



This is a repository copy of *Empirical analysis of diversity-preserving mechanisms on example landscapes for multimodal optimisation*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/136492/>

Version: Accepted Version

---

**Proceedings Paper:**

Covantes Osuna, E. [orcid.org/0000-0001-5991-6927](https://orcid.org/0000-0001-5991-6927) and Sudholt, D. [orcid.org/0000-0001-6020-1646](https://orcid.org/0000-0001-6020-1646) (2018) Empirical analysis of diversity-preserving mechanisms on example landscapes for multimodal optimisation. In: Auger, A., Fonseca, C., Lourenço, N., Machado, P., Paquete, L. and Whitley, D., (eds.) Parallel Problem Solving from Nature – PPSN XV. PPSN 2018: 15th International Conference on Parallel Problem Solving from Nature, 08-12 Sep 2018, Coimbra, Portugal. Lecture Notes in Computer Science, 11102 . Springer Verlag , pp. 207-219. ISBN 9783319992587

[https://doi.org/10.1007/978-3-319-99259-4\\_17](https://doi.org/10.1007/978-3-319-99259-4_17)

---

The final publication is available at Springer via  
[https://doi.org/10.1007/978-3-319-99259-4\\_17](https://doi.org/10.1007/978-3-319-99259-4_17)

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**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](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Empirical Analysis of Diversity-preserving Mechanisms on Example Landscapes for Multimodal Optimisation\*

Edgar Covantes Osuna and Dirk Sudholt

University of Sheffield, Sheffield, S1 4DP, United Kingdom  
ecovantes1@sheffield.ac.uk, d.sudholt@sheffield.ac.uk

**Abstract.** Many diversity-preserving mechanisms have been developed to reduce the risk of premature convergence in evolutionary algorithms and it is not clear which mechanism is best. Most multimodal optimisation problems studied empirically are restricted to real-parameter problems and are not accessible to theoretical analysis, while theoreticians analyse the simple bimodal function TWOMAX. This paper looks to narrow the gap between both approaches. We perform an extensive empirical study involving 9 common diversity mechanisms on Jansen-Zarges multimodal function classes (Jansen and Zarges, PPSN 2016) that allow to control important problem features while still being amenable to theoretical analysis. This allows us to study functions with various degrees of multimodality and to explain the results in the light of previous theoretical works. We show which mechanisms are able to find and maintain a large number of distant optima, escape from local optima, and which fail to locate even a single peak.

**Keywords:** diversity-preserving mechanisms, evolutionary algorithms, multimodal optimisation, empirical study, theory

## 1 Introduction

Many optimisation problems are multimodal, and finding global optima or high-quality local optima can become a challenge for any optimisation algorithm [15, 17]. Evolutionary algorithms (EAs) are well suited to dealing with multimodal problems due to their use of a population. A diverse population can explore several hills in the fitness landscape simultaneously and offer several good solutions to the user, a feature desirable for decision making, in multi-objective optimisation and in dynamic optimisation. However, a major difficulty when applying EAs is that the population may converge to a sub-optimal individual before the fitness landscape is explored properly.

Many diversity-preserving mechanisms have been developed to reduce the risk of such premature convergence, including fitness sharing, clearing, avoiding

---

\* The authors would like to thank the Consejo Nacional de Ciencia y Tecnología CONACYT (the Mexican National Council for Science and Technology) for the financial support under the grant no. 409151 and registration no. 264342.

duplicates, fitness diversity, crowding methods, restricted tournament selection, and many others [4,7,16,18]. These mechanisms seek to enable EAs to visit many different regions of the search space and generate solutions that differ from those seen before [10]. Given such a variety of mechanisms to choose from, it is often not clear which mechanism is the best choice for a particular problem.

Previous empirical analyses have considered real-parameter multimodal optimisation problems [5] like the 4 one-dimensional, five-peaked, sinusoidal, multimodal functions called  $M_{1-4}$  defined in [11, Sect. 5.3]. The single variable  $x$  is restricted to the real-value range  $[0, 1]$  encoded using binary representation and decoded by interpreting the bit string as unsigned binary integer and dividing it by  $2^n - 1$ , where  $n$  is the length of the bit string. Other studies used Gray codes [15]. The drawback of real-valued encodings is that it is not obvious how phenotypic features such as local optima appear in genotype space; for example what Hamming distance local optima have and how likely it is that mutation jumps from one basin of attraction to another. This makes the analysis of the population dynamics a very challenging task for theoretical analysis.

Previous theoretical studies that considered multimodal problems [1,2,6,13] compared the expected running time of different diversity mechanisms to find both global optima on the bimodal function  $TWOMAX(x) := \{\sum_{i=1}^n x_i, n - \sum_{i=1}^n x_i\}$  that has a straightforward mapping between genotypes (bit strings) and phenotypes (number of 1-bits).  $TWOMAX$  consists of two different symmetric slopes (or branches)  $ZEROMAX$  and  $ONEMAX$  with  $0^n$  and  $1^n$  as global optima, respectively, and the goal is to evolve a population that contains both optima<sup>1</sup>. This is challenging as the two optima have the maximum possible Hamming distance. Studying  $TWOMAX$  led to insights into the capabilities and weaknesses of various diversity mechanisms (see Section 2 and Sudholt’s survey [19]), however a question left open is how diversity mechanisms deal with many local optima.

Jansen and Zarges [9] addressed the need for more general classes of functions for multimodal optimisation for both empirical and theoretical analysis by defining multimodal landscapes with straightforward binary encodings (see Section 3). We seek to narrow the gap between theory and practice by performing an empirical study on the Jansen-Zarges multimodal function classes, complementing existing rigorous theory for  $TWOMAX$  [1,2,6,13] with empirical results of more complex functions with multiple different peaks, slopes and heights. The main goal is to provide insights into the working principles of these mechanisms by testing their ability to find and maintain many local optima in the population as well as their ability to escape from local optima with different basins of attraction. We use previous theoretical results to inform the choice of algorithm parameters and to discuss in how far our empirical results agree or disagree with theoretical results obtained for  $TWOMAX$ .

---

<sup>1</sup> In [6] an additional fitness value for  $1^n$  was added to distinguish between a local optimum  $0^n$  and a unique global optimum  $1^n$ . The discussion of previous work from [6] is adapted to a  $TWOMAX$  with two optima [1,2,13] (see Table 1 and [19]).

---

**Algorithm 1** ( $\mu+1$ ) EA

---

- 1: Initialise  $P$  with  $\mu$  individuals chosen uniformly at random.
  - 2: **while** stopping criterion **not** met **do**
  - 3:   Choose  $x \in P$  uniformly at random.
  - 4:   Create  $y$  by flipping each bit in  $x$  independently with probability  $1/n$ .
  - 5:   Choose  $z \in P$  uniformly at random from all individuals with worst fitness in  $P$ .
  - 6:   **if**  $f(y) \geq f(z)$  **then**  $P = P \setminus \{z\} \cup \{y\}$ .
- 

## 2 Diversity Mechanisms & Previous Results for TwoMax

Following previous theoretical work, we consider diversity mechanisms embedded in a bare-bones EA: a  $(\mu+1)$  EA with uniform parent selection in a population of size  $\mu$ , using standard mutation (and no crossover). The offspring  $y$  replaces a worst individual  $z$  from the population if  $f(y) \geq f(z)$  (see Algorithm 1).

Table 1 summarises previous work for the  $(\mu+1)$  EA with diversity mechanisms on TWOMAX (details for each  $(\mu+1)$  EA variant can be found in the respective publications and in [19]). Some mechanisms succeed in finding both optima on TWOMAX efficiently, that is, in (expected) time  $O(\mu n \log n)$ . Others have a very low success probability. Friedrich, Oliveto, Sudholt, and Witt [6] showed that the plain  $(\mu+1)$  EA (PL, Algorithm 1) is not able to maintain individuals on both branches for a long time; the whole population is likely to converge to one of the two peaks [6, Theorem 1].

Introducing other simple mechanisms to the  $(\mu+1)$  EA like *avoiding genotype duplicates (NGD)*, where (after initialisation) identical copies of individuals are prevented from entering the population [6, Algorithm 2], and *avoiding fitness duplicates (NFD)*, rejecting individuals with the same fitness [6, Algorithm 3], are not able to prevent the extinction of one branch, ending with the population converging to one optimum with high probability [6, Theorem 2 and 3, resp.].

In the  $(\mu+1)$  EA with *probabilistic crowding (PC)* [2, Algorithm 1], an offspring competes with its parent and the survivor is chosen with a probability proportional to their fitness [12]. Covantes Osuna and Sudholt showed that this mechanism is unable to evolve solutions of significantly higher fitness than that obtained during initialisation (or, equivalently, through random search), even when given exponential time [2, Theorem 2.2]. The reason is that fitness-proportional selection between parent and offspring results in an almost uniform choice as both have very similar fitness, hence fitness-proportional selection degrades to uniform selection for replacement. In *deterministic crowding (DC)*, the offspring competes against its parent and replaces it if the offspring is at least as good [11]. For the  $(\mu+1)$  EA with deterministic crowding [6, Algorithm 4], this mechanism with a sufficiently large population is able to reach both optima with high probability in expected time  $O(\mu n \log n)$  [6, Theorem 4].

In *restricted tournament selection (RTS)*, for every offspring created, RTS selects uniformly at random (u. a. r.)  $w$  (*window size*) members from the population with replacement. Each offspring competes with the closest element from this set and the offspring replaces it if its fitness is at least as good [8]. For the

Table 1: Overview of runtime analyses for the  $(\mu+1)$  EA with diversity mechanisms on TWOMAX, adapted from [2]. The success probability is the probability of finding both optima within (expected) time  $O(\mu n \log n)$ . Conditions include restrictions on the population size  $\mu$ , the sharing/clearing radius  $\sigma$ , the niche capacity  $\kappa$ , window size  $w$ , and  $\mu' := \min(\mu, \log n)$ .

Diversity Mechanism	Success prob.	Conditions
PL Plain $(\mu+1)$ EA [6]	$o(1)$	$\mu = o(n/\log n)$
NGD No Genotype Duplicates [6]	$o(1)$	$\mu = o(\sqrt{n})$
NFD No Fitness Duplicates [6]	$o(1)$	$\mu = \text{poly}(n)$
PC Probabilistic Crowding [2]	$2^{-\Omega(n)}$	all $\mu$
DC Deterministic Crowding [6]	$1 - 2^{-\mu+1}$	all $\mu$
RTS Restricted Tournament Selection [2]	$\geq 1 - 2^{-\mu'+3}$	$w \geq 2.5\mu \ln n$
PFS Population-based Fitness Sharing* ( $\sigma = n/2$ ) [6]	1	$\mu \geq 2$
FS Individual-based Fitness Sharing* ( $\sigma = n/2$ ) [13]	1	$\mu \geq 3$
CL Clearing ( $\sigma = n/2$ ) [1]	1	$\mu \geq \kappa n^2$

\* Fitness sharing uses phenotypic sharing based on the number of ones.

$(\mu+1)$  EA with RTS [2, Algorithm 2], the mechanism succeeds in finding both optima of TWOMAX in the same way as deterministic crowding, provided that  $w$  is chosen large enough [2, Theorem 3.1]. However, if  $w$  is too small, then it cannot prevent one branch taking over the other, leading to exponential running times with high probability [2, Theorem 3.4].

*Fitness sharing* derates the real fitness of an individual by an amount that represents the similarity to other population members. A *population-based fitness sharing (PFS)* approach [6, Algorithm 5], constructing the best possible new population amongst parents and offspring is able to find both optima in expected time  $O(\mu n \log n)$  for any population size  $\mu \geq 2$  [6, Theorem 5]. A drawback of this approach is that examining all possible new populations is computationally expensive. The conventional *fitness sharing (FS)*, where selection is based on individuals, was studied by Oliveto, Sudholt, and Zarges [13, Algorithm 1]. Population size  $\mu = 2$  is not sufficient to find both optima in polynomial time; the success probability is only  $1/2 - \Omega(1)$  [13, Theorem 1]. However, with  $\mu \geq 3$ , the  $(\mu+1)$  EA finds both optima in expected time  $O(\mu n \log n)$  [13, Theorem 3]. In all the above results, fitness sharing used a phenotypic distance: the distance between two search points  $x$  and  $y$  is the absolute difference in their number of ones. This choice is tailored to TWOMAX and is not applicable in our scenario. Hence our experiments must rely on fitness sharing with genotypic distances (Hamming distance), for which no runtime analyses are available.

In *clearing (CL)*, individuals are sorted in decreasing fitness and are processed in this order. Each individual is compared against other individuals according to its fitness with distance  $< \sigma$  (*clearing radius*) which determines if both individuals belong to the same subpopulation (niche) or not. Then, the procedure iterates through all remaining individuals (i. e., those with lower or equal fitness) that

havent been cleared yet, until  $\kappa$  (*niche capacity*) best individuals (also called *winners*) have been found, and all remaining individuals from the same niche are cleared to the minimum fitness value possible [14]. Finally, the individuals with best fitness are selected (set of winners) and individuals coming from the new generation are preferred [1, Algorithm 1 and 2]. Clearing, with a clearing radius of  $\sigma = n/2$ , niche capacity  $\kappa = 1$ , and  $\mu \geq \kappa n^2$  is able to find both optima in expected time  $O(\mu n \log n)$  [1, Theorem 5.6].

### 3 Jansen-Zarges Multimodal Function Classes

Jansen and Zarges [9] introduced several problem classes spanned by  $k$  peaks  $p_1, p_2, \dots, p_k \in \{0, 1\}^n$  for an arbitrary number  $k \in \mathbb{N}$  of peaks. Each peak  $i$  has a position  $p_i \in \{0, 1\}^n$ , a slope  $a_i \in \mathbb{R}^+$ , and an offset  $b_i \in \mathbb{R}_0^+$ . The fitness value of a search point depends on peaks in its vicinity as defined as follows.

**Definition 1 (Definition 3 in [9]).** *Let  $k \in \mathbb{N}$  and  $k$  peaks  $(p_1, a_1, b_1), (p_2, a_2, b_2), \dots, (p_k, a_k, b_k)$  be given, then*

- $\text{JZ}_1(x) := a_{\text{cp}(x)} \cdot G(x, p_{\text{cp}(x)}) + b_{\text{cp}(x)}$ , *called nearest peak function,*
- $\text{JZ}_2(x) := \max_{i \in \{1, 2, \dots, k\}} a_i \cdot G(x, p_i) + b_i$ , *called weighted nearest peak function,*

where  $\text{cp}(x) := \arg \min_{i \in \{1, 2, \dots, k\}} H(x, p_i)$  is defined by the closest peak to a search point, and  $G(x, p_i) := n - H(x, p_i)$  indicates the proximity of  $x$  to  $p_i$ .

For the nearest peak function,  $\text{JZ}_1(x)$ , the fitness of a search point  $x$  is determined by the proximity to the closest peak  $i = \text{cp}(x)$  along with its slope  $a_i$  and its offset  $b_i$ . In cases where multiple  $i$  minimise  $H(x, p_i)$ ,  $i$  should additionally maximise  $a_i \cdot G(x, p_i) + b_i$ .

The weighted nearest peak function,  $\text{JZ}_2(x)$ , takes the height of peaks into account. The peak  $i$  yielding the largest value  $a_i \cdot G(x, p_i) + b_i$  determines the function value. The bigger the height of a peak, the bigger its influence on the search space in comparison to smaller peaks. Note that, in case of equal slopes  $a_1 = \dots = a_k$  and equal heights  $b_1 = \dots = b_k$ , both functions  $\text{JZ}_1$  and  $\text{JZ}_2$  using parameters  $a_1, \dots, a_k, b_1, \dots, b_k$  are identical as for  $\text{JZ}_2$  the maximum over all terms  $a_i \cdot G(x, p_i) + b_i$  for all  $1 \leq i \leq k$  is attained for the closest peak  $i = \text{cp}(x)$ .

**Theorem 2.** *For  $\text{JZ}_1$  and  $\text{JZ}_2$  using the same parameters  $a_1 = \dots = a_k$  and  $b_1 = \dots = b_k$  we have  $\text{JZ}_1 = \text{JZ}_2$ .*

In the case of two peaks  $p_1$  and  $p_2$ , if these peaks are complementary, that is,  $p_2 = \overline{p_1}$ , then  $\text{JZ}_1$  and  $\text{JZ}_2$  generalise the TWOMAX function, with TWOMAX being the special case of  $p_1 = 0^n, p_2 = 1^n, a_1 = a_2 = 1$  and  $b_1 = b_2 = 0$  [9]. This setting was studied for the  $(\mu+1)$  EA with clearing in [3].

We consider peaks being placed independently and u. a. r., as this strategy is simple, fair, and it scales towards an arbitrary number of peaks. The slopes are chosen equal to 1 for all peaks for the sake of simplicity. Even though the peaks are placed randomly, if the peaks have moderately similar heights, the resulting

fitness landscape has a clear structure: with high probability all peaks are local optima, and all search points within a Hamming ball of radius  $\Omega(n)$  belong to a peak's basin of attraction. This holds for both functions  $\text{JZ}_1$  and  $\text{JZ}_2$  as they have equal fitness values within the mentioned Hamming balls (but may have different values on other search points).

**Theorem 3.** *Assume  $k$  peaks  $p_1, \dots, p_k$  chosen independently and u. a. r. from  $\{0, 1\}^n$ . If  $a_1 = \dots = a_k = 1$  and  $\max_{1 \leq i \leq k} b_i - \min_{1 \leq i \leq k} b_i \leq cn$  for a constant  $c < 1/2$  then with probability  $1 - k^2 e^{-\Omega(n)}$  for radius  $r := (1/2 - c)/3 \cdot n$  we have:*

1. all  $k$  peaks  $p_1, \dots, p_k$  are local optima in both  $f_1$  and  $f_2$ ,
2. for all  $1 \leq i \leq k$ , all search points in  $\mathcal{B}_i := \{x \mid \text{H}(x, p_i) \leq r\}$ , a Hamming ball of radius  $r$  around  $p_i$ , belong to the basin of attraction of  $p_i$  with respect to both  $\text{JZ}_1$  and  $\text{JZ}_2$ , that is, there is a Hamming path from  $x$  to  $p_i$  on which the values of  $\text{JZ}_1$  and  $\text{JZ}_2$  are strictly increasing, and
3. for all search points  $x \in \bigcup_{i=1}^k \mathcal{B}_i$ ,  $\text{JZ}_1(x) = \text{JZ}_2(x)$ .

*Proof.* Assume without loss of generality  $\min_{1 \leq i \leq k} b_i = 0$  (as adding a fixed value does not affect the problem structure), hence  $b_i \leq cn$  for all  $1 \leq i \leq k$ .

By Chernoff bounds, the probability that two different peaks will have Hamming distance at most  $n/2 - r$  is  $e^{-\Omega(n)}$ . By the union bound, the probability that this holds for any pair of peaks is at most  $k^2 \cdot e^{-\Omega(n)}$ . We assume in the following that every two peaks have a Hamming distance larger than  $n/2 - r$ .

Now consider a search point  $x \in \mathcal{B}_i$ , that is,  $\text{H}(x, p_i) \leq r$ . Since  $r \leq n/6$  we have  $n/2 \geq 3r$ , and thus for all  $j \neq i$  we have  $\text{H}(x, p_j) \geq \text{H}(p_i, p_j) - \text{H}(x, p_i) > n/2 - r - r \geq r \geq \text{H}(x, p_i)$ . So  $p_i$  is a unique closest peak,  $\text{cp}(x) = i$ . By definition of  $\text{JZ}_1$ , the second statement follows for  $\text{JZ}_1$  as on every shortest Hamming path from  $x$  to  $p_i$ , subsequently decreasing the Hamming distance to  $p_i$  increases the fitness by  $a_i = 1$ . Since  $r \geq 1$  if  $n$  is large enough,  $p_i$  is a local optimum for  $\text{JZ}_1$ .

It only remains to show the third statement as then the first two statements also apply to  $\text{JZ}_2$ . To prove that  $\text{JZ}_2(x) = \text{JZ}_1(x)$  for  $x \in \mathcal{B}_i$ , we need to show that the maximum over terms  $a_j \cdot \text{G}(x, p_j) + b_j = n - \text{H}(x, p_j) + b_j$  from the definition of  $\text{JZ}_2$  is attained for  $j = i$ . We have  $n - \text{H}(x, p_i) + b_i \geq n - r$  as  $\text{H}(x, p_i) \leq r$  and  $b_i \geq 0$ . For  $j \neq i$  we have  $n - \text{H}(x, p_j) + b_j < n/2 + 2r + cn$  as  $b_j \leq cn$  and  $\text{H}(x, p_j) \geq \text{H}(p_i, p_j) - \text{H}(x, p_i) > n/2 - r - r = n/2 - 2r$ . Noting that  $n/2 + 2r + cn = n/2 + 3r + cn - r = n/2 + (n/2 - cn) + cn - r = n - r$  establishes  $n - \text{H}(x, p_j) + b_j \leq n - \text{H}(x, p_i) + b_i$  and hence  $\text{JZ}_2(x) = \max_{1 \leq j \leq k} (n - \text{H}(x, p_j) + b_j) = n - \text{H}(x, p_i) + b_i = \text{JZ}_1(x)$ .  $\square$

## 4 Experimental Analysis

For the experimental analysis we test each of the algorithms from Table 1 (referred to by the acronyms defined in the first column) on Jansen-Zarges multimodal function classes. We consider a problem size  $n = 100$ , genotypic distance for all algorithms that require a dissimilarity measure and stop runs after  $10\mu n \ln n$  generations. This time limit is motivated by [2, Lemma 3.3] stating

that, loosely speaking,  $2e\mu n \ln n \approx 5.44\mu n \ln n$  generations are sufficient to perform hill climbing on two peaks with high probability.

The experimental framework is divided in 3 experimental set-ups. In Sect. 4.1 we assess the ability of each mechanism to find many peaks with equal height, and in Sect. 4.2, we assess the ability of each mechanism to maintain the population diversity when considering peaks with different heights to yield global and local optima. For both sections, the number of peaks was increased exponentially as  $k = \{2, 4, 8, \dots, 64\}$ . For each  $k$ , we generated 100 different instances choosing  $k$  peaks u. a. r. from  $\{0, 1\}^n$ . In each experiment, all algorithms are tested on the same set of 100 instances to ensure a fair comparison. The challenge for each mechanism is to find and maintain as many peaks as possible before reaching the  $10\mu n \ln n$  generations; we record the fraction of the peaks found. The population size is chosen large enough ( $\mu = 100$ ) to be able to accommodate all peaks.

The analysis in Sect. 4.3 is inspired by [3, Sect. 7.3] and focusses on landscapes with two peaks. In this section we take a closer look at the ability of the diversity mechanisms to deal with different basins of attraction, including a wider range of two-peaked landscapes than the ones likely to be generated by placing peaks u. a. r.. The goal is to observe which mechanisms are able to escape from local optima by tunnelling through the fitness valley that separates two peaks. We choose  $\mu = 32$  as in [3, Sect. 7.3] and also consider the same two initialisations: the standard uniform random initialisation and biased initialisation where the whole population is initialised with copies of one peak ( $0^n$  for TWOMAX). Biased initialisation is used in order to observe how the mechanisms are able to escape from a local optimum and how fast it is compared to a random initialisation.

Based on the theoretical analysis in [2] we define the window size  $w = 2.5\mu \ln n$  for RTS. We know from [6, 13] that both FS approaches with phenotypic sharing and  $\sigma = n/2$  are always efficient on TWOMAX but no theory for genotypic sharing is available. Preliminary experiments for genotypic sharing and  $\sigma = n/2$  on TWOMAX yielded poor results; however with  $\sigma = n$  (which implies that all individuals always share fitness) both peaks were found in most runs. This makes sense on other landscapes as well as if  $\sigma$  is set smaller than the radius or basin of attraction around a local optimum, then FS is unable to push individuals away from said local optimum. Thus it seems best to err on the side of choosing  $\sigma$  too large rather than too small.

For CL the situation is different. If  $\sigma$  is chosen too large, such that there are several optima within a distance of  $\sigma$ , then global optima may be cleared, making it impossible to maintain many optima in the population. So for CL it seems best to err on the side of choosing  $\sigma$  too small rather than too large. We choose  $\sigma = n/3$  for Sect(s). 4.1 and 4.2 as with high probability every two different peaks will have a Hamming distance larger than  $n/3$  (cf. Theorem 3). For Sect. 4.3, we use the recommendation  $\sigma = \min\{H(p_1, p_2), n/2\}$  from [3].

#### 4.1 Finding Peaks of Equal Height

We consider the JZ<sub>1</sub> function with equal slopes  $a_1 = \dots = a_k = 1$  and offsets  $b_1 = \dots = b_k = 0$ . We know from Theorem 2 that with equal parameters



$JZ_1 = JZ_2$ . In Fig. 1 (blue/left box plots), we show the fraction of peaks obtained in each of the 100 instances and its variance for each choice of  $k$ .

As can be seen, the PL, NGD and NFD perform poorly; these have already been proven to perform poorly on TWOMAX [6]. PC as predicted in [2] is not able to find even one peak. FS performs best for an intermediate number of peaks,  $k \in \{4, 8, 16\}$ , but still far worse than the best mechanisms. This is in contrast to theoretical results [6, 13] where FS in both variants was shown to be very effective on TWOMAX. These differences may be down to the differences between TWOMAX and  $JZ_1$  with random peaks and/or they may be caused by the differences between phenotypic and genotypic sharing. Interestingly, PFS performs far worse than the conventional FS. This is surprising as PFS uses a significant amount of computation time to search for the best possible population (in terms of shared fitness) it can create out of all parents and offspring, hence we would have expected it to perform better than FS. A possible explanation for the poor performance of FS is that even when the population is able to locate basins of attraction of several peaks, we found several individuals scattered around each peak, apparently repelling each other and preventing each other from reaching the peak.

Finally, DC, RTS and CL perform surprisingly well: they find all optima most of the time for  $k \leq 16$ , and find most optima for  $k = 32$ . Only for a large number of  $k = 64$  peaks, performance deteriorated to around 80% of peaks found. This deterioration is not surprising as the population size was fixed to  $\mu = 100$ . RTS with  $w = 2.5\mu \ln n$  seems to behave similarly to DC as predicted in [2].

## 4.2 Finding Peaks with Different Height

For this case we make use of the  $JZ_2$  function with  $a_1 = \dots = a_k = 1$  and  $b_1 \dots b_k$  chosen independently and u. a. r. from  $[0, 1, \dots, n/3]$ . This range is motivated by Theorem 3, as here two peaks differ in their heights by at most  $n/3$ , choosing the leading constant  $c := 1/3$  as the simplest constant smaller than  $1/2$ . Theorem 3 then yields that all search points within Hamming balls of radius  $n/18$  centred at a peak are located in the peak's basin of attraction. Fig. 1 (red/centre box plots) shows the fraction of peaks obtained in each of the 100 instances and its variance for each  $k = \{2, 4, 8, \dots, 64\}$  peaks. To gauge the quality of the peaks found, we also plot the normalised best fitness found (green/right box plots), formally  $f_i^*/\text{opt}_i$  where  $f_i^*$  is the fitness of the best peak found on instance  $i$  and  $\text{opt}_i$  is the optimal value of instance  $i$ .

In this setting PL, NGD and NFD manage to find the global optimum in up to 80% of instances. This suggests that on this function class it is fairly easy to find a global optimum. However, they rarely find more than one peak, hence they seem to suffer from premature convergence. PC continues to show the worst performance of all mechanisms. PFS and FS find fewer peaks on  $JZ_2$  compared to  $JZ_1$ . This makes sense since the former setting is more difficult than the latter; both mechanisms seem to suffer from the issues mentioned in Sect. 4.1.

Finally, DC, RTS and CL also find fewer peaks due to the difficulty of this setting, but still show the best performance of all mechanisms analysed in this

paper and they manage to find the global optimum in all instances. For  $k \leq 8$  is not possible to always find all peaks any more, but they still manage to find at least 50% of the peaks. Then, for  $k \geq 16$  the performance deteriorates in such a way that it is not possible to reach any more 50% of the peaks but still the mechanisms manage to find some of the peaks. The general cause of the drop in the performance seems to be that all mechanisms struggle to escape from the optimum found, also that low-quality optimums are being dropped when better peaks have been found.

### 4.3 Escaping from Local Optima

Theorem 3 and its proof suggest that when peaks are chosen u. a. r., they will have a Hamming distance close to  $n/2$ . We would like to investigate how the diversity mechanisms behave if peaks have different Hamming distances. Following [3], we focus on two peaks and vary their Hamming distance between 1 and  $n$  by choosing  $p_1 = 0^n$  and  $p_2 \in \{0^{n-1}1, 0^{n-2}1^2, \dots, 1^n\}$ , along with  $a_1 = a_2 = 1$  and  $b_1 = b_2 = 0$ . As argued in [3], this captures the performance across all possible  $JZ_1$  functions with two complementary peaks and the given slopes and offsets. In particular, it includes many functions that only have an exponentially small probability to be generated when choosing peaks independently and u. a. r. With biased initialisation, the algorithms have to find the other optimum by tunnelling through the fitness valley that separates these two peaks. This is a much harder task compared to hill climbing on various hills, where the aim is for the population to maintain a good spread over the search space.

We use the set-up and empirical data for CL from [3] and report the average number of generations of 100 runs, with two stopping criteria: both optima have been found or  $t = 10\mu n \ln n$  generations were reached.

From Fig. 2a all mechanisms are effective when the Hamming distance is so small that the peaks are very close together such that the second peak can be found by a mutation of the first peak found (except for PC, that is not able to reach a single peak). But as the distance increases, the time for some mechanisms increases rapidly; they are inefficient on all non-trivial settings. DC and RTS seem to be agnostic of Hamming distances as they show a very stable and equal performance across the whole range of Hamming distances. This makes sense as DC climbs up both peaks with equal probability (cf. the analysis on TWOMAX [6]) and RTS behaves similarly to DC. CL is very effective and only mildly worse than DC and RTS. We see that for FS with genotypic sharing is only effective if the peaks have a Hamming distance that is very close to  $n$  or trivially small. For intermediate values, FS fails badly.

With biased initialisation (Fig. 2b), CL is the only mechanism able to escape from local optima with different basins of attraction. As shown theoretically in [3], this is because cleared individuals are able to explore the fitness landscape by performing a random walk. We know from [6, 13] that both FS approaches with phenotypic sharing and sharing radius  $\sigma = n/2$  are able to escape from local optima as well, if the two peaks are complementary. With genotypic sharing both

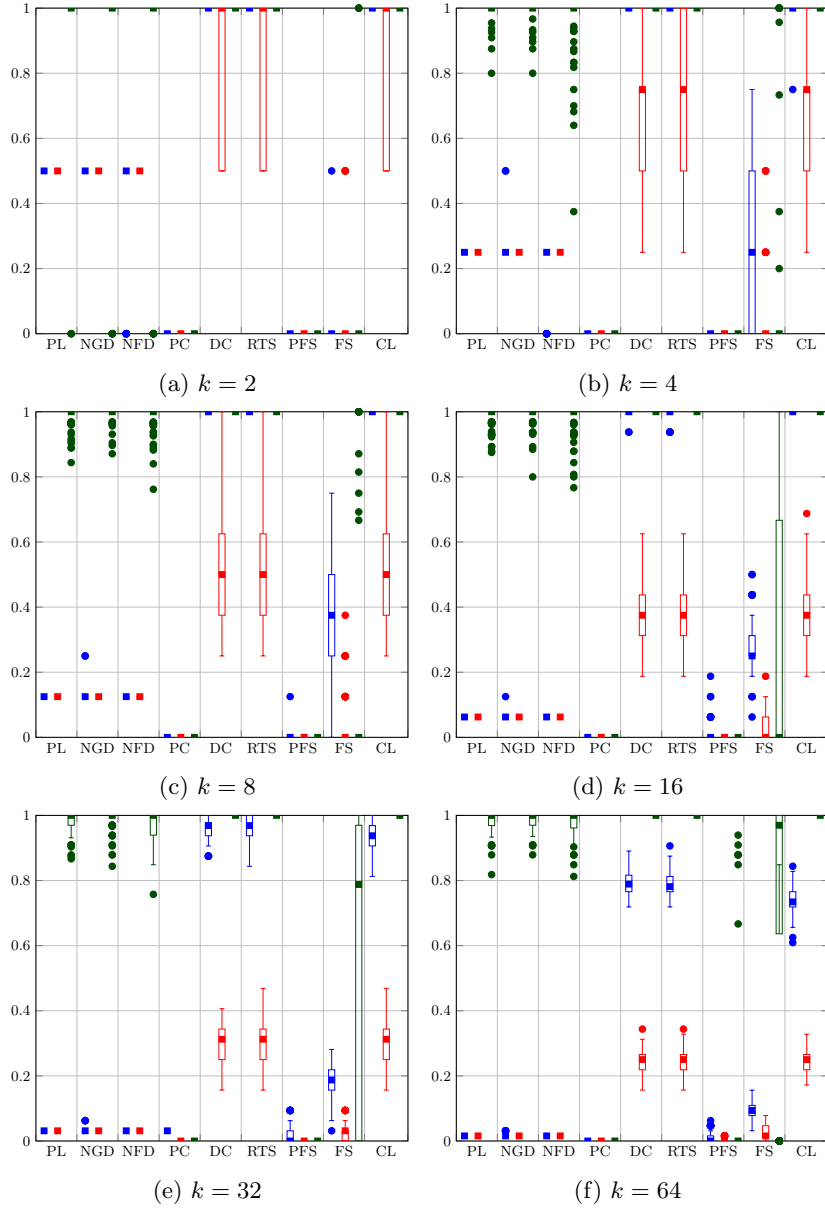


Fig. 1: Experimental results for all  $(\mu+1)$  EA variants from Table 1 among 100 instances generated u. a. r. for each number of peaks  $k = \{2, 4, 8, \dots, 64\}$ ,  $\mu = 100$  and  $n = 100$ , stopping runs after  $10\mu n \ln n$  generations. Blue/left: fraction of peaks found on  $JZ_1$  with peaks of equal height. Red/centre: fraction of peaks found on  $JZ_2$  with peaks with different heights,  $b_1 \dots b_k$  chosen u. a. r. from  $\{0, 1, \dots, n/3\}$ . Green/right: normalised best fitness found on  $JZ_2$  experiments. Squares indicate median values.

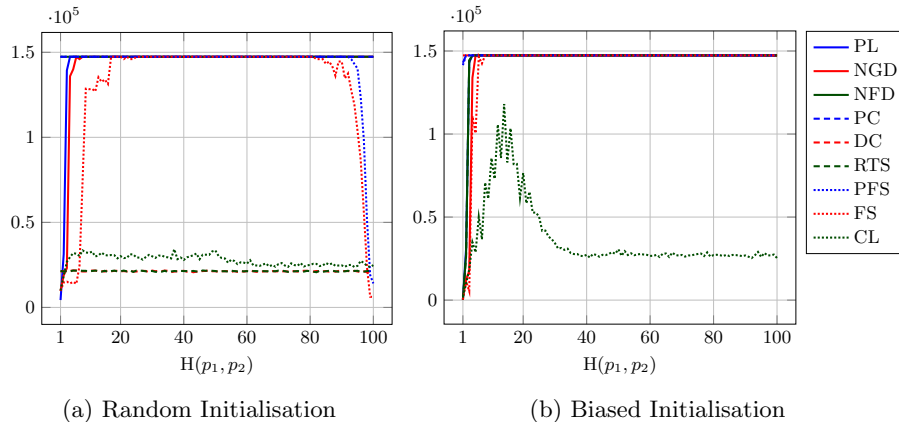


Fig. 2: The average number of generations among 100 runs for finding both peaks  $p_1 = 0^n$  and  $p_2 = \{0^{n-1}1, 0^{n-2}1^2, \dots, 1^n\}$  on the fitness landscape defined by  $JZ_2$  with  $a_1 = a_2 = 1$  and  $b_1 = b_2 = 0$  or  $t = 10\mu n \ln n$  generations were reached, for all  $(\mu+1)$  EA variants mentioned in Table 1, using  $n = 100$  and  $\mu = 32$ . Results for both random and biased initialisation are shown.

FS approaches perform very poorly and seem unable to escape from local optima. Also the other mechanisms fail as they are unable to accept worse search points.

## 5 Conclusions

We have performed an extensive empirical study involving 9 common diversity mechanisms on Jansen-Zarges multimodal function classes, covering various degrees of multimodality from 2 to 64 peaks and peaks having equal or different heights, reflected in their basins of attraction. Our results show that the plain  $(\mu+1)$  EA, the simple mechanisms: avoiding genotype and fitness duplicates cannot maintain subpopulations on several peaks; once a peak has been found it seems impossible to escape from such a peak. Probabilistic crowding shows a terrible performance as it is unable to locate even a single peak. These findings are in line with theoretical results on TWOMAX [2, 6].

Previous theoretical results have shown that both fitness sharing approaches are always efficient on TWOMAX if phenotypic distances are being used and parameters are set appropriately [6, 13]. This includes the ability to climb down a peak and to tunnel through fitness valleys to reach other niches. Unfortunately this is not the case for fitness sharing with genotypic distance. Only when the peaks have a Hamming distance that is trivially small or very close to  $n$  they seem to be effective; for any other intermediate case they show a poor performance.

Deterministic crowding, restricted tournament selection and clearing perform well for peaks with the same slope and height, much better than all other diversity mechanisms. Only for large numbers of peaks ( $k = 64$ ) and different heights

the performance starts to deteriorate. Finally, only clearing has shown the ability to escape from local optima since all other mechanisms seem unable to accept worse search points or unable to tunnel through fitness valleys.

## References

- [1] Covantes Osuna, E., Sudholt, D.: Analysis of the clearing diversity-preserving mechanism. In: Proc. of FOGA '17, ACM (2017) 55–63
- [2] Covantes Osuna, E., Sudholt, D.: Runtime analysis of probabilistic crowding and restricted tournament selection for bimodal optimisation. In: Proc. of GECCO '18. (2018) To appear. Preprint available from <http://arxiv.org/abs/1803.09766>.
- [3] Covantes Osuna, E., Sudholt, D.: On the runtime analysis of the clearing diversity-preserving mechanism. *Evol. Comput.* (Accepted with revisions. Preprint available from <http://arxiv.org/abs/1803.09715>)
- [4] Črepinšek, M., Liu, S.H., Mernik, M.: Exploration and exploitation in evolutionary algorithms: A survey. *ACM Comput. Surv.* **45**(3) (2013) 35:1–35:33
- [5] Das, S., Maity, S., Qu, B.Y., Suganthan, P.: Real-parameter evolutionary multimodal optimization — a survey of the state-of-the-art. *Swarm and Evolutionary Computation* **1**(2) (2011) 71–88
- [6] Friedrich, T., Oliveto, P.S., Sudholt, D., Witt, C.: Analysis of Diversity-preserving Mechanisms for Global Exploration. *Evol. Comput.* **17**(4) (2009) 455–476
- [7] Glibovets, N.N., Gulayeva, N.M.: A review of niching genetic algorithms for multimodal function optimization. *Cybernetics and Systems Analysis* **49**(6) (2013) 815–820
- [8] Harik, G.R.: Finding Multimodal Solutions Using Restricted Tournament Selection. In: Proc. of the 6th ICGA, Morgan Kaufmann Publishers Inc. (1995) 24–31
- [9] Jansen, T., Zarges, C.: Example landscapes to support analysis of multimodal optimisation. In: Proc. of PPSN XIV, Springer (2016) 792–802
- [10] Lozano, M., García-Martínez, C.: Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: Overview and progress report. *Computers & Operations Research* **37**(3) (2010) 481–497
- [11] Mahfoud, S.W.: Niching methods for genetic algorithms. PhD thesis, University of Illinois at Urbana-Champaign (1995)
- [12] Mengersheol, O., Goldberg, D.: Probabilistic crowding: Deterministic crowding with probabilistic replacement. In: Proc. of GECCO '99. (1999) 409–416
- [13] Oliveto, P.S., Sudholt, D., Zarges, C.: On the runtime analysis of fitness sharing mechanisms. In: Proc. of PPSN XIII, Springer (2014) 932–941
- [14] Pétrowski, A.: A clearing procedure as a niching method for genetic algorithms. In: Proc. of ICEC '96. (1996) 798–803
- [15] Sareni, B., Krahenbuhl, L.: Fitness sharing and niching methods revisited. *IEEE Transactions on Evolutionary Computation* **2**(3) (1998) 97–106
- [16] Shir, O.M.: Niching in evolutionary algorithms. In: Handbook of Natural Computing. Springer Berlin Heidelberg (2012) 1035–1069
- [17] Singh, G., Deb, K.: Comparison of multi-modal optimization algorithms based on evolutionary algorithms. In: Proc. of GECCO '06, ACM (2006) 1305–1312
- [18] Squillero, G., Tonda, A.: Divergence of character and premature convergence: A survey of methodologies for promoting diversity in evolutionary optimization. *Information Sciences* **329** (2016) 782–799
- [19] Sudholt, D.: The Benefits of Population Diversity in Evolutionary Algorithms: A Survey of Rigorous Runtime Analyses. (2018) <http://arxiv.org/abs/1801.10087>.