# On the Runtime Analysis of Fitness Sharing Mechanisms<sup>\*</sup>

Pietro S. Oliveto<sup>1</sup>, Dirk Sudholt<sup>1</sup>, and Christine Zarges<sup>2</sup>

<sup>1</sup> University of Sheffield, Sheffield, UK <sup>2</sup> University of Birmingham, Birmingham, UK

Abstract. Fitness sharing is a popular diversity mechanism implementing the idea that similar individuals in the population have to share resources and thus, share their fitnesses. Previous runtime analyses of fitness sharing studied a variant where selection was based on populations instead of individuals. We use runtime analysis to highlight the benefits and dangers of the original fitness sharing mechanism on the well-known test problem TwoMAX, where diversity is crucial for finding both optima. In contrast to population-based sharing, a (2+1) EA in the original setting does not guarantee finding both optima in polynomial time; however, a  $(\mu+1)$  EA with  $\mu \geq 3$  always succeeds in expected polynomial time. We further show theoretically and empirically that large offspring populations in  $(\mu+\lambda)$  EAs can be detrimental as overpopulation can make clusters of search points go extinct.

**Keywords:** Evolutionary computation, diversity mechanisms, fitness sharing, runtime analysis.

## 1 Introduction

Diversity mechanisms are used in evolutionary computation to tackle multimodal optimisation problems [7]. The main idea is to maintain dissimilar individuals in the population such that different *niches* explore different peaks of the fitness landscape. A popular diversity mechanism is *fitness sharing* [1,6]. In this scheme niche formation is induced by using a *sharing function* that derates the fitness of an individual by an amount related to its similarity to the rest of the population. Different fitness sharing functions are obtained according to how the distance between individuals is defined [7]. *Genotypic sharing* uses Hamming distance and is generally employed when no phenotypic knowledge is available [7]. In *phenotypic sharing* the distance is defined using problem specific knowledge.

Previous theoretical work on diversity mechanisms has concentrated on a somewhat unusual implementation of the sharing mechanism. Let P denote the union of parents and offspring. Rather than selecting *individuals* based on their

<sup>&</sup>lt;sup>\*</sup> The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 618091 (SAGE).

T. Bartz-Beielstein et al. (Eds.): PPSN XIII 2014, LNCS 8672, pp. 932–941, 2014.

<sup>©</sup> Springer International Publishing Switzerland 2014

shared fitness f(x, P), selection was done on a level of populations, creating a population that maximises the overall shared fitness of the population (i. e., creating  $P^* = \arg \max\{\sum_{x \in P'} f(x, P') \mid P' \subset P, |P'| = \mu\}$  [3,4,10,11]). The drawback of this approach is that all the possible subsets of P of size  $\mu$  need to be examined. For large  $\mu$  and  $\lambda$ , this is prohibitive.

In this paper we analyse the performance of the original fitness sharing approach used in many practical applications [5]. We consider a standard  $(\mu + \lambda)$  EA (see Algorithm 1) using the shared fitness values within the selection for replacement. We analyse the algorithm on the same TWOMAX function used in the literature for the analysis of the effectiveness of the previous approach for multimodal optimisation [4].

TWOMAX $(x) := \max \{\sum_{i=1}^{n} x_i, n - \sum_{i=1}^{n} x_i\}$  is a simple bimodal function consisting of two different symmetric branches (i. e., ONEMAX and ZEROMAX) and we have defined both  $0^n$  and  $1^n$  to be global optima. Since we aim at analysing the global exploration capabilities of the population-based EA, we call a run *successful* if it manages to find both optima (i. e., a population is reached that contains both  $0^n$  and  $1^n$ ) efficiently. The expected number of generations for this to happen is called *expected running time*. Apart from TwoMAX being the ideal benchmark function for the analysis (i. e., the simplest bimodal function), its choice also allows comparisons with the previous approaches.

A  $(\mu+1)$  EA using the unconventional approach (i. e., maximising the phenotypic shared fitness of the population) can efficiently optimise TwoMAX for any population size  $\mu \ge 2$  [4]. The reason is that, in any population, the individuals with the smallest and the largest number of ones are always accepted for the next generation. Our analysis shows that using the conventional (phenotypic) sharing approach leads to considerably different behaviours of evolutionary algorithms. We illustrate this by using the analytical framework presented in Sect. 2.

We first concentrate on the effects of the parent population in Sect. 3. A population of size  $\mu = 2$  is not sufficient to guarantee that the  $(\mu+1)$  EA finds both optima in polynomial time. If the two individuals are initialised on the same branch, then there is a high probability that they will both find the same local optimum. Furthermore, there is a chance that the algorithm fails also when the two individuals are initialised on opposite branches. This leads to a worse failure probability than that of a simple crowding algorithm or that of a (1+1) EA that is restarted twice. On the other hand Sect. 4 shows that for  $\mu \geq 3$ , once the population is close enough to one optimum, individuals descending the branch heading towards the other optimum are accepted. This threshold, that allows successful runs with probability 1, lies further away from the local optimum as the population size increases.

Concerning the effects of the offspring population, in Sect. 5 we show that large values of  $\lambda$  can be detrimental. We rigorously prove that increasing the offspring population of a  $(\mu+1)$  EA to a  $(\mu+\lambda)$  EA, with  $\mu = 2$  and  $\lambda \geq 2$  a constant, results in an overcrowding that can make a (sub-)population go extinct. For the special case of  $\lambda = 2$  we also prove an increased failure probability.

| <b>Algorithm 1.</b> $(\mu + \lambda)$ EA with fitness sharing                       |
|---|
| 1. Choose $\mu$ individuals uniformly at random from $\{0,1\}^n$ .                  |
| 2. repeat   |
| 3. for each $1 \le i \le \lambda$ do  |
| 4. Select a parent x uniformly at random from the population.                       |
| 5. Let $x^i := x$ . Flip each bit in $x^i$ independently with probability $1/n$ .   |
| 6. end for  |
| 7. Create a new population by selecting the $\mu$ best individuals according to the |
| shared fitness, breaking ties towards favouring offspring over parents, breaking    |
| remaining ties uniformly at random.   |
| 8. until stopping criterion met   |

We complement this result with an empirical analysis that suggests that the  $(\mu+1)$  EA is successful if  $\lambda < \lfloor \mu/2 \rfloor$  and that it almost always fails for  $\lambda \ge \mu$ .

In this extended abstract, some proofs are omitted due to space limitations.

### 2 Analytical Framework

Throughout this work, |x| denotes the number of 1-bits in x. The shared fitness of an individual x in population P is  $f(x, P) := \frac{f(x)}{\sum_{y \in P} sh(x,y)}$  and the sharing function is  $sh(x, y) := \max\{0, 1 - (d(x, y)/\sigma)^{\alpha}\}$ . Here, d is the distance between the two individuals,  $\sigma$  is the *sharing distance* beyond which individuals do not share fitness and  $\alpha$  is a constant, typically set to 1, that regulates the shape of the sharing function. We consider fitness sharing with phenotypic sharing as in [4], where the distance between individuals is based on the number of ones. We use  $\sigma = n/2$  (as in [4]) and the standard value  $\alpha = 1$  and obtain

$$f(x, P) := \frac{f(x)}{\sum_{y \in P} \max\left\{0, 1 - \frac{||x| - |y||}{n/2}\right\}}.$$

Let  $P := \{x_1, x_2, \ldots, x_s\}$  denote the extended population of current search points and the new offspring, labelled such that  $|x_1| \leq |x_2| \leq \cdots \leq |x_s|$ . Let  $D_j := \sum_{i=1}^s \min\{||x_j| - |x_i||, n/2\}$  denote the sum of phenotypic distances to all other members of the extended population. Individual distances are capped at the sharing distance n/2 so that the shared fitness can be written as

$$f(x_i, P) = \frac{f(x_i)}{s - 2D_i/n}$$

Phenotypic fitness sharing, along with the shape of the TwoMAX function, implies that a unique best individual will always survive, as it has a better fitness than the individual with the closest number of ones, and it has a larger phenotypic distance to other individuals. This means that in a  $(\mu+1)$  EA the current best fitness never decreases; this also holds if multiple individuals have the same current best fitness, as only one individual is removed by selection. **Lemma 1.** Let  $P = \{x_1, \ldots, x_s\}$  with  $|x_1| \leq \cdots \leq |x_s|$ . If  $f(x_1) > f(x_2)$  then  $f(x_1, P) > f(x_2, P)$ . Likewise, if  $f(x_{s-1}) < f(x_s)$  then  $f(x_{s-1}, P) < f(x_s, P)$ .

As a result, the  $(\mu+1)$  EA never decreases its current best fitness and finds at least one optimum in expected time  $O(\mu n \log n)$ .

The time bound follows from standard arguments, as used in [4]. The symmetry between  $f(x_1, P)$  vs.  $f(x_2, P)$  and  $f(x_{s-1}, P)$  vs.  $f(x_s, P)$  follows from swapping the meaning of zeros and ones. This also applies to further statements, where for simplicity we omit symmetric statements.

The following Main Lemma gives sufficient and necessary conditions on when the shared fitness of one individual is better than another.

Lemma 2 (Main Lemma). Let 
$$P = \{x_1, ..., x_s\}$$
 with  $|x_1| \le \dots \le |x_s|$  and  
fix  $1 \le i \le s - 1$ . If  $f(x_i) - f(x_{i+1}) = |x_{i+1}| - |x_i| > 0$  and  $|x_s| - |x_1| \le n/2$ ,  
 $f(x_{i+1}, P) \ge f(x_i, P) \Leftrightarrow f(x_i) \cdot (2i - s) + D_i \ge s \cdot n/2$   
 $\Leftrightarrow f(x_{i+1}) \cdot (2i - s) + D_{i+1} \ge s \cdot n/2$ .

The same holds if all inequalities " $\geq$ " are replaced by strict inequalities ">". Moreover, for i = s - 1

$$f(x_s, P) > f(x_{s-1}, P) \Leftrightarrow |x_s| > \sum_{i=1}^{s-1} |x_i| - n/2 \cdot (s-4).$$

*Proof.* Note that  $|x_s| - |x_1| \le n/2$  implies that all pairs of individuals do share fitness. Comparing  $D_i$  and  $D_{i+1}$ , for the latter the distance to  $x_1, \ldots, x_{i-1}$  is higher by  $|x_{i+1}| - |x_i|$ , and the distance to  $x_{i+2}, \ldots, x_s$  is lower by  $|x_{i+1}| - |x_i|$ :

$$\begin{split} D_{i+1} &= D_i + (i-1) \cdot (|x_{i+1}| - |x_i|) + (s-i-1) \cdot (|x_i| - |x_{i+1}|) \\ &= D_i + (2i-s) \cdot (|x_{i+1}| - |x_i|). \end{split}$$

Using the shorthand  $h := |x_{i+1}| - |x_i|$ ,

$$f(x_{i+1}, P) = \frac{f(x_{i+1})}{s - \frac{D_{i+1}}{n/2}} = \frac{f(x_i) - h}{s - \frac{D_i + (2i - s)h}{n/2}}.$$

Now  $f(x_{i+1}, P) \ge f(x_i, P)$  is equivalent to

$$\frac{f(x_i) - h}{s - \frac{D_i + (2i - s)h}{n/2}} \ge \frac{f(x_i)}{s - \frac{D_i}{n/2}}$$
  

$$\Leftrightarrow (f(x_i) - h) \cdot (sn/2 - D_i) \ge f(x_i) \cdot (sn/2 - D_i - (2i - s)h)$$
  

$$\Leftrightarrow f(x_i) \cdot (2i - s)h + h \cdot D_i \ge h \cdot sn/2$$
  

$$\Leftrightarrow f(x_i) \cdot (2i - s) + D_i \ge sn/2.$$

In the last step we used h > 0. The same calculations hold if " $\geq$ " is replaced by ">" throughout. The second equivalence from the statement follows from

$$f(x_i) \cdot (2i - s) + D_i = (f(x_{i+1}) + h) \cdot (2i - s) + D_{i+1} - h(2i - s)$$
$$= f(x_{i+1}) \cdot (2i - s) + D_{i+1}.$$

The last statement follows from simple manipulations.

The Main Lemma gives a condition for the individual of lowest raw fitness (i. e.,  $x_s$ ) to be accepted by selection. Concerning the  $(\mu+1)$  EA, the condition clearly shows that for  $\mu = 2$  at least n/2 bits have to flip (i. e.,  $|x_3| - |x_2| \ge n/2$ ). On the other hand, for  $\mu \ge 3$  offspring with lower fitness values are accepted once the population is close enough to the optimum  $0^n$ . This threshold is further away from the optimum as the population size increases. If mutation was only allowed to flip one bit and  $\mu = 3$ , then it is necessary that both  $x_1$  and  $x_2$  reach the local optimum before decreasing moves are accepted (i. e.,  $|x_1| + |x_2| = 0$ ). For  $\mu = 4$  the sum of 1-bits in the first 4 individuals can be up to  $|x_1| + |x_2| + |x_3| + |x_4| \le n/2$  for any decreasing move to be accepted by the  $(\mu+1)$  EA.

In general, the conditions from Lemma 2 are true for  $x_{s-1}$  and  $x_s$  if  $|x_{s-1}| < n/2$  and two individuals are in the optimum  $0^n$  as then  $s^{-2}$ 

$$f(x_{s-1})(s-2) + D_{s-1} \ge (n - |x_{s-1}|)(s-2) + (s-2)|x_{s-1}| - \sum_{i=1}^{n} |x_i|$$
  
>  $n(s-2) - (s-4)n/2 = sn/2.$ 

**Lemma 3.** If  $P = \{x_1, \ldots, x_s\}$ ,  $|x_1| \le \cdots \le |x_s|$ , with  $|x_{s-1}| < n/2$  and  $|x_1| = |x_2| = 0$  then  $f(x_{s-1}, P)(s-2) + D_{s-1} > sn/2$ .

#### 3 Population Size $\mu = 2$ Is Not Enough

We first investigate the (2+1) EA, showing that a population size of  $\mu = 2$  is not sufficient to guarantee finding both optima.

The following lemma gives sufficient and necessary conditions for a single individual on a branch to survive. For  $|x_3| = |x_2|$  the statement implies that  $x_1$  survives if the distance from n/2 to  $x_2$  is less than around 3/2 the distance from n/2 to  $x_1$ . The condition for survival sharpens when  $|x_3| > |x_2|$ ; however, as  $x_2$  and  $x_3$  result from a mutation of one another,  $|x_3| - |x_2|$  is bounded from above by the number of bits flipped in that mutation.

**Lemma 4.** Let  $\mu = 2$  and  $P = \{x_1, x_2, x_3\}$  with  $|x_1| < n/2 < |x_2| \le |x_3|$  and  $|x_3| - |x_1| \le n/2$ . Let  $d_1 := n/2 - |x_1|$  and  $d_2 := |x_2| - n/2$ , then

$$\begin{aligned} &f(x_1,P) > f(x_2,P) \Leftrightarrow \\ &d_2 < \left(\frac{3}{2} + \frac{7d_1}{n+6|x_1|}\right) \cdot d_1 + \frac{(|x_3|+|x_2|)(f(x_2)-f(x_1))}{n/2+3|x_1|} \end{aligned}$$

The following theorem states that with a probability greater than 1/2, the (2+1) EA will end up with both individuals in the same optimum, leading to an exponential running time from there. This performance is worse than having two independent runs of a (1+1) EA, as in deterministic crowding, for which the probability of finding both optima is exactly 1/2 [4].

**Theorem 1.** The (2+1) EA with fitness sharing with probability  $1/2 + \Omega(1)$  will reach a population with both members in the same optimum, and then the expected time for finding both optima from there is  $\Omega(n^{n/2})$ .

Proof. Using that  $2^{-n} \binom{n}{i} \leq 2^{-n} \binom{n}{n/2} = \Theta(1/\sqrt{n})$  for any  $0 \leq i \leq n$ , it is easy to show that with probability  $1 - O(n^{1/3}/\sqrt{n}) = 1 - o(1)$  for both initial search points  $x_1, x_2$  we have  $|x_1|, |x_2| \notin [n/2 - n^{1/3}, n/2 + n^{1/3}]$ . By symmetry, with probability 1/2 - o(1),  $x_1$  and  $x_2$  are on the same branch. The probability of a mutation jumping from one branch to the other is then at most  $1/(n^{1/3}!) = 2^{-\Omega(n^{1/3}\log n)}$ , and the probability of this happening in expected polynomial time is still of the same order. This implies that w.o.p. no individuals on the opposite branch will be created in polynomial time as long as no offspring of decreasing fitness are ever accepted on the branch. In the following we prove by contradiction that such offspring are always rejected.

Assuming both search points and the offspring are all on the same branch, w.l.o.g. the left branch, by Lemma 2

$$f(x_3, P) \ge f(x_2, P) \Leftrightarrow f(x_2) + D_2 \ge 3 \cdot \frac{n}{2} \tag{1}$$

where  $D_2 = (|x_2| - |x_1|) + (|x_3| - |x_2|) = |x_3| - |x_1|$ . Then  $f(x_2) + D_2 = n - |x_2| + |x_3| - |x_1| \le n + |x_3| - |x_2|$ . This implies that (1) only holds if  $|x_3| - |x_2| \ge n/2$ , which is a contradiction since there are no points on the left branch differing in more than n/2 one-bits. Hence, the claim that no offspring on the left branch of worse fitness than  $x_2$  are ever accepted, is proved. By Lemma 1,  $0^n$  will be reached in expected time  $O(n \log n)$ . In a further expected  $2 \cdot (1 - 1/n)^n = O(1)$  generations, the extended population will contain a clone of  $0^n$ , and from then on any offspring  $x_3$  with  $0 < |x_3| \le n/2$  will be rejected. Then the expected time to create an individual on the other branch is  $\Omega(n^{n/2})$  since at least n/2 bits need to flip.

The claimed probability  $1/2 + \Omega(1)$  follows from considering the following additional event, which is disjoint from the above. The algorithm also fails if, using the notation from Lemma 4,  $3\sqrt{n}/4 \le d_2 \le \sqrt{n}$  (probability at least 0.02) and  $\sqrt{n}/3 \ge d_1 \ge 0$  (probability at least 0.21). If then in the first generation a clone of  $x_2$  is generated (probability at least  $1/2 \cdot (1 - 1/n)^n > 1/8$ ), we have

$$\left(\frac{3}{2} + \frac{7d_1}{n+6|x_1|}\right) \cdot d_1 + \frac{(x_3 + x_2)(f(x_2) - f(x_1))}{n/2 + 3x_1} \le \frac{\sqrt{n}}{3} \cdot \frac{3}{2} + O(1) < \frac{3\sqrt{n}}{4} \le d_2$$

if n is large enough. Now Lemma 4 implies  $f(x_1, P) < f(x_2, P) = f(x_3, P)$ , hence  $x_1$  will be removed. Then we are in the same situation as when initialising two individuals on the same branch.

However, there is still a constant probability that the (2+1) EA finds both optima in polynomial expected time. This holds if the EA is initialised with its two search points on different branches, and if these two search points maintain similar fitness values throughout the run.

**Theorem 2.** The (2+1) EA with fitness sharing with probability  $\Omega(1)$  will find both optima in time  $O(n \log n)$ .

Due to space restrictions, we only sketch the proof. Let  $x_1, x_2$  be the two initial search points and  $d_1 := n/2 - |x_1|$  and  $d_2 := |x_2| - n/2$ . With probability  $\Omega(1)$ ,  $x_1$  and  $x_2$  are on opposite branches and have similar fitness:  $\frac{3}{4}\sqrt{n} \le d_1, d_2 \le \sqrt{n}$ .

Now, assume w. l. o. g. that when a new offspring is created and the population contains  $x_1, x_2, x_3$  in order of their numbers of ones, that  $x_2$  and  $x_3$  are on the same branch. If  $f(x_1) > f(x_2)$ , Lemma 1 implies that  $f(x_1, P) > f(x_2, P)$  and  $f(x_2, P) < f(x_3, P)$  if  $|x_3| > |x_2|$ . Then  $x_1$  is guaranteed to survive.

Now assume  $f(x_1) \leq f(x_2)$ . It is easy to derive from Lemma 4 and further arguments for  $|x_3| - |x_1| > n/2$  that  $f(x_1, P) > f(x_2, P)$  follows if  $d_1 \geq (2/3) \cdot d_2$ .

For a current population  $P = \{x_1, x_2\}$  define a potential  $g(P) := \min\{d_1, d_2\}$  $-(2/3) \cdot \max\{d_1, d_2\}$ . Intuitively, the potential indicates a distance to a population where the lower-fitness individual is at risk of dying. For  $d_1 \leq d_2$  we have, using Lemma 4,

$$g(P) \ge \frac{\sqrt{n}}{24} \iff d_1 \ge \frac{2}{3} \cdot d_2 \implies f(x_1, P) > f(x_2, P).$$

For the initial population  $P_0$  we have  $g(P_0) \ge 3/4 \cdot \sqrt{n} - 2/3 \cdot \sqrt{n} \ge \sqrt{n}/12$ . If  $d_1 \le d_2 - k$  for some  $k \in \mathbb{N}$ , the potential increases by k if  $d_1$  increases by k. However, the potential only decreases by  $2/3 \cdot k$  if  $d_2$  increases by k. Moreover, increasing  $d_1$  is easier than increasing  $d_2$  as the former contains more "incorrect" bits (cf. Lemma 13 in [2]). This shows that, whenever the potential changes, it increases in expectation by 1/3.

A straightforward application of the simplified drift theorem [8,9] shows that with probability  $2^{-\Omega(\sqrt{n})}$  the potential never decreases below  $\sqrt{n}/24$  in  $2^{\Omega(\sqrt{n})}$ steps. So, with overwhelming probability  $x_1$  survives until both optima are reached.

#### 4 Population Size $\mu \geq 3$ Succeeds

A population of size  $\mu = 2$  may fail, but we show that a  $(\mu+1)$  EA with  $\mu \ge 3$  always finds both optima in expected time  $O(\mu n \log n)$ .

The following lemma is an extension of the Main Lemma to the case where an individual  $x_{\mu+1}$  is on the other branch compared to the rest of the population. In particular, a stronger condition is given such that  $x_{\mu+1}$  will survive selection when  $f(x_{\mu}) > f(x_{\mu+1})$ . The proof is similar to the one for the Main Lemma.

**Lemma 5.** Let  $|x_{\mu}| < n/2$ ,  $|x_{\mu+1}| > n/2$  and  $f(x_{\mu}) > f(x_{\mu+1})$ . Also let  $h_{\mu} := n/2 - |x_{\mu}|$  and  $h_{\mu+1} := |x_{\mu+1}| - n/2$ . Then

$$f(x_{\mu}) \cdot (\mu - 1) \cdot \frac{h_{\mu}}{h_{\mu} - h_{\mu+1}} + D_{\mu} \ge (\mu + 1) \cdot n/2 \Rightarrow f(x_{\mu+1}, P) \ge f(x_{\mu}, P).$$

The following lemma states that if there is a bounded number r of individuals in one optimum then they will have better shared fitness than the next suboptimal individual. This implies that r such individuals survive in the  $(\mu+1)$  EA; the same holds if there are more than r such individuals in the extended population as only one individual is being removed.

**Lemma 6.** Let  $P = \{x_1, ..., x_s\}$  with  $|x_1| \le \cdots \le |x_s|$ . Assume  $|x_1| = \cdots = |x_r| = 0 < |x_{r+1}|$  and  $|x_s| < n$ . If  $r \le 2$  or if both  $|x_{r+1}| \ge n/2$  and  $r \le s/2$ ,

then for all  $1 \leq i \leq r$  we have  $f(x_i, P) > f(x_{r+1}, P)$ . In particular, if the current population of the  $(\mu+1)$  EA contains at least two individuals  $0^n$ , two such individuals always survive.

With these lemmas we are ready to prove the main result of this section.

**Theorem 3.** Let  $\mu \ge 3$ . The  $(\mu+1)$  EA with fitness sharing will find both optima of TWOMAX with probability 1 in expected time  $O(\mu n \log n)$ .

*Proof.* By Lemma 1, in expected time  $O(\mu n \log n)$  one of the two optima is found. W.l.o.g. we assume the  $0^n$  optimum is found. In expected time  $O(\mu)$  a clone of  $0^n$  is created (i.e.,  $|x_2| = 0$ ) and by Lemma 6  $x_1$  and  $x_2$  (or clones thereof) will survive for the rest of the run.

We show that then the individual with the largest number of ones,  $x_{\mu+1}$ (or a clone thereof), will always survive. If  $|x_{\mu}| = |x_{\mu+1}|$  then  $x_{\mu+1}$  or a clone survive. If  $n/2 \leq |x_{\mu}| < |x_{\mu+1}|$  then  $f(x_{\mu+1}) > f(x_{\mu})$  and the claim follows from Lemma 1. If  $|x_{\mu}| < n/2$  then Lemma 3 implies  $f(x_{s-1})(s-2) + D_{s-1} > sn/2$ (where  $s = \mu + 1$ ). If  $|x_{\mu+1}| \leq n/2$ , by the Main Lemma this condition is equivalent to  $f(x_{\mu+1}, P) > f(x_{\mu}, P)$ . Otherwise, the same conclusion follows from Lemma 5 as  $h_{\mu}/(h_{\mu} - h_{\mu+1}) > 1$ . So, in all cases  $x_{\mu+1}$  survives. The expected time for  $x_{\mu+1}$  reaching  $1^n$  is again  $O(\mu n \log n)$  as in [4].

Our analysis has revealed two very different behaviours. It is possible that the whole population climbs up one branch. But once a sufficiently large overall fitness value has been obtained – at the latest when two individuals have found an optimum – then the population expands towards lower fitness values as then the individuals with the smallest and the largest numbers of 1-bits always survive.

#### 5 Too Large Offspring Population Sizes

Fitness sharing works for the  $(\mu+1)$  EA, but for larger offspring populations it can have undesirable effects: if a cluster of individuals creates too many offspring, sharing decreases the shared fitness of all individuals in the cluster, and the cluster may go extinct. We consider this problem of overpopulation for  $\mu = 2$ and  $\lambda \ge \mu$  with  $\lambda = O(1)$ . In this setting we cannot guarantee convergence to populations with both optima any more, i. e., depending on  $\lambda$  we can lose one or even both optima.

Assume that all individuals are in the same optimum. With probability  $\Omega(1)$ , we create  $\lambda - 1$  copies and one point with distance 1 to the optimum. Then,  $f(x_1, P) = \ldots = f(x_{\lambda+1}, P) = n/((\lambda+2)-2/n)$  and  $f(x_{\lambda+2}, P) = (n-1)/((\lambda+2)-(\lambda+1)\cdot 2/n)$ . We see that  $f(x_i, P) < f(x_{\lambda+2}, P)$  for all  $i \in \{1, \ldots, \lambda+1\}$ and  $\lambda \geq 2$ . Thus, selection picks  $x_{\lambda+2}$  and one of the optimal points. Following the same argumentation, we lose both optima if  $\lambda \geq 6$ : If mutation creates  $\lambda - 2$  copies and two points with distance 1 to the optimum (also with probability  $\Omega(1)$ ), we have  $f(x_1, P) = \ldots = f(x_\lambda, P) = n/((\lambda+2) - 2 \cdot 2/n) < (n-1)/((\lambda+2) - \lambda \cdot 2/n) = f(x_{\lambda+1}, P) = f(x_{\lambda+2}, P)$  for  $\lambda \geq 6$ . In exactly the same way we show that both optima are lost with probability  $\Omega(1)$  if  $\lambda \geq 6$  even if they are on different branches, i. e., we create  $\lfloor \lambda/2 \rfloor$  offspring on the left branch and  $\lceil \lambda/2 \rceil$  on the right branch where exactly one offspring on each branch has distance 1 to the optimum and the remaining offspring are copies.

Offspring populations can also decrease diversity in the following way.

**Lemma 7.** With probability 1 - o(1) the  $(2 + \lambda)$  EA with fitness sharing,  $\lambda \ge 2$  and  $\lambda = O(1)$  will, at some point of time before an optimum is reached, obtain a population with both members on the same branch.

Proof (Proof sketch). The proof mainly uses that in a single iteration with probability  $\Omega(1)$  only copies of  $x_1$  and  $x_2$  are created. We show that if  $f(x_1) \neq f(x_2)$ and if we have a surplus of offspring on the branch with smaller fitness (also probability  $\Omega(1)$ ), this branch goes extinct. If  $f(x_1) = f(x_2)$  in iteration t we have  $f(x_1) \neq f(x_2)$  in iteration t+1 with probability  $\Omega(1)$  and if  $f(x_1) \neq f(x_2)$ in iteration t we still have  $f(x_1) \neq f(x_2)$  in iteration t+1 with probability  $\Omega(1)$ . Thus, with probability  $1 - 2^{-\Omega(n)}$  there are  $\Omega(n)$  iterations with  $f(x_1) \neq f(x_2)$ before an optimum is reached and consequently, with probability  $1 - 2^{-\Omega(n)}$ , one branch will take over the whole population before an optimum is reached.  $\Box$ 

In order to show that the  $(2 + \lambda)$  EA also reaches a population with both members in the same optimum we additionally need to show that the population will not be stuck somewhere on the branch and that individuals cannot traverse back to the other branch. We consider this for the special case of  $\lambda = 2$ .

**Theorem 4.** With probability 1-o(1) the (2+2) EA with fitness sharing will, at some point of time, reach a population with both members in the same optimum. The expected time for finding both optima from there is  $\Omega(n^{n/2})$ .

*Proof (Proof sketch).* Due to Lemma 7 both individuals are on the same branch with probability 1 - o(1) before an optimum is reached.

We show that a current best individual is never lost. Due to Lemma 1  $f(x_1, P) > f(x_2, P)$  holds. We apply Lemma 2 and have  $f(x_3, P) \ge f(x_2, P) \Leftrightarrow D_2 \ge 2n$  where  $D_2 = d_{2,1} + d_{2,3} + d_{2,4}$  since  $d_{2,2} = 0$ . Since all individuals are on the same branch  $d_{i,j} \le n/2$ . This implies that  $D_2 \le 3n/2$  and thus,  $f(x_3, P) < f(x_2, P) < f(x_1, P)$ . Thus, a single best individual will always survive. Moreover, in case of 2 best individuals at least one of them will be selected for the next iteration. Since  $\mu = 2$  we are guaranteed to select at least one of the best individuals if there are 3 or 4 best. Following the same argumentation, we see that a single improved offspring of a best individual will always be accepted. Thus, we will reach a population with both members in the same optimum. The claim about the expected time to find both optima follows as in Theorem 1.

#### 6 Experiments

Our final contribution is a set of experiments, shown in Table 1, where we ran  $(\mu+\lambda)$  EAs for n = 100 bits and varying values of  $2 \le \mu \le 12$  and  $1 \le \lambda \le 12$ . We recorded the success rate as the number of runs where both optima were

**Table 1.** Success rates of the  $(\mu + \lambda)$  EA with fitness sharing on TwoMAX in 1000 runs, stopped after 100000 generations, and once both optima were found

| $\mu$ | $\lambda = 1$ | $\lambda = 2$ | $\lambda = 3$ | $\lambda = 4$ | $\lambda = 5$ | $\lambda = 6$ | $\lambda = 7$ | $\lambda = 8$ | $\lambda = 9$ | $\lambda = 10$ | $\lambda = 11$ | $\lambda = 12$ |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|----------------|----------------|
| 2     | 0.23          | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0            | 0.0            | 0.0            |
| 3     | 1.0           | 0.277         | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0            | 0.0            | 0.0            |
| 4     | 1.0           | 0.602         | 0.32          | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0            | 0.0            | 0.0            |
| 5     | 1.0           | 0.793         | 0.644         | 0.025         | 0.0           | 0.0           | 0.0           | 0.0           | 0.0           | 0.0            | 0.0            | 0.0            |
| 6     | 1.0           | 1.0           | 0.824         | 0.687         | 0.261         | 0.0           | 0.0           | 0.0           | 0.0           | 0.0            | 0.0            | 0.0            |
| 7     | 1.0           | 1.0           | 0.936         | 0.861         | 0.768         | 0.156         | 0.0           | 0.0           | 0.0           | 0.0            | 0.0            | 0.0            |
| 8     | 1.0           | 1.0           | 1.0           | 0.926         | 0.874         | 0.816         | 0.064         | 0.0           | 0.0           | 0.0            | 0.0            | 0.0            |
| 9     | 1.0           | 1.0           | 1.0           | 0.996         | 0.957         | 0.894         | 0.828         | 0.039         | 0.0           | 0.0            | 0.0            | 0.0            |
| 10    | 1.0           | 1.0           | 1.0           | 1.0           | 0.972         | 0.957         | 0.918         | 0.843         | 0.032         | 0.0            | 0.0            | 0.0            |
| 11    | 1.0           | 1.0           | 1.0           | 1.0           | 1.0           | 0.98          | 0.945         | 0.929         | 0.805         | 0.02           | 0.001          | 0.0            |
| 12    | 1.0           | 1.0           | 1.0           | 1.0           | 1.0           | 0.99          | 0.978         | 0.972         | 0.945         | 0.738          | 0.029          | 0.0            |

found within 100000 generations. The table shows a clear distinction between efficient and inefficient behaviour: for  $\lambda < \lfloor \mu/2 \rfloor$  runs were always successful, whereas runs for  $\lambda \ge \mu$  always failed (except for one run with  $\lambda = \mu = 11$ ).

#### References

- Cioppa, A.D., Stefano, C.D., Marcelli, A.: On the role of population size and niche radius in fitness sharing. IEEE Trans. Evol. Comput. 8(6), 580–592 (2004)
- Doerr, B., Johannsen, D., Winzen, C.: Multiplicative drift analysis. Algorithmica 64, 673–697 (2012)
- Fischer, S., Wegener, I.: The one-dimensional Ising model: Mutation versus recombination. Theor. Comput. Sci. 344(2–3), 208–225 (2005)
- Friedrich, T., Oliveto, P.S., Sudholt, D., Witt, C.: Analysis of diversity-preserving mechanisms for global exploration. Evol. Comput. 17(4), 455–476 (2009)
- 5. Goldberg, D.E.: Genetic Algorithms for Search, Optimization, and Machine Learning. Addison-Wesley (1989)
- Goldberg, D.E., Richardson, J.: Genetic algorithms with sharing for multimodalfunction optimization. In: Proc. of ICGA, pp. 41–49. Lawrence Erlbaum Associates (1987)
- Mahfoud, S.W.: Niching methods. In: Bäck, T., Fogel, D.B., Michalewicz, Z., (eds.) Handbook of Evolutionary Computation, pp. C6.1:1–C6.1:4. IOP Publishing and Oxford University Press (1997)
- Oliveto, P.S., Witt, C.: Simplified drift analysis for proving lower bounds in evolutionary computation. Algorithmica 59(3), 369–386 (2011)
- 9. Oliveto, P.S., Witt, C.: Erratum: Simplified Drift Analysis for Proving Lower Bounds in Evolutionary Computation. ArXiv e-prints (2012)
- Oliveto, P.S., Zarges, C.: Analysis of diversity mechanisms for optimisation in dynamic environments with low frequencies of change. In: Proc. of GECCO, pp. 837–844. ACM (2013)
- Sudholt, D.: Crossover is provably essential for the Ising model on trees. In: Proc. of GECCO, pp. 1161–1167. ACM Press (2005)