}}{3} \right) P(S = s) \\
 &\quad - \frac{1}{2} \cdot \left(\ell_2 \cdot \frac{n_{01}}{3n_{10} + 3n_{01}} + \frac{-n_{01} + F_{00}}{3} \right) P(S = \ell_2).
 \end{aligned}$$

By adding terms for $\ell_1 \leq s \leq n_{10} + n_{01}$ to the second line and subtracting them from the first line, we get

$$\begin{aligned}
 & E(\Delta \mid P_{12}, F_{00}, F_{11}) \\
 &= \sum_{s=\ell_1}^{n_{10}+n_{01}} \left(\frac{2F_{00}}{3} - \frac{2F_{11}}{3} - \frac{2n_{10}}{3} + \frac{2s}{3} \right) P(S = s) \\
 &\quad + \sum_{s=\ell_2}^{n_{10}+n_{01}} \left(s \cdot \frac{n_{01}}{3n_{10} + 3n_{01}} + \frac{-n_{01} + F_{00}}{3} \right) P(S = s) \\
 &\quad - \frac{1}{2} \cdot \left(\ell_2 \cdot \frac{n_{01}}{3n_{10} + 3n_{01}} + \frac{-n_{01} + F_{00}}{3} \right) P(S = \ell_2) \\
 &= P(S \geq \ell_1) \cdot \frac{2F_{00} - 2F_{11} - 2n_{10}}{3} + \frac{2}{3} \sum_{s=\ell_1}^{n_{10}+n_{01}} s \cdot P(S = s) \\
 &\quad + P(S \geq \ell_2) \left(\frac{-n_{01} + F_{00}}{3} \right) + \frac{n_{01}}{3n_{10} + 3n_{01}} \sum_{s=\ell_2}^{n_{10}+n_{01}} s \cdot P(S = s) \\
 &\quad - \frac{1}{2} \cdot \left(\ell_2 \cdot \frac{n_{01}}{3n_{10} + 3n_{01}} + \frac{-n_{01} + F_{00}}{3} \right) P(S = \ell_2).
 \end{aligned}$$

Applying Lemma 12 twice, to the first and second line, yields the following two equations:

$$\begin{aligned}
 \sum_{s=\ell_1}^{n_{10}+n_{01}} s \cdot P(S = s) &= \frac{\ell_1}{2} P(S = \ell_1) + \frac{n_{10} + n_{01}}{2} P(S \geq \ell_1) \\
 \sum_{s=\ell_2}^{n_{10}+n_{01}} s \cdot P(S = s) &= \frac{\ell_2}{2} P(S = \ell_2) + \frac{n_{10} + n_{01}}{2} P(S \geq \ell_2).
 \end{aligned}$$

Plugging this in yields

$$\begin{aligned}
 E(\Delta \mid P_{12}, F_{00}, F_{11}) &= P(S \geq \ell_1) \cdot \frac{2F_{00} - 2F_{11} - n_{10} + n_{01}}{3} \\
 &\quad + P(S = \ell_1) \cdot \frac{\ell_1}{3} \\
 &\quad + P(S \geq \ell_2) \cdot \frac{2F_{00} - n_{01}}{6} \\
 &\quad + P(S = \ell_2) \cdot \frac{n_{01} - F_{00}}{6}.
 \end{aligned}$$

□

7 Potential Drift When No Zeros Flip

We now consider the potential drift when no zeros flip. When the distance to the optimum is $o(n)$, this case is by far the most frequent case. This also means that our drift bounds have to be precise, as even a small error term may have a big impact and spoil the analysis.

7.1 Selecting the Same Parent Twice

We start by considering the drift conditional on selecting the same parent twice.

Lemma 15 *For all populations with $n_{10} = n_{01}$,*

$$E(\Delta \mid P_{11}, F_{0*} = 0)P(F_{0*} = 0) = -\frac{n_{01}}{9} \left(1 - \frac{c}{n}\right)^n$$

$$E(\Delta \mid P_{22}, F_{*0} = 0)P(F_{*0} = 0) = -\frac{n_{01}}{9} \left(1 - \frac{c}{n}\right)^n.$$

For all populations with $n_{10} > n_{01}$,

$$E(\Delta \mid P_{11}, F_{0*} = 0)P(F_{0*} = 0) = -\frac{n_{01}}{3} \left(1 - \frac{c}{n}\right)^n$$

$$E(\Delta \mid P_{22}, F_{*0} = 0)P(F_{*0} = 0) = 0.$$

Proof First assume $n_{10} = n_{01}$ and consider the event P_{11} . Given $F_{0*} = 0$, that is, if no 0-bit is flipped, the offspring can only be accepted if no 1-bit is flipped, i.e., $F_{1*} = 0$. These events happen with probability $P(F_{0*} = 0)P(F_{1*} = 0) = (1 - c/n)^n$ and they lead to an offspring that is identical to x^1 . Since all search points have equal fitness, x^2 is removed with probability $1/3$. This leads to a monomorphic population and the potential decreases by $n_{01}/3$. Multiplying the above terms proves the claimed equality. The case of P_{22} follows analogously, considering F_{*0} and F_{*1} instead.

For $n_{10} > n_{01}$, if the fitter parent x^1 is selected twice, a copy of it is created with probability $(1 - c/n)^n$ and then x^2 is removed. This decreases the potential by $n_{01}/3$. Multiplying the above terms yields an expectation of $-n_{01}/3 \cdot (1 - c/n)^n$. Note that other operations cannot increase the potential since no 0-bit is being flipped and flipping 1-bits in x^1 does not decrease the potential.

If x^2 is selected twice as parent, the potential cannot increase since no 0-bits are flipped, and it cannot decrease as any 1-bit being flipped will lead to the offspring being rejected. □

Now we consider the event that two different parents are selected. The remainder of this section is split into drift bounds when the parents have equal fitness, $n_{10} = n_{01}$, and the case where parents have unequal fitness, $n_{10} > n_{01}$.

7.2 Selecting Different Parents of Equal Fitness ($n_{10} = n_{01}$)

We use the closed formula from Lemma 13 to show that, to get an upper bound on the drift when $F_{00} = 0$, we only need to consider $F_{11} \in \{0, 1\}$ as larger values lead to a non-positive drift. This is not obvious, but follows from a lengthy and trite calculation. The proof is placed in the appendix to keep the main part streamlined.

Lemma 16 For all $n_{10} = n_{01}$ and all $i \geq 2$,

$$E(\Delta \mid P_{12}, F_{00} = 0, F_{11} = i) \leq 0.$$

Now we are able to give an upper bound on the potential drift, assuming that different parents are selected (P_{12}) and no common zero-bits flip ($F_{00} = 0$).

Lemma 17 For every population with $n_{10} = n_{01}$,

$$E(\Delta \mid P_{12}, F_{00} = 0) \leq \begin{cases} 0 & \text{if } n_{10} = 0 \\ \frac{1}{9} \cdot P(F_{11} = 0) & \text{if } n_{10} = 1 \\ \frac{1}{8} \cdot P(F_{11} = 0) + \frac{1}{64} \cdot P(F_{11} = 1) & \text{if } n_{10} \geq 2. \end{cases}$$

Proof By the law of total probability and Lemma 13,

$$\begin{aligned} E(\Delta \mid P_{12}, F_{00}) &= \sum_{i=0}^{\infty} E(\Delta \mid P_{12}, F_{00}, F_{11} = i) \cdot P(F_{11} = i) \\ &= \sum_{i=0}^{\infty} P(F_{11} = i) \left(P(S = n_{10} + i) \cdot \frac{17n_{10} + 13i}{36} - P(S \geq n_{10} + i) \cdot \frac{6n_{10} + 24i}{36} \right) \\ &= \sum_{i=0}^{n_{10}-1} P(F_{11} = i) \left(P(S = n_{10} + i) \cdot \frac{17n_{10} + 13i}{36} - P(S \geq n_{10} + i) \cdot \frac{6n_{10} + 24i}{36} \right) \end{aligned}$$

as the term in brackets is 0 for all values $i \geq n_{10}$.

For $n_{10} = 0$, the above is 0 and for $n_{10} = 1$, is it $P(F_{11} = 0)/9$ as claimed. For $n_{10} \geq 2$, by Lemma 16 all terms for $i \geq 2$ are non-positive and can be dropped. Thus, we get

$$\begin{aligned} E(\Delta \mid P_{12}, F_{00}) &\leq \sum_{i=0}^1 E(\Delta \mid P_{12}, F_{00}, F_{11} = i) \cdot P(F_{11} = i) \\ &= \sum_{i=0}^1 P(F_{11} = i) \left(P(S = n_{10} + i) \cdot \frac{17n_{10} + 13i}{36} - P(S \geq n_{10} + i) \cdot \frac{6n_{10} + 24i}{36} \right) \\ &= P(F_{11} = 0) \left(P(S = n_{10}) \cdot \frac{17n_{10}}{36} - P(S \geq n_{10}) \cdot \frac{6n_{10}}{36} \right) \end{aligned}$$

$$+ P(F_{11} = 1) \left(P(S = n_{10} + 1) \cdot \frac{17n_{10} + 13}{36} - P(S \geq n_{10} + 1) \cdot \frac{6n_{10} + 24}{36} \right).$$

Writing the above as

$$P(F_{11} = 0)a_0 + P(F_{11} = 1)a_1,$$

we have $a_0 = 0$ for $n_{10} = 0$ and $a_0 = 1/9$ for $n_{10} = 1$. The maximum value is $a_0 = 1/8$ for $n_{10} = 2$. This can be seen from evaluating a_0 for $0 \leq n_{10} \leq 10$ (shown in Table 1) and from the following analytical arguments for $n_{10} \geq 11$. We use the bound on the largest binomial coefficient $\binom{2n_{10}}{k} \leq \binom{2n_{10}}{n_{10}} \leq 2^{2n_{10}} / (\sqrt{\pi n_{10}})$ for all k according to Lemma 7 and the fact that $P(S \geq n_{10}) \geq 1/2$ by symmetry of the binomial distribution. Then a_0 is at most

$$\begin{aligned} & P(S = n_{10}) \cdot \frac{17n_{10}}{36} - P(S \geq n_{10}) \cdot \frac{6n_{10}}{36} \\ & \leq P(S = n_{10}) \cdot \frac{17n_{10}}{36} - \frac{n_{10}}{12} \\ & \leq \frac{1}{\sqrt{\pi n_{10}}} \cdot \frac{17n_{10}}{36} - \frac{n_{10}}{12} \\ & = \frac{1}{\sqrt{\pi}} \cdot \frac{17\sqrt{n_{10}}}{36} - \frac{n_{10}}{12}. \end{aligned}$$

This upper bound becomes negative for $n_{10} \geq 11$ (the real value a_0 that is being bounded from above already becomes negative for $n_{10} \geq 7$). The term a_1 is 0 for $n_{10} \leq 1$ and at most $1/64$ for $n_{10} \geq 2$. This can be seen by evaluating the above formula for $2 \leq n_{10} \leq 10$ (see Table 1) and bounding it above as follows for $n_{10} \geq 3$, using that $P(S \geq n_{10} + 1)$ is minimised for $n_{10} = 3$, where it is $5/16$.

$$\begin{aligned} & P(S = n_{10} + 1) \cdot \frac{17n_{10} + 13}{36} - P(S \geq n_{10} + 1) \cdot \frac{6n_{10} + 24}{36} \\ & \leq \frac{1}{\sqrt{\pi n_{10}}} \cdot \frac{17n_{10} + 13}{36} - \frac{5}{16} \cdot \frac{6n_{10} + 24}{36} \\ & = \frac{1}{\sqrt{\pi n_{10}}} \cdot \frac{17n_{10} + 13}{36} - \frac{5}{16} \cdot \frac{6n_{10} + 24}{36}, \end{aligned}$$

which is at most $n_{10}/64$ for $n_{10} \geq 11$. □

7.3 Selecting Different Parents of Unequal Fitness ($n_{10} > n_{01}$)

The case of P_{12} , different parents being selected, no common 0-bits flipping ($F_{00} = 0$) and parents having unequal fitness is easy to handle as in this scenario the potential drift is always non-positive. This is shown in the following lemma. It also states a closed formula for the potential drift as this will be used later on, in the proof of Lemma 22.

Table 1 Numerical values for coefficients a_0 and a_1 defined in the proof of Lemma 17, for different values of $n_{10} = n_{01}$

| n_{10} | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---------------|----------------|-----------------|------------------|--------------------|--------------------|-----------------------|-------------------------|-------------------------|---------------------------|
| a_0 | 0 | $\frac{1}{9}$ | $\frac{1}{8}$ | $\frac{11}{96}$ | $\frac{53}{576}$ | $\frac{95}{1536}$ | $\frac{27}{1024}$ | $-\frac{161}{12288}$ | $-\frac{1369}{24576}$ | $-\frac{13219}{131072}$ | $-\frac{349465}{2359296}$ |
| a_1 | 0 | 0 | $\frac{1}{72}$ | $\frac{1}{64}$ | $\frac{1}{128}$ | $-\frac{11}{1536}$ | $-\frac{85}{3072}$ | $-\frac{2585}{49152}$ | $-\frac{23927}{294912}$ | $-\frac{44317}{393216}$ | $-\frac{115495}{786432}$ |

Lemma 18 For all $n_{10} > n_{01}$ and all $F_{11} \in \mathbb{N}_0$,

$$\begin{aligned}
 & E(\Delta \mid P_{12}, F_{00} = 0, F_{11}) \\
 &= P(S > \ell_1) \frac{-2F_{11} - n_{10} + n_{01}}{3} + P(S = \ell_1) \frac{-F_{11} + n_{01}}{3} + P(S > \ell_2) \frac{-n_{01}}{6} \leq 0.
 \end{aligned}$$

The proof is not very insightful, hence it is placed in the appendix.

7.4 Combined Drift Bound When No Zero Flips

Assembling the previous drift bounds under various conditions, we get the following drift bounds.

Lemma 19 Assume $c \leq 56/9$, $n_{10} = n_{01}$ and $n_{11} \leq n - c$.

$$\begin{aligned}
 & E(\Delta \mid P_{11}, F_{0*} = 0) \cdot P(P_{11}, F_{0*} = 0) \\
 &+ E(\Delta \mid P_{22}, F_{*0} = 0) \cdot P(P_{22}, F_{*0} = 0) \\
 &+ E(\Delta \mid P_{12}, F_{00} = 0) \cdot P(P_{12}, F_{00} = 0) \\
 &\leq \begin{cases} \frac{cn_{10}^2}{9n} & \text{if } n_{10} = n_{01} \\ -\frac{n_{01}}{12} \left(1 - \frac{c}{n}\right)^n & \text{if } n_{10} > n_{01}. \end{cases}
 \end{aligned}$$

Proof For $n_{10} = n_{01}$, we argue that Lemma 17 implies

$$E(\Delta \mid P_{12}, F_{00} = 0) \leq \frac{n_{01}}{9} \left(1 - \frac{c}{n}\right)^{n_{11}}.$$

This is obvious for $n_{10} \leq 1$; for $n_{10} \geq 2$ it follows from $P(F_{11} = 1) = cn_{11}/n \cdot (1 - c/n)^{n-1} = cn_{11}/(n - c) \cdot (1 - c/n)^n \leq c(1 - c/n)^n$ and $1/8 + c/64 \leq 2/9 \leq n_{01}/9$ using $c \leq 56/9$.

By Lemmas 15 and 17, along with $P(P_{12}) = P(P_{11} \cup P_{22}) = 1/2$ and $P(F_{00} = 0) = (1 - c/n)^{n_{00}}$, the left-hand side is at most

$$\begin{aligned}
 & \frac{n_{10}}{18} \left(1 - \frac{c}{n}\right)^{n_{11}+n_{00}} - \frac{n_{10}}{18} \left(1 - \frac{c}{n}\right)^n \\
 &= \frac{n_{10}}{18} \left(1 - \frac{c}{n}\right)^{n_{11}+n_{00}} \left(1 - \left(1 - \frac{c}{n}\right)^{n_{10}+n_{01}}\right)
 \end{aligned}$$

$$\begin{aligned} &\leq \frac{n_{10}}{18} \left(1 - \frac{c}{n}\right)^{n_{11}+n_{00}} \left(1 - \left(1 - \frac{c(n_{10} + n_{01})}{n}\right)\right) \\ &= \frac{n_{10}}{18} \left(1 - \frac{c}{n}\right)^{n_{11}+n_{00}} \frac{c(n_{10} + n_{01})}{n} \leq \frac{cn_{10}^2}{9n}. \end{aligned}$$

The bound for the case $n_{10} > n_{01}$ follows immediately from Lemmas 15 and 18, along with $P(P_{11}) = 1/4$. □

8 Potential Drift When One Zero Flips

In this section we show that the drift is bounded by a term that yields the leading constant we are aiming for in our main result. Note that here we can afford to include error terms of lower order.

We first consider the case that two equal parents are selected.

Lemma 20 *For all populations with $n_{10} = n_{01}$,*

$$\begin{aligned} E(\Delta \mid P_{11}, F_{0*} = 1) &\leq P(F_{1*} = 0) \left(1 - \frac{n_{01}}{6}\right) + P(F_{1*} = 1) \left(\frac{2}{9} - \frac{n_{01}}{9}\right) \\ E(\Delta \mid P_{22}, F_{*0} = 1) &\leq P(F_{*1} = 0) \left(1 - \frac{n_{01}}{6}\right) + P(F_{*1} = 1) \left(\frac{2}{9} - \frac{n_{01}}{9}\right). \end{aligned}$$

For all populations with $n_{10} > n_{01}$,

$$\begin{aligned} E(\Delta \mid P_{11}, F_{0*} = 1) &\leq \begin{cases} \frac{1}{3} + \frac{2}{3} \cdot P(F_{1*} = 0) & \text{if } n_{01} = 0 \\ \left(1 - \frac{n_{01}}{3}\right) \cdot P(F_{1*} = 0) + \left(\frac{1}{3} - \frac{n_{01}}{3}\right) \cdot P(F_{1*} = 1) & \text{otherwise.} \end{cases} \\ E(\Delta \mid P_{22}, F_{*0} = 1) &\leq \frac{1}{3} \cdot P(F_{*1} = 0) + \frac{1}{6} \cdot P(F_{*1} = 1). \end{aligned}$$

Proof First assume $n_{10} = n_{01}$ and consider the event P_{11} . Given $F_{0*} = 1$, the offspring is accepted with certainty if no 1-bit is flipped. With probability $1/2$ the parent survives and then the new potential is at most $n_{11} + n_{10} + 1$. Consequently, the potential changes by at most $1 - n_{01}/3$ due to the loss of diversity. With the remaining probability $1/2$, the potential increases by at most 1. The overall expected change in potential in this case is thus at most $1 - n_{01}/6$. If a single 1-bit also flips together with the 0-bit, then the offspring has the same fitness as both parents and the individual to be removed is selected uniformly at random. If the offspring is removed, the potential does not change. If the parent is removed, the potential increases by at most $1/3$. If the other population member is removed, the potential increases by at most $1/3 - n_{10}/3$ due to the loss of diversity. The expected change in potential in this case is thus at most $2/9 - n_{10}/9$. If more than one 1-bits flip, then the offspring will have lower fitness than both other members and it will be rejected. Summing up the terms proves the first claim and the case of P_{22} follows analogously, considering F_{*0} and F_{*1} instead.

For $n_{10} > n_{01}$, given $F_{0*} = 1$ and that the fitter parent x^1 is selected twice, we consider separate cases according to the size of n_{01} . If $n_{01} = 0$, then no diversity can be lost. Thus, if no 1-bits flip, the potential increases by 1 because the new offspring has higher fitness than x^1 and x^2 is rejected. If at least one 1-bit flips, then the best fitness does not change and the potential increases by at most $1/3$. Overall, for the case when $n_{01} = 0$, the potential changes by $P(F_{1*} = 0) + 1/3 \cdot P(F_{1*} > 0) = 1/3 + 2/3 \cdot P(F_{1*} = 0)$. If $n_{01} > 0$ and no 1-bits flip, then the potential changes by at most $1 - n_{01}/3$ since the diversity is lost because x^2 is removed. If instead at least one 1-bit flips, then the potential changes by at most $1/3 - n_{01}/3$ since the best fitness does not change and the diversity may be lost if x^2 is removed. Since for $n_{01} > 0$ these terms are negative, and the offspring is accepted with probability 1 if $F_{1*} = 1$ the claim follows by summing up the two terms.

If x^2 is selected twice and no 1-bits are flipped, then the potential increases by at most $1/3$ (i.e., if an n_{00} bit is flipped) since the parent is removed and the diversity is kept. If a single 1-bit is flipped then the potential increases again by at most $1/3$. However, since the offspring has the same fitness as its parent, it is necessary that the parent is removed which happens with probability $1/2$. If more than one 1-bits are flipped, then the offspring is rejected. Summing up the terms completes the proof. \square

The proof of Lemma 20 has revealed a counterintuitive effect. A population of individuals with very different fitness values $f(x^1) \gg f(x^2)$ can have an advantage over a population where both members have the same fitness $f(x^1)$. This is because, conditioning on a 0-bit flipping, if the fitter parent is chosen twice, a near-arbitrary number of 1-bits can flip at the same time and the outcome may still be accepted. This increases the potential and explains why Lemma 20 contains an unexpectedly large potential drift in the case $n_{10} > n_{01} = 0$.

Now we consider the case P_{12} that two different parents are selected.

When $F_{11} \geq 2$ and the parents have equal fitness, the drift under P_{12} is non-positive for $n_{10} \geq 10$, except for $F_{11} = n_{10}-1$, where it is exponentially small in n_{10} . The proof of the following lemma is given in the appendix.

Lemma 21 For all $n_{10} = n_{01}$, $n_{10} \geq 10$ and all $2 \leq i \leq n_{10} - 1$,

$$E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i) \leq \begin{cases} 5n_{10}^3 2^{-2n_{10}}/3 & \text{if } i = n_{10} - 1 \\ 0 & \text{otherwise.} \end{cases}$$

When the two parents have unequal fitness, the following uniform upper bound on the potential drift applies.

Lemma 22 For all $n_{10} > n_{01}$ and all $F_{11} \in \mathbb{N}_0$,

$$E(\Delta \mid P_{12}, F_{00} = 1, F_{11}) \leq \frac{2}{3}.$$

Proof We start again with the drift formula from Lemma 14 and plug in $F_{00} = 1$. Then

$$E(\Delta \mid P_{12}, F_{00} = 1, F_{11}) = P(S > \ell_1) \cdot \frac{2 - 2F_{11} - n_{10} + n_{01}}{3}$$

$$\begin{aligned}
 &+ P(S = \ell_1) \cdot \frac{1 - F_{11} + n_{01}}{3} \\
 &+ P(S > \ell_2) \cdot \frac{2 - n_{01}}{6} \\
 &+ P(S = \ell_2) \cdot \frac{1}{6}.
 \end{aligned}$$

For $n_{10} < 12$ the claim is verified numerically for all pairs $n_{01} < n_{10}$, see Table 4. We now turn to an analytical argument for $n_{10} \geq 12$.

From Lemma 18 we already know that

$$P(S > \ell_1) \frac{-2F_{11} - n_{10} + n_{01}}{3} + P(S = \ell_1) \frac{-F_{11} + n_{01}}{3} + P(S > \ell_2) \frac{-n_{01}}{6} \leq 0$$

regardless of F_{11} . Hence,

$$\begin{aligned}
 E(\Delta \mid P_{12}, F_{00} = 1, F_{11}) &\leq P(S > \ell_1) \cdot \frac{2}{3} + P(S = \ell_1) \cdot \frac{1}{3} + P(S > \ell_2) \cdot \frac{1}{3} + P(S = \ell_2) \cdot \frac{1}{6} \\
 &\leq \frac{1}{3} + P(S > \ell_2) \cdot \frac{1}{3} + P(S = \ell_2) \cdot \frac{1}{6}
 \end{aligned}$$

since $P(S \geq \ell_1) \leq 1/2$ (using that ℓ_1 is strictly greater than the mean of S). Altogether,

$$E(\Delta \mid P_{12}, F_{00} = 1, F_{11}) \leq \frac{1}{3} + P(S \geq \ell_2) \frac{1}{3} \leq \frac{2}{3}$$

no matter how ℓ_1 and ℓ_2 turn out. □

To obtain a combined drift formula, we need to consider probabilities for flipping a certain number of 1-bits. The following simple lemma gives bounds for these.

Lemma 23 *For every mutation rate c/n and every i ,*

$$\left(\frac{n_{11} - i + 1}{n - c} \right)^i \cdot \frac{c^i}{i!} \left(1 - \frac{c}{n} \right)^{n_{11}} \leq P(F_{11} = i) \leq \left(\frac{n_{11}}{n - c} \right)^i \cdot \frac{c^i}{i!} \left(1 - \frac{c}{n} \right)^{n_{11}}.$$

Proof.

$$\begin{aligned}
 P(F_{11} = i) &= \binom{n_{11}}{i} \left(\frac{c}{n} \right)^i \left(1 - \frac{c}{n} \right)^{n_{11} - i} \\
 &= \binom{n_{11}}{i} \left(\frac{c}{n - c} \right)^i \left(1 - \frac{c}{n} \right)^{n_{11}} \\
 &= \left(\frac{n_{11}}{n - c} \cdot \frac{n_{11} - 1}{n - c} \cdots \frac{n_{11} - i + 1}{n - c} \right) \cdot \frac{c^i}{i!} \cdot \left(1 - \frac{c}{n} \right)^{n_{11}}.
 \end{aligned}$$

The first bracket is bounded as

$$\left(\frac{n_{11} - i + 1}{n}\right)^i \leq \left(\frac{n_{11}}{n - c} \cdot \frac{n_{11} - 1}{n - c} \cdots \frac{n_{11} - i + 1}{n - c}\right) \leq \left(\frac{n_{11}}{n - c}\right)^i. \quad \square$$

Using Lemmas 20 and 21 and considering the drift under P_{12} separately for $F_{11} \in \{0, 1\}$ and for all F_{11} when $n_{10} \leq 10$, we get:

Lemma 24 Assume $c \leq 2.71$, $n_{10} + n_{01} \leq \log^2 n$, $n_{11} \geq n - n/\log^3 n$ and $n_{00} \geq n_{10} \log n$.

$$\begin{aligned} & E(\Delta \mid P_{11}, F_{0*} = 1) \cdot P(P_{11}, F_{0*} = 1) \\ & + E(\Delta \mid P_{22}, F_{*0} = 1) \cdot P(P_{22}, F_{*0} = 1) \\ & + E(\Delta \mid P_{12}, F_{00} = 1) \cdot P(P_{12}, F_{00} = 1) \\ & \leq \begin{cases} \frac{c(2c+9)}{9e^c} \cdot \frac{n_{00}}{n} \cdot (1 + O(1/\log n)) & \text{if } n_{10} = n_{01} \\ \frac{c}{e^c} \left(\frac{1}{4} + \frac{c}{48} + \frac{e^c}{4} + O(1/\log n)\right) \frac{n_{00}}{n} & \text{if } n_{10} > n_{01} = 0 \\ \frac{c}{e^c} \left(\frac{1}{4} + \frac{c}{24} + \frac{e^c}{3} + O(1/n)\right) \frac{n_{00}}{n} & \text{if } n_{10} > n_{01} > 0. \end{cases} \end{aligned}$$

Proof We consider different cases.

Case $n_{10} = n_{01}$: In case both parents have equal fitness,

$$E(\Delta \mid P_{11}, F_{0*} = 1) \cdot P(P_{11}, F_{0*} = 1) = E(\Delta \mid P_{22}, F_{*0} = 1) \cdot P(P_{22}, F_{*0} = 1)$$

and we bound the sum of both terms, using $P(P_{11}) = P(P_{22}) = 1/4$ and Lemma 20 (noting that the formulas for P_{11} and P_{22} are identical) as

$$\frac{P(F_{0*} = 1)}{2} \left(P(F_{1*} = 0) \left(1 - \frac{n_{01}}{6}\right) + P(F_{1*} = 1) \left(\frac{2}{9} - \frac{n_{01}}{9}\right) \right). \quad (6)$$

For P_{12} we consider probabilities $P(F_{11} = \cdot)$, hence we aim to relate this with probabilities $P(F_{1*} = \cdot)$ as follows. (Note that, for $n_{10} = 0$, $F_{1*} = F_{11}$.)

$$P(F_{1*} = 0) = \left(1 - \frac{c}{n}\right)^{n_{11} + n_{10}} = P(F_{11} = 0) \left(1 - \frac{c}{n}\right)^{n_{10}}$$

and the term $\left(1 - \frac{c}{n}\right)^{n_{10}}$ is at least $1 - \frac{cn_{10}}{n}$ and at most 1. Similarly,

$$P(F_{1*} = 1) = \frac{c(n_{11} + n_{10})}{n} \left(1 - \frac{c}{n}\right)^{n_{11} + n_{10} - 1} = P(F_{11} = 1) \frac{n_{11} + n_{10}}{n_{11}} \left(1 - \frac{c}{n}\right)^{n_{10}}$$

and the term $\frac{n_{11} + n_{10}}{n_{11}} \left(1 - \frac{c}{n}\right)^{n_{10}}$ is at least $1 - \frac{cn_{10}}{n}$ and at most $1 + O(n_{10}/n)$. Likewise,

$$\frac{P(F_{0*} = 1)}{P(F_{00} = 1)} = \frac{n_{00} + n_{01}}{n_{00}} \left(1 - \frac{c}{n}\right)^{n_{01}},$$

which is at least $1 - cn_{01}/n$ and at most $1 + O(1/\log n)$ owing to the assumption $n_{00} \geq n_{10} \log n \geq n_{01} \log n$.

Thus, (6) can be written as

$$\frac{P(F_{00} = 1)}{2} \left(P(F_{11} = 0) \left(1 - \frac{n_{01}}{6} \right) + P(F_{11} = 1) \left(\frac{2}{9} - \frac{n_{01}}{9} \right) \right) + \xi \cdot P(F_{00} = 1) \tag{7}$$

for an error term $\xi \in [-O(1/\log n), +O(1/\log n)]$, regardless of the signs of the factors the terms $P(F_{1*} = 0)$ and $P(F_{1*} = 1)$ are multiplied with. Here we used that these factors are in $[-O(n_{10}), O(n_{10})]$ and the absolute value of the error from replacing each term $P(F_{1*} = \cdot)$ with $P(F_{11} = \cdot)$ is at most $O(n_{10}^2/n) = O(1/\log n)$.

The third term from the statement is

$$\begin{aligned} & E(\Delta \mid P_{12}, F_{00} = 1) \cdot P(P_{12}, F_{00} = 1) \\ &= \frac{P(F_{00} = 1)}{2} E(\Delta \mid P_{12}, F_{00} = 1) \\ &= \frac{P(F_{00} = 1)}{2} \sum_{i=0}^{n_{10}+1} E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i) P(F_{11} = i). \end{aligned} \tag{8}$$

We simplify the expectations in the above terms for $i \in \{0, 1\}$ using Lemma 13 and add them to the terms in (7). This yields an upper bound of

$$P(F_{00} = 1, F_{11} = 0) \cdot a_0 + P(F_{00} = 1, F_{11} = 1) \cdot a_1$$

for coefficients a_0, a_1 defined as follows:

$$\begin{aligned} a_0 &:= \frac{1}{2} \left(1 - \frac{n_{10}}{6} + P(S \geq n_{10} - 1) \left(1 - \frac{n_{10}}{6} \right) + P(S = n_{10} - 1) \cdot \frac{17n_{10} - 17}{36} \right) \\ a_1 &:= \frac{1}{2} \left(\frac{2}{9} - \frac{n_{10}}{9} + P(S \geq n_{10}) \left(\frac{1}{3} - \frac{n_{10}}{6} \right) + P(S = n_{10}) \cdot \frac{17n_{10} - 4}{36} \right). \end{aligned}$$

It is easy to verify that $a_0 = 1$ for $n_{10} = 0$ and $a_0 \leq 1$ for $1 \leq n_{10} \leq 9$. For $n_{10} \geq 10$, a_0 becomes negative since the term $P(S > n_{10} - 1) \left(1 - \frac{n_{10}}{6} \right)$ is non-positive and using $P(S = n_{10} - 1) \leq 1/(\sqrt{\pi n_{10}})$ according to Lemma 7, we get

$$a_0 \leq \frac{1}{2} \left(1 - \frac{n_{10}}{6} + \frac{1}{\sqrt{\pi n_{10}}} \cdot \frac{11n_{10} + 19}{36} \right) < 0.$$

Similarly, it is easy to check that $a_1 = 2/9$ for $n_{10} = 0$ and $a_1 \leq 2/9$ for $1 \leq n_{10} \leq 6$. For $n_{10} \geq 7$, a_1 becomes negative as

$$a_1 = \frac{1}{2} \left(\frac{2}{9} - \frac{n_{10}}{9} + P(S > n_{10}) \left(\frac{1}{3} - \frac{n_{10}}{6} \right) + P(S = n_{10}) \cdot \frac{11n_{10} + 8}{36} \right)$$

$$\leq \frac{1}{2} \left(\frac{2}{9} - \frac{n_{10}}{9} + \frac{1}{\sqrt{\pi n_{10}}} \cdot \frac{11n_{10} + 8}{36} \right) < 0.$$

By (8), for larger values of F_{11} , coefficients a_i for $P(F_{00} = 1, F_{11} = i)$, with $2 \leq i \leq n_{10} + 1$, can be bounded from considering only

$$a_i = \frac{E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i)}{2}.$$

Then we can bound the sought expression as

$$\begin{aligned} & E(\Delta \mid P_{11}, F_{0*} = 1) \cdot P(P_{11}, F_{0*} = 1) \\ & + E(\Delta \mid P_{22}, F_{*0} = 1) \cdot P(P_{22}, F_{*0} = 1) \\ & + E(\Delta \mid P_{12}, F_{00} = 1) \cdot P(P_{12}, F_{00} = 1) \\ & \leq \xi \cdot P(F_{00} = 1) + \sum_{i=0}^{n_{10}+1} a_i P(F_{00} = 1, F_{11} = i) \\ & = P(F_{00} = 1) \left(\xi + \sum_{i=0}^{n_{10}+1} a_i P(F_{11} = i) \right). \end{aligned}$$

We bound a_i by $\max\{0, a_i\}$ to ensure non-negative terms, which enables us to use the upper bound on $P(F_{11} = i)$ provided by Lemma 23:

$$P(F_{11} = i) \leq \binom{n_{11}}{n-c} \cdot \frac{c^i}{i!} \left(1 - \frac{c}{n}\right)^{n_{11}} \leq \frac{c^i}{i!} \left(1 - \frac{c}{n}\right)^{n_{11}}$$

where the last inequality follows from $n_{11} \leq n - c$.

Thus,

$$\begin{aligned} & P(F_{00} = 1) \left(\xi + \sum_{i=0}^{n_{10}+1} a_i P(F_{11} = i) \right) \\ & \leq P(F_{00} = 1) \left(\xi + \sum_{i=0}^{n_{10}+1} \max\{a_i, 0\} P(F_{11} = i) \right) \\ & \leq \frac{cn_{00}}{n} \left(1 - \frac{c}{n}\right)^{n_{00}-1} \left(\xi + \sum_{i=0}^{n_{10}+1} \max\{a_i, 0\} \cdot \frac{c^i}{i!} \cdot \left(1 - \frac{c}{n}\right)^{n_{11}} \right) \\ & = \frac{cn_{00}}{n} \left(1 - \frac{c}{n}\right)^{n_{00}+n_{11}-1} \left(\xi \left(1 - \frac{c}{n}\right)^{-n_{11}} + \sum_{i=0}^{n_{10}+1} \max\{a_i, 0\} \cdot \frac{c^i}{i!} \right). \end{aligned}$$

Using

$$\left(1 - \frac{c}{n}\right)^{n_{00}+n_{11}} = \left(1 - \frac{c}{n}\right)^{n-n_{10}-n_{01}} \leq e^{-c} \left(1 + \frac{c}{n-c}\right)^{n_{10}+n_{01}}$$

$$\leq e^{-c} \left(1 + \frac{2cn_{10}}{n - c} \right),$$

and recalling $\xi \leq O(1/\log n)$ we can write this as

$$\frac{cn_{00}}{e^{cn}} (1 + O(1/\log n)) \sum_{i=0}^{n_{10}+1} \max\{a_i, 0\} \cdot \frac{c^i}{i!}$$

assuming that $\max\{a_i, 0\} \cdot \frac{c^i}{i!} = \Omega(1)$ (if it is 0 or $o(1)$, we have proved the claim).

Now we obtain the claimed bound if we can show that

$$\sum_{i=0}^{n_{10}+1} \max\{a_i, 0\} \cdot \frac{c^i}{i!} \leq 1 + \frac{2}{9} \cdot c. \tag{9}$$

For $n_{10} = 0$ we have $a_0 = 1, a_1 = 2/9$, hence (9) holds with equality for $n_{10} = 0$. For $0 \leq n_{10} \leq 10$ we evaluate the coefficients $a_0, \dots, a_{n_{10}+1}$ with the closed formulas given above and verify (9) for $c \leq 2.71$. The coefficients and the ranges of c for which (9) holds are shown in Table 2. Since we assume $c \leq 2.71$, the inequality holds for all considered n_{10} .

For $n_{10} > 10$, we already established above that $a_0 \leq 0$ and $a_1 \leq 0$. By Lemma 21, $a_i \leq 0$ for all $2 \leq i \leq n_{10} + 1$, except for $i = n_{10} - 1$, where it is at most $5n_{10}^3 2^{-2n_{10}}/3$. For $n_{10} \geq 11$, this is at most 0.00053 and thus we get an upper drift bound of

$$P(F_{00} = 1) (\xi + 0.00053) = \frac{cn_{00}}{n} \cdot (\xi + 0.00053) \leq \frac{c(2c + 9)}{9e^c} \cdot \frac{n_{00}}{n} \cdot (1 + O(1/\log n))$$

where the last inequality holds for $c \leq 6.09$. This completes the proof for the case $n_{10} = n_{01}$.

Case $n_{10} > n_{01} = 0$: Now we consider unequal fitness, $n_{10} > n_{01}$, in the special case $n_{01} = 0$ (no diversity). By Lemma 20

$$\begin{aligned} & E(\Delta \mid P_{11}, F_{0*} = 1) \cdot P(P_{11}, F_{0*} = 1) \\ &= \frac{P(F_{0*} = 1)}{4} \cdot \left(\frac{1}{3} + \frac{2}{3} \cdot P(F_{1*} = 0) \right) \\ &= \frac{1}{12} \cdot P(F_{0*} = 1) + \frac{1}{6} \cdot P(F_{0*} = 1, F_{1*} = 0) \end{aligned}$$

and

Table 2 Values for coefficients a_i from the proof of Lemma 24 under P_{12} , $F_{00} = 1$ and $n_{10} = n_{01}$, rounded to 4 digits

| Condition on c | n_{10} | F_{11} | | | | | | | | | | | |
|------------------|----------|----------|---------|---------|---------|---------|---------|---------|---------|--------|--------|----------|----------|
| | | a_0 | a_1 | a_2 | a_3 | a_4 | a_5 | a_6 | a_7 | a_8 | a_9 | a_{10} | a_{11} |
| $c \leq \infty$ | 0 | 1.0000 | 0.2222 | | | | | | | | | | |
| $c \leq 2.71$ | 1 | 0.8333 | 0.2083 | 0.0556 | | | | | | | | | |
| $c \leq 3.08$ | 2 | 0.7048 | 0.1562 | 0.0903 | 0.0139 | | | | | | | | |
| $c \leq 3.54$ | 3 | 0.5833 | 0.0937 | 0.1042 | 0.0260 | 0.0035 | | | | | | | |
| $c \leq 4.08$ | 4 | 0.4641 | 0.0258 | 0.1046 | 0.0326 | 0.0074 | 0.0009 | | | | | | |
| $c \leq 4.47$ | 5 | 0.3460 | -0.0455 | 0.0957 | 0.0332 | 0.0098 | 0.0021 | 0.0002 | | | | | |
| $c \leq 4.93$ | 6 | 0.2282 | -0.1194 | 0.0799 | 0.0286 | 0.0097 | 0.0028 | 0.0006 | 0.0001 | | | | |
| $c \leq 5.85$ | 7 | 0.1106 | -0.1951 | 0.0588 | 0.0193 | 0.0070 | 0.0027 | 0.0008 | 0.0002 | 0.0000 | | | |
| $c \leq 8.36$ | 8 | -0.0069 | -0.2723 | 0.0334 | 0.0060 | 0.0014 | 0.0011 | 0.0007 | 0.0002 | 0.0000 | 0.0000 | | |
| $c \leq 12.53$ | 9 | -0.1246 | -0.3508 | 0.0045 | -0.0107 | -0.0068 | -0.0020 | -0.0001 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | |
| $c \leq \infty$ | 10 | -0.2423 | -0.4302 | -0.0274 | -0.0305 | -0.0176 | -0.0068 | -0.0017 | -0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

The first column shows sufficient conditions on c that imply $\sum_i \max\{a_i, 0\}c^i / (i!) \leq 1 + 2c/9$

$$\begin{aligned}
 & E(\Delta \mid P_{22}, F_{*0} = 1) \cdot P(P_{22}, F_{*0} = 1) \\
 &= \frac{P(F_{*0} = 1)}{4} \cdot \left(\frac{1}{3} \cdot P(F_{*1} = 0) + \frac{1}{6} \cdot P(F_{*1} = 1) \right) \\
 &= \frac{1}{12} \cdot P(F_{*0} = 1, F_{*1} = 0) + \frac{1}{24} \cdot P(F_{*0} = 1, F_{*1} = 1).
 \end{aligned}$$

Moving from events F_{0*}/F_{*0} and F_{1*}/F_{*1} to events F_{00} and F_{11} at the expense of an additive error term $\xi' \in [-O(1/\log n), +O(1/\log n)]$ as before and adding these two expressions shows that the first two lines of the formula from the statement are bounded by

$$\left(\frac{1}{12} + \xi' \right) P(F_{00} = 1) + \frac{1}{4} \cdot P(F_{00} = 1, F_{11} = 0) + \frac{1}{24} \cdot P(F_{00} = 1, F_{11} = 1). \tag{10}$$

Note that the first of these summands is multiplied by $P(F_{00} = 1)$, which does *not* include a statement on F_{11} .

Reusing the above calculations, the third term from the statement is

$$\begin{aligned}
 & E(\Delta \mid P_{12}, F_{00} = 1) \cdot P(P_{12}, F_{00} = 1) \\
 &= \frac{P(F_{00} = 1)}{2} \sum_{i=0}^{n_{10}+1} E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i) P(F_{11} = i).
 \end{aligned}$$

Using Lemma 14, the expectation simplifies to

$$\begin{aligned}
 & E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i) \\
 &= P(S > n_{10} + F_{11}) \cdot \frac{2 - 2F_{11} - n_{10}}{3} + P(S = n_{10} + F_{11}) \cdot \frac{1 - F_{11}}{3} \\
 &\quad + P(S > F_{11}) \cdot \frac{1}{3} + P(S = F_{11}) \cdot \frac{1}{6}.
 \end{aligned}$$

Since $n_{10} + n_{01} = n_{10} \leq n_{10} + F_{11}$, the event $S > n_{10} + F_{11}$ is impossible. The event $S = n_{10} + F_{11}$ is only possible for $F_{11} = 0$, where $P(S = n_{10} + F_{11}) = P(S = n_{10}) = P(S = 0)$ by symmetry of the binomial distribution. For $F_{11} = 0$ we thus get

$$\begin{aligned}
 E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = 0) &= P(S = n_{10}) \cdot \frac{1}{3} + P(S > 0) \cdot \frac{1}{3} + P(S = 0) \cdot \frac{1}{6} \\
 &= \frac{1}{3} + P(S = 0) \cdot \frac{1}{6}.
 \end{aligned}$$

For all $i > 0$, we get

$$E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i) = P(S > F_{11}) \cdot \frac{1}{3} + P(S = F_{11}) \cdot \frac{1}{6}.$$

Now we can bound the sought expression as

$$\begin{aligned} & E(\Delta \mid P_{11}, F_{0*} = 1) \cdot P(P_{11}, F_{0*} = 1) \\ & + E(\Delta \mid P_{22}, F_{*0} = 1) \cdot P(P_{22}, F_{*0} = 1) \\ & + E(\Delta \mid P_{12}, F_{00} = 1) \cdot P(P_{12}, F_{00} = 1) \\ & \leq P(F_{00} = 1) \left(\frac{1}{12} + \xi' + \frac{1}{2} \sum_{i=0}^{n_{10}+1} a_i P(F_{11} = i) \right) \end{aligned}$$

where, putting together (10) and the above bounds on $E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i)$,

$$\begin{aligned} a_0 &= \frac{7}{12} + P(S = 0) \cdot \frac{1}{6} \\ a_1 &= \frac{1}{24} + P(S > 1) \cdot \frac{1}{3} + P(S = 1) \cdot \frac{1}{6} \\ \forall i \geq 2: a_i &= P(S > i) \cdot \frac{1}{3} + P(S = i) \cdot \frac{1}{6}. \end{aligned}$$

Since all coefficients are positive, we can bound $P(F_{11} = i)$ as before and obtain a drift bound of

$$\frac{cn_{00}}{e^c n} (1 + O(1/\log n)) \left(\frac{1}{12} + \xi' + \frac{1}{2} \sum_{i=0}^{n_{10}+1} a_i \cdot \frac{c^i}{i!} \right).$$

Bounding $a_0 \leq 2/3$ (since $P(S = 0) \leq 1/2$), $a_1 \leq 1/24 + 1/3 = 3/8$ and $a_i \leq 1/3$ for $i \geq 2$, we have

$$\begin{aligned} \sum_{i=0}^{n_{10}+1} a_i \cdot \frac{c^i}{i!} &\leq \frac{2}{3} + \frac{3}{8} \cdot c + \frac{1}{3} \sum_{i=2}^{\infty} \frac{c^i}{i!} \\ &= \frac{2}{3} + \frac{3}{8} \cdot c + \frac{1}{3} \cdot (e^c - c - 1) \\ &= \frac{1}{3} + \frac{1}{24} \cdot c + \frac{1}{3} \cdot e^c. \end{aligned}$$

Plugging this in yields an upper drift bound of

$$\begin{aligned} & \frac{cn_{00}}{e^c n} (1 + O(1/\log n)) \left(\frac{1}{12} + \xi' + \frac{1}{6} + \frac{c}{48} + \frac{e^c}{6} \right) \\ & = \frac{cn_{00}}{e^c n} \left(\frac{1}{4} + \frac{c}{48} + \frac{e^c}{4} + O(1/\log n) \right). \end{aligned}$$

Case $n_{10} > n_{01} > 0$: Finally, we deal with the case $n_{10} > n_{01} > 0$. By Lemma 20

$$E(\Delta \mid P_{11}, F_{0*} = 1) \cdot P(P_{11}, F_{0*} = 1)$$

$$\begin{aligned}
 &= \frac{P(F_{0*} = 1)}{4} \left(\left(1 - \frac{n_{01}}{3}\right) \cdot P(F_{1*} = 0) + \left(\frac{1}{3} - \frac{n_{01}}{3}\right) \cdot P(F_{1*} = 1) \right) \\
 &\leq \frac{P(F_{0*} = 1)}{4} \left(1 - \frac{n_{01}}{3}\right) P(F_{1*} = 0) \\
 &\leq \frac{cn_{00}}{n} \left(1 - \frac{c}{n}\right)^{n-1} \left(\frac{1}{4} - \frac{n_{01}}{12}\right).
 \end{aligned}$$

and

$$\begin{aligned}
 &E(\Delta \mid P_{22}, F_{*0} = 1) \cdot P(P_{22}, F_{*0} = 1) \\
 &= \frac{P(F_{*0} = 1)}{4} \cdot \left(\frac{1}{3} \cdot P(F_{*1} = 0) + \frac{1}{6} \cdot P(F_{*1} = 1)\right) \\
 &= \frac{cn_{00}}{n} \left(\left(1 - \frac{c}{n}\right)^{n-1} \frac{1}{12} + \frac{c(n_{11} + n_{01})}{n} \left(1 - \frac{c}{n}\right)^{n-1} \frac{1}{24} \right) \\
 &\leq \frac{cn_{00}}{n} \left(\left(1 - \frac{c}{n}\right)^{n-1} \frac{1}{12} + \left(1 - \frac{c}{n}\right)^n \frac{c}{24} \right) \\
 &= \frac{cn_{00}}{n} \left(1 - \frac{c}{n}\right)^{n-1} \left(\frac{1}{12} + \frac{c}{24}\right).
 \end{aligned}$$

We use Lemma 22 to bound the third term in the statement as

$$\begin{aligned}
 &E(\Delta \mid P_{12}, F_{00} = 1) \cdot P(P_{12}, F_{00} = 1) \\
 &= \frac{P(F_{00} = 1)}{2} \cdot E(\Delta \mid P_{12}, F_{00} = 1) \\
 &\leq \frac{P(F_{00} = 1)}{3} \leq \frac{cn_{00}}{n} \cdot \frac{1}{3}.
 \end{aligned}$$

Together, we get a drift bound of

$$\begin{aligned}
 &\left(\frac{1}{3} + \left(1 - \frac{c}{n}\right)^{n-1} \left(\frac{1}{3} - \frac{n_{10}}{12} + \frac{c}{24}\right)\right) \frac{cn_{00}}{n} \\
 &\leq \left(\frac{e^c}{3} + \frac{1}{4} + \frac{c}{24} + O(1/n)\right) \frac{cn_{00}}{e^c n}.
 \end{aligned}$$

□

9 Putting Everything Together

Combining results from previous sections, for different numbers of flipping zeros, yields the following unconditional drift bound.

Lemma 25 *If $c \leq 1.422$, $n_{10} + n_{01} \leq \log^2 n$, $n_{11} \geq n - n/\log^3 n$ and $n_{00} \geq \log^5 n$, then*

$$E(\Delta) \leq \frac{c(2c + 9)}{9e^c} \cdot \frac{n_{00}}{n} \cdot (1 + O(1/\log n)).$$

Proof This follows from adding drift bounds from Lemmas 10, 19 and 24. For $n_{10} = n_{01}$, Lemma 24 gives the stated bound. The terms $O(n_{10}^2/n) = O((\log^4 n)/n)$ from Lemma 19 and $O(n_{0*}/(n \log n)) = O(n_{00}/(n \log n))$ can both be absorbed in the $O(1/\log n)$ term since $n_{00}/n \geq \log^5 n/n$.

For $n_{10} > n_{01} = 0$, the bound from Lemma 24 is at most the bound from the same lemma for $n_{10} = n_{01}$ since

$$\frac{c}{e^c} \left(\frac{1}{4} + \frac{c}{48} + \frac{e^c}{4} \right) \leq \frac{c(2c + 9)}{9e^c}$$

for $c \leq 1.422$. Then the claim follows as above.

For $n_{10} > n_{01} > 0$ note that Lemma 19 yields a negative upper bound of $-n_{01}/8 \cdot (1 - c/n)^n = -\Omega(1)$. Since $n_{00} \leq n/\log^3 n$, we obtain a negative drift bound if n is large enough. □

To translate the upper bound from Lemma 25 into a lower bound on the expected runtime, we use the lower-bound version of the multiplicative drift theorem from Theorem 4. This theorem requires an upper bound on the drift of the potential function and a sufficiently small probability for large jumps of this value. Such large jumps can occur if the two individuals of the (2+1) GA have a large Hamming distance. Recall that Lemma 9 shows this to be unlikely.

The following lemma shows that a drift of the potential can be translated into a lower bound on the expected optimisation time. This is not immediate since the potential function is a weighted combination of two quantities.

Lemma 26 *Let N_t denote the number of n_{00} -bits at time t of the (2+1) GA and T the first point in time where $n_{00} \leq \log^5 n$. Assume that $n_{01} + n_{10} \leq \log^2 n$ for all points in time before T , and $N_0 \geq \log^5 n$. If $E(\varphi_{t+1} - \varphi_t \mid \varphi_t) \leq \delta N_t$ for some $n^{-O(1)} \leq \delta < 1$ and all $t < T$, then*

$$E(T \mid \varphi_0) \geq (1 - O(1/\log n)) \frac{\ln(n) - O(\ln \ln n)}{\delta}.$$

Proof We introduce a distance function $\bar{\varphi}_t := n - \varphi_t = n_{00} + \frac{2}{3}n_{01}$ as the mirror image of our potential to obtain a function to be minimised, as required in Theorem 4. Moreover, we write $\Delta_t = \bar{\varphi}_t - \bar{\varphi}_{t+1}$. The key idea is to show that the statement on $E(T \mid \varphi_0)$ holds under the slightly different drift condition

$$E(\Delta_t \mid \bar{\varphi}_t) \leq \delta \bar{\varphi}_t \tag{11}$$

and then to prove that the actual drift condition $E(\varphi_{t+1} - \varphi_t \mid \varphi_t) \leq \delta N_t$ leads to the same result, up to lower-order terms.

We consider the process from the first point in time where $\bar{\varphi}_t \leq n/\log^3 n$, assume (11) to hold for all $t < T$ and estimate the remaining expected time to minimise the distance $\bar{\varphi}_t$. Lemma 9 and the facts that at most $\log n$ bits flip per generation and that the distance does not drop below $n/\log^3 n$ within the first $\log^2 n$ generations (each happening with probability $1 - n^{-\Omega(\log n)}$), imply that $\bar{\varphi}_0 \geq n/(2 \log^3 n)$ with respect to our time count. We assume this to happen. By Lemma 9, with probability $1 - n^{-\Omega(\log n)}$ it holds that $n_{10} + n_{01} \leq \log^2 n$ for any polynomial number of steps. Assuming this for a sufficiently long period obtained from applying Markov's inequality on $E(T \mid \bar{\varphi}_0)$, our assumptions only change the bound on the expected value by a $1 - n^{-\Omega(\log n)}$ factor.

Clearly, since $n_{10} + n_{01} \leq \log^2 n$, crossover followed by a neutral mutation can change the $\bar{\varphi}$ -value by at most $\log^2 n$. Moreover, each mutation flips k or more bits with probability at most

$$\binom{n}{k} \left(\frac{c}{n}\right)^k \leq \frac{c^k}{k!}; \tag{12}$$

in particular it flips at most $\log^2 n$ bits with probability $1 - n^{-\Omega(\log n)}$. Adding up these effects, we arrive at $P(\bar{\varphi}_t - \bar{\varphi}_{t+1} \geq 2 \log^2 n) = n^{-\Omega(\log n)}$. The time to minimise the distance function is no smaller than the time to reach a distance of at most $x_{\min} := \log^5 n$ (we stop at this point to fulfill the condition on n_{00} in Lemma 25). Along with $\beta := 2/\log n$ and using $X_t := \bar{\varphi}_t$, we verify the second condition of Theorem 4 by estimating

$$\begin{aligned} P(X_t - X_{t+1} \geq \beta X_t) &\leq \\ P(X_t - X_{t+1} \geq \beta x_{\min}) &= P(X_t - X_{t+1} \geq 2 \log^2 n) = n^{-\Omega(\log n)}. \end{aligned}$$

Finally, we estimate this bound as $n^{-\Omega(n)} \leq \beta\delta/\log n \leq \beta\delta/\log(X_t)$ since both β and δ are at least inversely polynomial in n . Hence, $P(X_t - X_{t+1} \geq \beta X_t) \leq \beta\delta/\log(X_t)$, which satisfies the condition. Applying the theorem and recalling our assumption $\bar{\varphi}_0 \geq n/(2 \log n)$,

$$E(T \mid X_0) \geq \frac{1 - \beta}{1 + \beta} \cdot \frac{\ln(\bar{\varphi}_0/x_{\min})}{\delta} = (1 - O(1/\log n)) \frac{\ln(n/\log^4 n)}{\delta}$$

which is $(1 - O(1/\log n)) \frac{\ln(n) - O(\ln \ln n)}{\delta}$. Recall that the unconditional expected time is only by a factor $1 - n^{-\Omega(\log n)}$ smaller.

Finally, we relate N_t to the distance function $\bar{\varphi}_t$ and note first that $\bar{\varphi}_t \geq N_t$. Since, for $t < T$, our assumptions imply $\bar{\varphi}_t = n_{00} + (2/3)n_{01} \leq (1 + 1/\log n)N_t$, the prerequisite

$$E(\varphi_{t+1} - \varphi_t \mid \varphi_t) \leq \delta N_t$$

along with the fact $\Delta_t = \varphi_{t+1} - \varphi_t$ imply

$$E(\Delta_t \mid \bar{\varphi}_t) \leq \delta(1 - 1/\log n)\bar{\varphi}_t.$$

Hence, (11) has been established with parameter $\delta(1 - 1/\log n)$ so that

$$E(T \mid \varphi_0) \geq (1 - O(1/\log n)) \frac{\ln(n/\log^4 n)}{(1 - 1/\log n)\delta},$$

which is again $(1 - O(1/\log n)) \frac{\ln(n/\log^4 n)}{\delta}$ as claimed. □

We are now ready to prove our main result.

Proof of Theorem 2 We apply drift analysis to a “typical” run, that is, a run where events mentioned in the following that happen with overwhelming probability do occur. The contrary is called a failure event and if a failure occurs, we pessimistically assume that the runtime is 0. By the law of total probability, the expected runtime is bounded from below by the expected runtime in a typical run, multiplied by the probability that no failure occurs.

With probability $1 - 2^{-\Omega(n)}$ the initial population has a maximum number of $(2/3)n$ one-bits. The random number of one-bits that crossover creates among the relevant n_{10} and n_{01} bits follows a binomial distribution with parameters $n_{10} + n_{01}$ and $1/2$. The expected value of this number equals $(n_{10} + n_{01})/2 \leq n/2$, and using Chernoff bounds (Theorem 5) with $\delta = 2n^{-1/3}$ and the upper bound $n/2$ on the expectation, the number of one-bits of every fixed offspring is at most $U := (n_{10} + n_{01})/2 + n^{2/3}$ with probability $1 - 2^{-\Omega(n^{1/3})}$. By a union bound, the probability that at least one offspring has more than U one-bits is still $2^{-\Omega(n^{1/3})}$. Since the fittest parent has at least $(n_{10} + n_{01})/2$ one-bits, this means that the number of one-bits at the relevant positions, and thereby the number of one-bits of the fitter offspring, grows by at most $n^{2/3}$ with probability $1 - 2^{-\Omega(n^{1/3})}$. The same holds for each mutation with constant rate c as seen in (12). Hence, the subsequent $2 \log^2 n$ generations do not increase the ONEMAX-value to more than $(3/4)n$ with probability $1 - 2^{-\Omega(n^{1/3})}$.

We consider the first point in time when $n_{11} \geq n - n/\log^3 n$. By the same arguments as before, we then have $n_{11} \leq n - n/(2 \log^3 n)$ with overwhelming probability. Now Lemma 9 is in force, implying that we can apply Lemma 26 with

$$\delta \leq \frac{c(2c + 9)}{9e^c n} \cdot (1 + O(1/\log n)),$$

according to the drift bound from Lemma 25. Hence, if none of the above-mentioned failure events occur, the expected optimization time is bounded from below by

$$\left(1 - O\left(\frac{1}{\log n}\right)\right) \frac{9e^c n \ln(n) - O(n \ln \ln n)}{c(2c + 9)},$$

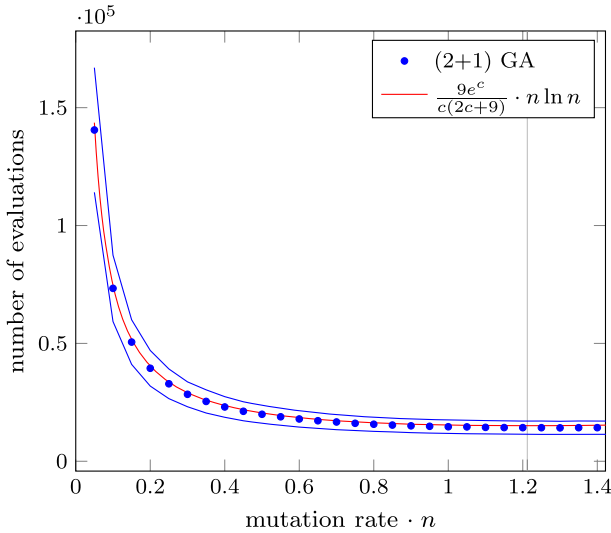


Fig. 1 Average optimisation times over 10,000 runs of the (2+1) GA on OneMax with $n = 1000$ bits. The mutation rate is set to c/n with $c \in \{0.05, 0.1, \dots, 1.4\}$. The thin blue lines show mean \pm standard deviation (Color figure online)

which is $\frac{9e^c}{c(2c+9)} \cdot n \ln(n) - O(n \ln \ln n)$ as claimed. This still holds when multiplying the above with the probability of no failure event occurring, as the union of all failure events is superpolynomially small.

10 Experiments

Our lower bound is tight up to lower-order terms. We ran experiments to see how close the results are to the dominant term of $\frac{9e^c}{c(2c+9)} \cdot n \ln n$, when varying c and n .

We first fixed $n = 1000$ and varied c in steps of 0.1. For each value of c , we ran the (2+1) GA 10000 times to obtain smooth curves. For the range of c -values, $c \in (0, 1.422]$, covered by Theorem 2, Fig. 1 shows that the empirical averages are very close to the dominant term. The vertical line indicates the c -value of $\frac{\sqrt{97}-5}{4} \approx 1.2122$ that was determined to be optimal (up to small-order terms) in Theorem 3.

An obvious question arises: what if values of c are chosen that are larger than the limit 1.422 in Theorem 2? Fig. 2 shows results of the previous experiment for a larger range of c -values, $c \in \{0.1, 0.2, 0.3, \dots, 5.0\}$. The first vertical line in Fig. 2 again indicates the c -value of $\frac{\sqrt{97}-5}{4} \approx 1.2122$; the second vertical line indicates the limit of 1.422 from Theorem 2.

We see that as c increases beyond 1.422, a gap starts to appear between the dominant term and the average runtime. The average runtime seems to be smaller than $\frac{9e^c}{c(2c+9)} \cdot n \ln n$. This could indicate that the lower bound does not hold for $c > 1.422$, or that there are small-order terms that affect the plots for $n = 1000$.

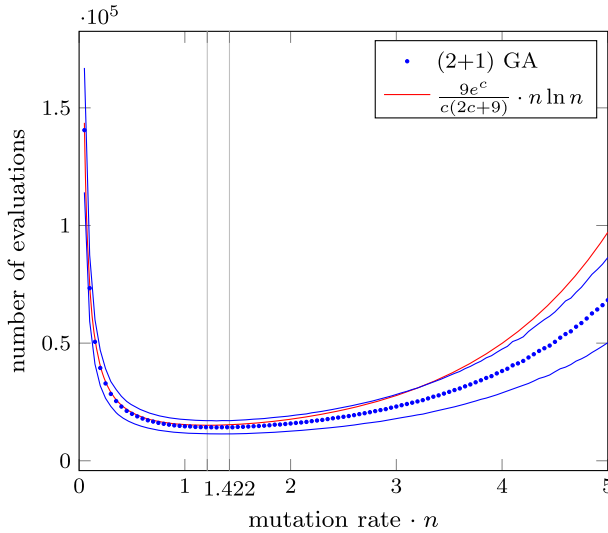


Fig. 2 Average optimisation times over 10,000 runs of the (2+1) GA on OneMax with $n = 1000$ bits. The mutation rate is set to c/n with $c \in \{0.1, 0.2, \dots, 5.0\}$. The thin blue lines show mean \pm standard deviation (Color figure online)

To see how quickly the runtime approaches the dominant term of $\frac{9e^c}{c(2c+9)} \cdot n \ln n$ as n grows, we ran experiments with n increasing exponentially in powers of 2 from $n = 2^3 = 8$ to $n = 2^{13} = 8192$. Figure 3 shows the normalised average runtimes, which is the runtime divided by $n \ln n$. One can see that, for both the default and the optimal mutation rates, $c = 1$ and $c = \frac{\sqrt{97}-5}{4}$, respectively, the curves approach their respective dominant terms. However, the approach is quite slow. Even for $n = 8192$ there is still a significant gap to the leading constant of the dominant term. This might indicate that the average runtime includes a significant, negative small-order term.

Recall that our lower bound (Theorem 2) includes a negative term of $-O(n \log \log n)$, while the upper bound (Theorem 1) contains an additive term $+O(n)$. The term $-O(n \log \log n)$ is an artefact of the fact that we only considered around $n/\log^3 n$ fitness levels and we excluded the most difficult $\log^5 n$ ones. It is easy to show that optimising these fitness levels takes expected time $O(n \log \log n)$ for both (1+1) EA and (2+1) GA. We conjecture that the expected running time is close to $\frac{9e^c}{c(2c+9)} \cdot n \ln(n) + bn$ for some negative constant b . This is based on related work on the (1+1) EA whose expected runtime on ONEMAX is known to be $en \ln(n) - 1.8925 \dots n + O(\log n)$ (cf. [14], who even determine the expected runtime up to additive errors of $O(\log n/n)$). Intuitively, the negative linear term appears in the bound for the (1+1) EA since the algorithm does not start with the largest possible value of the underlying potential function, which is the number of one-bits. More precisely, the initial number of one-bits X_0 has an expected value of $n/2$ and the multiplicative drift theorems both for lower and upper bounds involve the term $\ln(X_0/x_{\min})$. For $X_0 \approx n/2$ this becomes roughly $\ln(n) - \ln 2$, so that the crucial term $\ln(X_0/x_{\min})/\delta$, where $\delta = \Theta(1/n)$, becomes at most $(\ln n)/\delta - (\ln 2)/\delta = \Theta(n \ln n) - O(n)$. The expected initial value

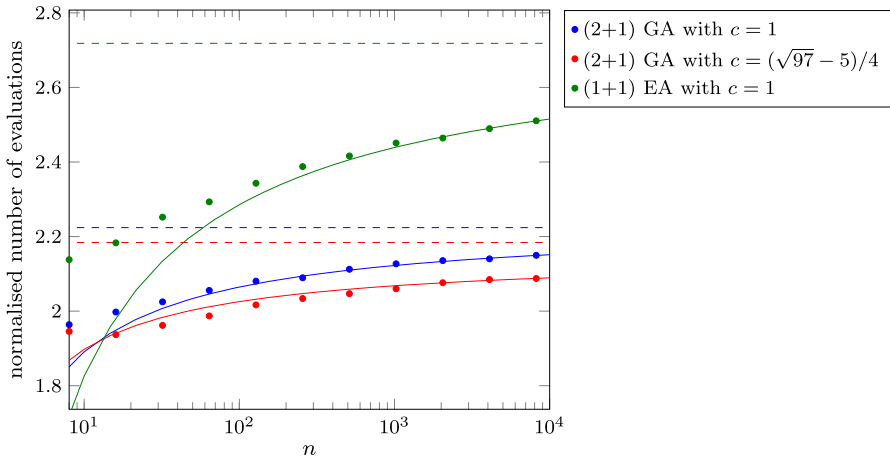


Fig. 3 Average optimisation times over 10,000 runs of the (2+1) GA and (1+1) EA on OneMax for different values of n and different mutation rates, normalised by dividing by $n \ln n$. The solid lines show the best fit obtained by a non-linear regression fit to a model of $an \ln(n) + bn$, using the non-normalised running times. The dashed lines show the leading constants provided by theoretical results

of our potential function for the (2+1) GA equals $(1/4 + 1/4 + 1/12)n = (7/12)n$, so we believe there is a non-negligible negative term in the expression for the expected running as well.

To see how closely the empirical data matches a conjectured bound with a negative linear term, we performed a non-linear regression to fit the average runtimes to a model of $an \ln(n) + bn$, for parameters a, b . We used the software R, version 4.0.3, calling the `nls` command with starting values of $a = 1$ and $b = 0$. The results of the regression were values of a, b that minimised the residual sum of squares.

For the (2+1) GA with $c = 1$, this resulted in a fitted function of $2.2381n \ln(n) - 0.7995n$. The leading constant 2.2381 is close to the value 2.224 from our theoretical analysis. Likewise, for the optimal value $c = \frac{\sqrt{97}-5}{4} \approx 1.2122$, the best fit is for $2.1532n \ln(n) - 0.5883n$ and the leading constant is close to 2.18417 from our analysis and smaller than the leading constant for $c = 1$. In both cases, the linear term has a negative sign. For comparison, the same regression for the (1+1) EA returned the best fit for the function $2.745n \ln(n) - 2.116n$. The leading constant is close to the theoretically proven one of $e = 2.71828$ and the coefficient of the linear term is also close to the theoretical one of $-1.8925 \dots$ [14].

11 Conclusions

Proving lower bounds for crossover-based GAs is a notoriously hard problem. We have provided such a lower bound for the (2+1) GA on ONEMAX through a careful analysis of a potential function that captures both the current best fitness and the potential for finding improvements through crossover combining different “building blocks” of good solutions. Our lower bound is tight up to small-order terms. This for

the first time proves rigorously that populations are beneficial for standard steady-state genetic algorithms. We also identified the optimal mutation rate for the (2+1) GA as $(\sqrt{97} - 5)/(4n)$ for the considered range of mutation rates c/n .

Our lower bound applies for $c \leq 1.422$ and an obvious open question is whether the leading constant in the expected runtime remains at $9e^c / (c(2c+9))$ when this threshold is exceeded. Our empirical results suggest that the expected runtime is smaller than the stated bound, albeit it is not clear whether these results are affected by negative small-order terms. We conjecture that the runtime is very close to $\frac{9e^c}{c(2c+9)} \cdot n \ln(n) + bn$, where bn is a linear term with a negative constant b . Our drift estimates subsumed errors that were by a factor of $O(1/\log n)$ smaller than the dominant term in asymptotic notation, hence tighter analyses would be required to obtain rigorous bounds on the conjectured linear term. Another avenue for future work would be to simplify our approach, possibly by exploiting that states with $n_{10} > 1$ are rarely reached in the late stages of a run, or to generalise our analysis to larger parent population sizes.

Acknowledgements This work was initiated at the Dagstuhl Seminar 19431 and supported by the EPSRC under Grant EP/M004252/1.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Omitted Proofs

In this appendix we present proofs omitted from the main part.

A.1 Proof of Lemma 12

Recall that Lemma 12 simplified the expected surplus of a binomially distributed random variable. To prove Lemma 12, we use an equation that is attributed to Gruder [13]. We therefore refer to it as *Gruder’s equation*.

Lemma 27 (Gruder’s equation [13]) *For every $0 \leq p \leq 1$ and all integers a, b ,*

$$\sum_{i=a}^b \binom{b}{i} p^i (1-p)^{b-i} (i-bp) = a \binom{b}{a} p^a (1-p)^{b-a+1}.$$

Proof. The claim is trivial for $a > b$, hence we assume $a \leq b$. The following proof is stated in [17], attributed to Gruder [13].

$$\begin{aligned} \sum_{i=a}^b \binom{b}{i} p^i (1-p)^{b-i} (i-bp) &= \sum_{i=a}^b \binom{b}{i} p^i (1-p)^{b-i} (i(1-p) - (b-i)p) \\ &= \sum_{i=a}^b \binom{b}{i} \left(b \binom{b-a}{i-1} p^i (1-p)^{b-i+1} \right. \\ &\quad \left. - b \binom{b-1}{i} p^{i+1} (1-p)^{b-i} \right) \\ &= b \binom{b-1}{a-1} p^a (1-p)^{b-a+1} \\ &= a \binom{b}{a} p^a (1-p)^{b-a+1}. \quad \square \end{aligned}$$

We use Gruder’s equation to prove Lemma 12.

Proof of Lemma 12

$$\begin{aligned} E(S \mid S \geq \ell) P(S \geq \ell) &= \sum_{s=\ell}^n s \cdot P(S = s) \\ &= \sum_{s=\ell}^n \left(s - \frac{n}{2} \right) \cdot P(S = s) + \frac{n}{2} \cdot P(S \geq \ell) \\ &= \sum_{s=\ell}^n \binom{n}{s} 2^{-n_{10}-n_{01}} \cdot \left(s - \frac{n}{2} \right) + \frac{n}{2} \cdot P(S \geq \ell). \end{aligned}$$

Applying Gruder’s equation (Lemma 27) with $p = 1/2$, $a = \ell$, $b = n$ yields

$$\begin{aligned} &= \ell \binom{n}{\ell} 2^{-n-1} + \frac{n}{2} \cdot P(S \geq \ell) \\ &= \frac{\ell}{2} \cdot P(S = \ell) + \frac{n}{2} \cdot P(S \geq \ell). \end{aligned}$$

A.2 Proof of Lemma 16

Recall that Lemma 16 claims that if $F_{00} = 0$ and $F_{11} \geq 2$ then under P_{12} the potential drift is non-positive.

Proof of Lemma 16 By Lemma 13,

$$\begin{aligned} &E(\Delta \mid P_{12}, F_{00} = 0, F_{11} = i) \\ &= \left(P(S = n_{10} + i) \cdot \frac{17n_{10} + 13i}{36} - P(S \geq n_{10} + i) \cdot \frac{6n_{10} + 24i}{36} \right). \end{aligned}$$

The above is 0 for $i \geq n_{10}$, hence the claim is trivial for $n_{10} \leq 2$. We assume $n_{10} \geq 3$ in the following.

The claim is equivalent to

$$P(S = n_{10} + i) \cdot \frac{17n_{10} + 13i}{36} \leq P(S \geq n_{10} + i) \cdot \frac{6n_{10} + 24i}{36}.$$

Multiplying by $36 \cdot 2^{2n_{10}}$ and spelling out the binomial distributions involving S yields

$$\binom{2n_{10}}{n_{10} + i} \cdot (17n_{10} + 13i) \leq \sum_{j=i}^{n_{10}} \binom{2n_{10}}{n_{10} + j} \cdot (6n_{10} + 24i).$$

Dividing both sides by $\binom{2n_{10}}{n_{10}+i}$ yields

$$(17n_{10} + 13i) \leq \sum_{j=i}^{n_{10}} \frac{(n_{10} + i)!(n_{10} - i)!}{(n_{10} + j)!(n_{10} - j)!} \cdot (6n_{10} + 24i).$$

Note that this is equivalent to

$$(11n_{10} - 11i) \leq \sum_{j=i+1}^{n_{10}} \frac{(n_{10} + i)!(n_{10} - i)!}{(n_{10} + j)!(n_{10} - j)!} \cdot (6n_{10} + 24i)$$

which holds trivially for $i = n_{10}$. For $i = n_{10} - 1$, the above simplifies to

$$11 \leq \frac{(2n_{10} - 1)!}{(2n_{10})!} \cdot (6n_{10} + 24n_{10} - 24) \Leftrightarrow 11 \leq 15 - \frac{12}{n_{10}}$$

which is true for $n_{10} \geq 3$. For $i \leq n_{10} - 2$, we have

$$\begin{aligned} & \sum_{j=i+1}^{n_{10}} \frac{(n_{10} + i)!(n_{10} - i)!}{(n_{10} + j)!(n_{10} - j)!} \cdot (6n_{10} + 24i) \\ & \geq \sum_{j=i+1}^{i+2} \frac{(n_{10} + i)!(n_{10} - i)!}{(n_{10} + j)!(n_{10} - j)!} \cdot (6n_{10} + 24i) \\ & = \left(\frac{(n_{10} + i)!(n_{10} - i)!}{(n_{10} + i + 1)!(n_{10} - i - 1)!} + \frac{(n_{10} + i)!(n_{10} - i)!}{(n_{10} + i + 2)!(n_{10} - i - 2)!} \right) \cdot (6n_{10} + 24i) \\ & = \left(\frac{n_{10} - i}{n_{10} + i + 1} + \frac{(n_{10} - i)(n_{10} - i - 1)}{(n_{10} + i + 1)(n_{10} + i + 2)} \right) \cdot (6n_{10} + 24i) \\ & = (n_{10} - i) \left(\frac{1}{n_{10} + i + 1} + \frac{n_{10} - i - 1}{(n_{10} + i + 1)(n_{10} + i + 2)} \right) \cdot (6n_{10} + 24i) \\ & = (n_{10} - i) \left(\frac{2n_{10} + 1}{(n_{10} + i + 1)(n_{10} + i + 2)} \right) \cdot (6n_{10} + 24i) \end{aligned}$$

$$= 6(n_{10} - i) \cdot \frac{(2n_{10} + 1)(n_{10} + 4i)}{(n_{10} + i + 1)(n_{10} + i + 2)}.$$

Now the claim follows if we can show that the above fraction is at least $11/6$.

$$\begin{aligned} & \frac{(2n_{10} + 1)(n_{10} + 4i)}{(n_{10} + i + 1)(n_{10} + i + 2)} \\ &= \frac{(n_{10} + i + 1)(n_{10} + i + 2) + (n_{10} - i)(n_{10} + i + 2) + (3i - 2)(2n_{10} + 1)}{(n_{10} + i + 1)(n_{10} + i + 2)} \\ &= 1 + \frac{(n_{10} - i)(n_{10} + i + 2) + (3i - 2)(2n_{10} + 1)}{(n_{10} + i + 1)(n_{10} + i + 2)} \\ &\geq 1 + \frac{(n_{10} - i)(n_{10} + i + 2) + (3i - 2)(n_{10} + i + 2)}{(n_{10} + i + 1)(n_{10} + i + 2)} \\ &= 1 + \frac{n_{10} + 2i - 2}{n_{10} + i + 1} \\ &\geq 1 + \frac{n_{10} + 2}{n_{10} + 3} \\ &\geq 1 + \frac{5}{6} = \frac{11}{6}, \end{aligned}$$

where in the last two inequalities we have used $i \geq 2$ and $n_{10} \geq 3$.

A.3 Proof of Lemma 18

Proof of Lemma 18 We investigate the formula

$$\begin{aligned} E(\Delta \mid P_{12}, F_{00} = 0, F_{11}) &= P(S > \ell_1) \cdot \frac{-2F_{11} - n_{10} + n_{01}}{3} \\ &\quad + P(S = \ell_1) \cdot \frac{-F_{11} + n_{01}}{3} \\ &\quad + P(S > \ell_2) \cdot \frac{-n_{01}}{6}. \end{aligned} \tag{13}$$

that follows from Lemma 14 for $F_{00} = 0$. For $n_{10} < 12$ the claim is verified numerically for all pairs $n_{01} < n_{10}$, see Table 3. We now turn to an analytical argument for $n_{10} \geq 12$.

Since the first term of the rhs. in (13) is clearly negative, we omit it in the following. We use our assumption $n_{10} \geq 12$. The aim is to compare different probabilities occurring in (13) to each other. More precisely, we want to show the claim that $P(S > \ell_2) \geq 2P(S = \ell_1)$. Along with the fact that the third term is clearly negative, we then have

$$E(\Delta \mid P_{12}, F_{00} = 0, F_{11}) \leq P(S = \ell_1) \left(\frac{-F_{11} + n_{01}}{3} \right)$$

Table 3 Numerical values for the potential drift given P_{12} , $n_{10} > n_{01}$ and $F_{00} = F_{11} = 0$, rounded to 4 digits

| n_{10} | n_{01} | | | | | | | | | | |
|----------|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 0.0000 | | | | | | | | | | |
| 2 | 0.0000 | 0.0000 | | | | | | | | | |
| 3 | 0.0000 | -0.0729 | -0.0208 | | | | | | | | |
| 4 | 0.0000 | -0.1146 | -0.1354 | -0.0521 | | | | | | | |
| 5 | 0.0000 | -0.1380 | -0.2109 | -0.1960 | -0.0898 | | | | | | |
| 6 | 0.0000 | -0.1510 | -0.2591 | -0.2988 | -0.2565 | -0.1322 | | | | | |
| 7 | 0.0000 | -0.1582 | -0.2891 | -0.3698 | -0.3822 | -0.3176 | -0.1779 | | | | |
| 8 | 0.0000 | -0.1621 | -0.3073 | -0.4173 | -0.4736 | -0.4628 | -0.3795 | -0.2263 | | | |
| 9 | 0.0000 | -0.1642 | -0.3182 | -0.4484 | -0.5382 | -0.5727 | -0.5418 | -0.4421 | -0.2769 | | |
| 10 | 0.0000 | -0.1654 | -0.3246 | -0.4682 | -0.5827 | -0.6536 | -0.6684 | -0.6197 | -0.5054 | -0.3293 | |
| 11 | 0.0000 | -0.1660 | -0.3284 | -0.4807 | -0.6127 | -0.7115 | -0.7646 | -0.7616 | -0.6968 | -0.5694 | -0.3833 |

Table 4 Numerical values for the potential drift given P_{12} , $n_{10} > n_{01}$ and $F_{00} = 1$, $F_{11} = 0$, rounded to 4 digits

| n_{10} | n_{01} | | | | | | | | | | | |
|----------|----------|--------|--------|---------|---------|---------|---------|---------|---------|---------|---------|--|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 1 | 0.6667 | | | | | | | | | | | |
| 2 | 0.5000 | 0.5833 | | | | | | | | | | |
| 3 | 0.4167 | 0.4167 | 0.5052 | | | | | | | | | |
| 4 | 0.3750 | 0.3125 | 0.3281 | 0.4297 | | | | | | | | |
| 5 | 0.3542 | 0.2500 | 0.2070 | 0.2402 | 0.3555 | | | | | | | |
| 6 | 0.3437 | 0.2135 | 0.1276 | 0.1035 | 0.1536 | 0.2819 | | | | | | |
| 7 | 0.3385 | 0.1927 | 0.0771 | 0.0086 | 0.0024 | 0.0682 | 0.2086 | | | | | |
| 8 | 0.3359 | 0.1810 | 0.0459 | -0.0552 | -0.1070 | -0.0966 | -0.0162 | 0.1355 | | | | |
| 9 | 0.3346 | 0.1745 | 0.0269 | -0.0970 | -0.1837 | -0.2197 | -0.1938 | -0.0999 | 0.0623 | | | |
| 10 | 0.3340 | 0.1709 | 0.0156 | -0.1237 | -0.2361 | -0.3088 | -0.3297 | -0.2894 | -0.1830 | -0.0109 | | |
| 11 | 0.3337 | 0.1689 | 0.0090 | -0.1406 | -0.2711 | -0.3719 | -0.4309 | -0.4375 | -0.3837 | -0.2656 | -0.0842 | |

$$\begin{aligned}
 &+ 2P(S = \ell_1) \left(\frac{-n_{01}}{6} \right) \\
 &\leq P(S = \ell_1) \left(\frac{-F_{11}}{3} + \frac{n_{01}}{3} - \frac{n_{01}}{3} \right) \leq 0
 \end{aligned}$$

since $F_{11} \geq 0$.

Hence, we are left with the claim. We first assume $n_{10} \geq n_{01} + 2$ and treat the case $n_{10} = n_{01} - 1$ separately below. Note that nothing is to show if $\ell_1 > n_{10} + n_{01}$ since then only negative terms have non-zero probability. Hence, we assume $\ell_1 \leq n_{10} + n_{01}$ in the following. Since $n_{01} \leq n_{10} - 2$ and $F_{00} = 0$, we have $\ell_1 = n_{10} - F_{00} + F_{11} \geq (n_{10} + n_{01})/2 + 1$ and $\ell_2 = n_{10} - F_{00} + F_{11} \leq \ell_1 - 2$. Since S follows a binomial distribution with parameters $n_{10} + n_{01}$ and $1/2$, we have shown $\ell_1 \geq E(S) + 1$. Using the well-known monotonicity of the binomial distribution, we have $P(S = \ell_1 - 1) \geq P(S = \ell_1)$, and, since $P(S > \ell_2) \geq P(S = \ell_1) + P(S = \ell_1 - 1)$, we obtain $P(S > \ell_2) \geq 2P(S = \ell_1)$ as claimed in the case $n_{10} \geq n_{01} + 2$.

If $n_{10} = n_{01} + 1$ we have to argue more carefully. A relatively straightforward subcase (for $n_{10} \geq 12$) is $F_{11} = 0$. Then $\ell_2 = n_{01}$ and therefore

$$P(S > \ell_2) = P(S \geq n_{01} + 1) \geq P(S \geq (n_{10} + n_{01})/2) \geq \frac{1}{2}$$

by symmetry of the binomial distribution $\text{Bin}(n_{10} + n_{01}, 1/2)$. Also, for any $k \in \{0, \dots, n_{10} + n_{01}\}$, we have

$$P(S = k) \leq \frac{1}{\sqrt{\pi(n_{10} + n_{01})/2}} \leq \frac{1}{\sqrt{\pi n_{10}/2}} \leq 0.231$$

according to Lemma 7 for $n_{10} \geq 12$, which proves $P(S = \ell_1) \leq 1/4$ and, therefore, again $P(S > \ell_2) \geq 2P(S = \ell_1)$ as above.

We are left with $n_{10} = n_{01} + 1$ and $F_{11} > 0$. Here we take into account the first line of (13) and obtain due to $\ell_1 = \ell_2 + 1$ that

$$\begin{aligned}
 E(\Delta \mid P_{12}, F_{00} = 0, F_{11}) &= P(S > \ell_1) \cdot \frac{-2F_{11} - n_{10} + n_{01}}{3} \\
 &+ P(S = \ell_1) \cdot \frac{-F_{11} + n_{01}}{3} \\
 &+ P(S \geq \ell_1) \cdot \frac{-n_{01}}{6} \\
 &\leq \Pr(S = \ell_1 + 1) \left(-\frac{2F_{11}}{3} - \frac{n_{01}}{6} \right) \\
 &+ \Pr(S = \ell_1) \left(-\frac{F_{11}}{3} + \frac{n_{01}}{6} \right). \tag{14}
 \end{aligned}$$

We will show that $\Pr(S = \ell_1 + 1) / \Pr(S = \ell_1) = \alpha$ for some sufficiently large α and obtain

$$E(\Delta \mid P_{12}, F_{00} = 0, F_{11}) \leq \Pr(S = \ell_1) \left(-\frac{2\alpha F_{11}}{3} - \frac{\alpha n_{01}}{6} - \frac{F_{11}}{3} + \frac{n_{01}}{6} \right). \tag{15}$$

The ratio α will reflect the following trade-off: if F_{11} is small then $\ell_1 = n_{10} + F_{11}$ is close to the middle value of the binomial distribution with parameters $N := n_{10} + n_{01}$ and $\Pr(S = \ell_1 + 1)$ is not much smaller than $\Pr(S = \ell_1)$. Then the positive contribution of $n_{01}/6$ can almost be compensated by the negative $-\alpha n_{01}/6$. However, if F_{11} is big then α must be relatively small due to the tail of the binomial distribution. Then we exploit that the negative terms involving F_{11} make the drift bound small.

We now consider several ranges for F_{11} to relate the drift and α . The first range is $F_{11} \geq n_{01}/2$. Then have (even without bounding α) (Table 4)

$$E(\Delta \mid P_{12}, F_{00} = 0, F_{11}) \leq \Pr(S = \ell_1) \left(-\frac{n_{01}/2}{3} + \frac{n_{01}}{6} \right) \leq 0.$$

If $F_{11} < n_{01}/2$ and $F_{11} \geq 1$ then we have

$$\frac{n_{01}}{k^* + 1} \leq F_{11} < \frac{n_{01}}{k^*}$$

for some $k^* \geq 2$. Plugging the lower bound into (15), we have

$$E(\Delta \mid P_{12}, F_{00} = 0, F_{11}) \leq \Pr(S = \ell_1) \left(-\frac{2\alpha n_{01}}{3(k^* + 1)} - \frac{\alpha n_{01}}{6} - \frac{n_{01}}{3(k^* + 1)} + \frac{n_{01}}{6} \right).$$

The bracket equals

$$\frac{-4\alpha - \alpha(k^* + 1) - 2 + (k^* + 1)}{6(k^* + 1)} \cdot n_{01},$$

which becomes non-positive (and along with it also our drift bound) for

$$\alpha \geq \frac{k^* - 1}{k^* + 5}.$$

Hence, we are left with establishing this bound on α . Note that, since $n_{01} = n_{10} - 1$, the sum $N := n_{10} + n_{01}$ is odd, and it holds that

$$\ell_1 \leq n_{10} + \frac{n_{01}}{k^*} = \left(\frac{N}{2} + \frac{1}{2} \right) + \frac{\left(\frac{N}{2} - \frac{1}{2} \right)}{k^*} \leq \left(\frac{N}{2} + \frac{1}{2} \right) + \frac{N}{2k^*} = \left(\frac{1}{2} + \frac{1}{2k^*} \right) N + \frac{1}{2}.$$

Using the representation $P(S = k) = \binom{N}{k}2^{-N}$, and the identity

$$\binom{a}{b+1} / \binom{a}{b} = \frac{a-b}{b+1},$$

we have $P(S = \ell_1 + 1) \geq P(S = \ell_1) \frac{N/2 - N/(2k^*) - 1/2}{N/2 + N/(2k^*) + 1/2 + 1}$, hence

$$\alpha = \frac{N/2 - N/(2k^*) - 1/2}{N/2 + N/(2k^*) + 3/2} = \frac{k^* - 1 - k^*/N}{k^* + 1 + 3k^*/N} \geq \frac{k^* - 1 - k^*/12}{k^* + 1 + k^*/4},$$

where we used $N \geq 12$. Now,

$$\alpha - \frac{k^* - 1}{k^* + 5} \geq \frac{(11/12)k^* - 1}{(5/4)k^* + 1} - \frac{k^* - 1}{k^* + 5} = \frac{11(k^*)^2 - 23k^* + 12}{15(k^*)^2 + 87k^* + 60} \geq 0$$

for $k^* \geq 1$, which proves $\alpha \geq \frac{k^*-1}{k^*+5}$ as required. □

A.4 Proof of Lemma 21

Lemma 21 is similar to Lemma 16, saying that the drift is non-positive if at least 2 1-bits flip. Strangely enough, this holds for all $i \geq 2$, except for $i = n_{10} - 1$ where the drift is exponentially small in n_{10} .

Proof of Lemma 21 Using Lemma 13,

$$\begin{aligned} & E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = i) \\ &= P(S \geq n_{10} - 1 + i) \cdot \left(1 - \frac{2i}{3} - \frac{n_{10}}{6}\right) + P(S = n_{10} - 1 + i) \cdot \frac{17n_{10} - 17 + 13i}{36} \\ &= P(S \geq n_{10} - 1 + i) \cdot \frac{36 - 24i - 6n_{10}}{36} + P(S = n_{10} - 1 + i) \cdot \frac{17n_{10} - 17 + 13i}{36}. \end{aligned} \tag{16}$$

Noting that the first fraction is negative and the second one positive because of $i \geq 2$, the whole expression is non-positive if

$$\alpha := \frac{P(S \geq n_{10} - 1 + i)}{P(S = n_{10} - 1 + i)} \geq \frac{17 - 13i - 17n_{10}}{36 - 24i - 6n_{10}}. \tag{17}$$

We will now bound α depending on i . It is easy to see that $\frac{17-13i-17n_{10}}{36-24i-6n_{10}} < 1$ for $i \geq n_{10}$ while trivially $\alpha \geq 1$. This proves the statement for $i \geq n_{10}$. Hence, we now focus on $2 \leq i \leq n_{10} - 2$ and write $i = n_{10}/r$ for some $r \in [n_{10}/(n_{10} - 2), n_{10}/2]$. Then

$$\frac{17 - 13i - 17n_{10}}{36 - 24i - 6n_{10}} = \frac{17 - 13n_{10}/r - 17n_{10}}{36 - 24n_{10}/r - 6n_{10}} = \frac{13 + 17r - 17r/n_{10}}{24 + 6r - 36r/n_{10}}.$$

We now use a similar argument as in the proof of Lemma 18. Since $n_{10} + i = (1 + 1/r)n_{10}$ and S is based on a binomial distribution with $2n_{10}$ trials and success probability $1/2$, we obtain that

$$\begin{aligned} P(S = n_{10} + i)/P(S = n_{10} + i - 1) &= \binom{2n_{10}}{(1 + 1/r)n_{10}} \bigg/ \binom{2n_{10}}{(1 + 1/r)n_{10} - 1} \\ &= \frac{r - 1 + r/n_{10}}{r + 1} \end{aligned}$$

using the identity

$$\binom{a}{b + 1} \bigg/ \binom{a}{b} = \frac{a - b}{b + 1}$$

and, similarly,

$$\frac{P(S = n_{10} + i + 1)}{P(S = n_{10} + i)} = \frac{r - 1}{r + 1 + r/n_{10}}$$

as well as

$$\frac{P(S = n_{10} + i + 2)}{P(S = n_{10} + i + 1)} = \frac{r - 1 - r/n_{10}}{r + 1 + 2r/n_{10}}.$$

Hence, since $i \leq n_{10} - 2$

$$\begin{aligned} &\frac{P(S \geq n_{10} + i - 1)}{P(S = n_{10} + i - 1)} \\ &\geq \left(1 + \frac{r - 1 + r/n_{10}}{r + 1} + \frac{r - 1 + r/n_{10}}{r + 1} \cdot \frac{r - 1}{r + 1 + r/n_{10}} \right. \\ &\quad \left. + \frac{r - 1 + r/n_{10}}{r + 1} \cdot \frac{r - 1}{r + 1 + r/n_{10}} \cdot \frac{r - 1 - r/n_{10}}{r + 1 + 2r/n_{10}} \right) =: q, \end{aligned}$$

in other words $\alpha \geq q$, with equality holding for $i = n_{10} - 2$.

We now show $q - \frac{13 + 17r - 17r/N}{24 + 6r - 36r/N} \geq 0$ for the desired ranges of $N := n_{10}$ and i , which implies (17). Simple algebraic manipulations (supported by a computer algebra system) yield that

$$\begin{aligned} q - \frac{13 + 17r - 17r/N}{24 + 6r - 36r/N} &= \frac{(7r^4 + 32r^3 - 66r^2 + 40r - 13)N^4 + (-118r^4 + 174r^3 - 114r^2 + 26r)N^3}{6((r + 4)N - 6r)(r + 1)N(N(r + 1) + 2r)(N(r + 1) + r)} \\ &\quad + \frac{(-295r^4 + 102r^3 + r^2)N^2 + (-242r^4 + 10r^3)N - 72r^4}{6((r + 4)N - 6r)(r + 1)N(N(r + 1) + 2r)(N(r + 1) + r)}. \end{aligned}$$

Clearly, the denominator of the last fraction is positive for $N \geq 6$. We now prove that its numerator, hereinafter called $v(r, N)$, is positive for $N \geq 12$ and $N/(N - 2) \leq r \leq N/2$. To see this, we will compute its first and second partial derivative with respect to r :

$$\frac{\partial}{\partial r} v(r, N) = (28r^3 + 96r^2 - 132r + 40)N^4 + (-472r^3 + 522r^2 - 228r + 26)N^3 + (-1180r^3 + 306r^2 + 2r)N^2 + (-968r^3 + 30r^2)N - 288r^3$$

and

$$\frac{\partial^2}{\partial^2 r} v(r, N) = (84r^2 + 192r - 132)N^4 + (-1416r^2 + 1044r - 228)N^3 + (-3540r^2 + 612r + 2)N^2 + (-2904r^2 + 60r)N - 864r^2.$$

If we evaluate $v(r, N)$ and its first and second partial derivative at the minimum viable $r = N/(N - 2)$, we have

$$v(N/(N - 2), N) = \frac{8N^4(4N^3 - 42N^2 - 61N - 63)}{(N - 2)^4}$$

$$\left. \frac{\partial}{\partial r} v(r, N) \right|_{r=N/(N-2)} = \frac{2N^3(16N^4 - 28N^3 - 604N^2 - 1239N - 274)}{(N - 2)^3}$$

and

$$\left. \frac{\partial^2}{\partial^2 r} v(r, N) \right|_{r=N/(N-2)} = \frac{2N^2(72N^4 - 228N^3 - 2315N^2 - 2494N - 488)}{(N - 2)^2},$$

all of which are clearly non-decreasing for $N \geq 2$ (since the respective lead coefficients are positive and all other coefficients negative) and verified as positive (with values 1144.6272, 626482.9440 and 2117468.1601, respectively) for all $N \geq 12$, without claiming this to be tight in all three cases. Hence, if $N \geq 12$, $v(r, N)$ is positive at $r = N/(N - 2)$ and increases with r at this point. If the second partial derivative is positive for all $r \in [N/(N - 2), N/2]$ then $v(r, N)$ is monotone increasing on that interval. Otherwise the second derivative, a quadratic function, will be monotone increasing until its maximum at some $r \in [N/(N - 2), N/2]$ and then monotone decreasing. In particular this means that $v(r, N)$ has at most one root right of $r = N/(N - 2)$. We evaluate $v(N/2, N)$, i.e., at the maximum r -value considered and obtain the polynomial

$$\frac{7}{16}N^8 - \frac{27}{8}N^7 - \frac{211}{16}N^6 - \frac{87}{8}N^5 - 3N^4,$$

which is non-decreasing in N . Now, $v(N/2, N) = 1144.6272 > 0$ for $N = 12$. Due to the monotonicity of $v(N/2, N)$, this holds for all greater N as well, and together with

the observations regarding monotonicity and roots in the interval $[N/(N-2), N/2]$, this concludes the proof that $\nu(r, N) \geq 0$ for $N \geq 12$ and $N/(N-2) \leq r \leq N/2$, implying (17).

Finally, for $i = n_{10} - 1$ and $n_{10} \geq 10$ the formula simplifies to

$$\begin{aligned}
 E(\Delta \mid P_{12}, F_{00} = 1, F_{11} = n_{10} - 1) & \\
 &= P(S \geq 2n_{10} - 2) \cdot \left(1 - \frac{2(n_{10} - 1)}{3} - \frac{n_{10}}{6}\right) \\
 &\quad + P(S = 2n_{10} - 2) \cdot \frac{17n_{10} - 17 + 13(n_{10} - 1)}{36} \\
 &\leq P(S = 2n_{10} - 2) \cdot \frac{30n_{10} - 30}{36} \\
 &\leq P(S = 2n_{10} - 2) \cdot \frac{5n_{10}}{6} \\
 &= \binom{2n_{10}}{2} \cdot 2^{-2n_{10}} \cdot \frac{5n_{10}}{6} \\
 &\leq \frac{(2n_{10})^2}{2} \cdot 2^{-2n_{10}} \cdot \frac{5n_{10}}{6} \\
 &= \frac{5n_{10}^3}{3} \cdot 2^{-2n_{10}}.
 \end{aligned}$$

□

References

1. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, tenth GPO printing edition Dover, Ninth Dover Printing, New York (1964)
2. Auger, A., Doerr, B. (eds.): Theory of Randomized Search Heuristics. World Scientific Publishing, Springer (2011)
3. Carvalho Pinto, E., Doerr, C.: A simple proof for the usefulness of crossover in black-box optimization. In: Auger, A., Fonseca, C.M., Lourenço, N., Machado, P., Paquete, L., Whitley, D. (eds) Parallel Problem Solving From Nature - PPSN XV, volume 11102 of LNCS, pp. 29–41. Springer (2018)
4. Chernoff, H.: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23**, 493–507 (1952)
5. Corus, D., Oliveto, P.S.: Standard steady state genetic algorithms can hillclimb faster than mutation-only evolutionary algorithms. *IEEE Trans. Evol. Comput.* **22**(5), 720–732 (2018)
6. Corus, D., Oliveto, P.S.: On the benefits of populations for the exploitation speed of standard steady-state genetic algorithms. *Algorithmica* **82**(12), 3676–3706 (2020)
7. Dang, D.-C., Friedrich, T., Kötzing, T., Krejca, M.S., Lehre, P.K., Oliveto, P.S., Sudholt, D., Sutton, A.M.: Escaping local optima using crossover with emergent diversity. *IEEE Trans. Evol. Comput.* **22**(3), 484–497 (2018)
8. Dang, D.-C., Friedrich, T., Krejca, M.S., Kötzing, T., Lehre, P.K., Oliveto, P.S., Sudholt, D., Sutton, A.M.: Escaping local optima with diversity-mechanisms and crossover. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2016), pp. 645–652. ACM Press (2016)
9. Doerr, B.: Probabilistic tools for the analysis of randomized optimization heuristics. In: Doerr, B., Neumann, F. (eds) Theory of Evolutionary Computation—Recent Developments in Discrete Optimization, Natural Computing Series, pp. 1–87. Springer (2020)
10. Doerr, B., Happ, E., Klein, C.: Crossover can provably be useful in evolutionary computation. *Theoret. Comput. Sci.* **425**, 17–33 (2012)

11. Doerr, B., Neumann, F. (eds.): Theory of Evolutionary Computation—Recent Developments in Discrete Optimization. Springer, Natural Computing Series (2020)
12. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Longman, London (1989)
13. Gruder, O.: The theory of risk. In: 9th International Congress of Actuaries, vol 2, pp. 222 (1930)
14. Hwang, H.-K., Panholzer, A., Rolin, N., Tsai, T.-H., Chen, W.-M.: Probabilistic analysis of the (1+1)-evolutionary algorithm. *Evolut. Comput.* **26**(2) (2018)
15. Jansen, T.: Analyzing Evolutionary Algorithms—The Computer Science Perspective. Springer, Berlin (2013)
16. Jansen, T., Wegener, I.: On the analysis of evolutionary algorithms—a proof that crossover really can help. *Algorithmica* **34**(1), 47–66 (2002)
17. Johnson, N.L.: A note on the mean deviation of the binomial distribution. *Biometrika* **44**(3–4), 532–533 (1957)
18. Kötzing, T., Sudholt, D., Theile, M.: How crossover helps in pseudo-Boolean optimization. In: Proceedings of the 13th Annual Genetic and Evolutionary Computation Conference (GECCO 2011), pp. 989–996. ACM Press (2011)
19. Lehre, P.K., Witt, C.: Black-box search by unbiased variation. *Algorithmica* **64**(4), 623–642 (2012)
20. Lehre, P.K., Yao, X.: Crossover can be constructive when computing unique input–output sequences. *Soft. Comput.* **15**(9), 1675–1687 (2011)
21. Lengler, J.: A general dichotomy of evolutionary algorithms on monotone functions. *IEEE Trans. Evol. Comput.* **24**(6), 995–1009 (2020)
22. Neumann, F., Oliveto, P.S., Rudolph, G., Sudholt, D.: On the effectiveness of crossover for migration in parallel evolutionary algorithms. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2011), pp. 1587–1594. ACM Press (2011)
23. Neumann, F., Witt, C.: Bioinspired Computation in Combinatorial Optimization—Algorithms and Their Computational Complexity. Springer, Berlin (2010)
24. Oliveto, P.S., Sudholt, D., Witt, C.: A tight lower bound on the expected runtime of standard steady state genetic algorithms. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2020), pp. 1323–1331. ACM (2020)
25. Oliveto, P.S., Witt, C.: On the runtime analysis of the simple genetic algorithm. *Theoret. Comput. Sci.* **545**, 2–19 (2014)
26. Oliveto, P.S., Witt, C.: Improved time complexity analysis of the simple genetic algorithm. *Theoret. Comput. Sci.* **605**, 21–41 (2015)
27. Sudholt, D.: How crossover speeds up building-block assembly in genetic algorithms. *Evol. Comput.* **25**(2), 237–274 (2017)
28. Sutton, A.M.: Fixed-parameter tractability of crossover: steady-state GAs on the closest string problem. *Algorithmica* **83**(4), 1138–1163 (2021)
29. Witt, C.: Tight bounds on the optimization time of a randomized search heuristic on linear functions. *Comb. Probab. Comput.* **22**, 294–318 (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.