

A First Runtime Analysis of the PAES-25: An Enhanced Variant of the Pareto Archived Evolution Strategy

Andre Opris

University of Passau, Passau, Germany

ABSTRACT

This paper presents a first mathematical runtime analysis of PAES-25, an enhanced version of the original Pareto Archived Evolution Strategy (PAES) coming from the study of telecommunication problems over two decades ago to understand the dynamics of local search of MOEAs on many-objective fitness landscapes. We derive tight expected runtime bounds of PAES-25 with one-bit mutation on *m*-LEADINGONESTRAILINGZEROS until the entire Pareto front is found: $\Theta(n^3)$ iterations if $m = 2$, $\Theta(n^3 \log^2(n))$ iterations if $m = 4$ and $\Theta(n \cdot (2n/m)^{m/2} \log(n/m))$ iterations if $m > 4$ where n is the problem size and m the number of objectives. To the best of our knowledge, these are the first known tight runtime bounds for an MOEA outperforming the best known upper bound of $O(n^{m+1})$ for (G)SEMO on *m*-LOTZ when $m \geq 4$. We also show that *archivers*, such as the *Adaptive Grid Archiver* (AGA), *Hypervolume Archiver* (HVA) or *Multi-Level Grid Archiver* (MGA), help to distribute the set of solutions across the Pareto front of *m*-LOTZ efficiently. We also show that PAES-25 with standard bit mutation optimizes the bi-objective LEADINGONESTRAILINGZEROS benchmark in expected $O(n^4)$ iterations, and we discuss its limitations on other benchmarks such as ONEMINMAX or COUNTINGONESCOUNTINGZEROS.

CCS CONCEPTS

• Theory of computation → Evolutionary algorithms.

KEYWORDS

Evolutionary multiobjective optimization, runtime analysis, PAES-25, archiving

ACM Reference Format:

Andre Opris. 2025. A First Runtime Analysis of the PAES-25: An Enhanced Variant of the Pareto Archived Evolution Strategy. In *Foundations of Genetic Algorithms XVIII (FOGA '25)*, August 27–29, 2025, Leiden, Netherlands. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3729878.3746618>

1 INTRODUCTION

In many real-world scenarios, optimization involves handling multiple objectives that often conflict with each other. A common approach to tackle such multi-objective optimization problems is to generate a diverse set of Pareto-optimal solutions. These are solutions where no objective can be improved without degrading at

least one other. A decision maker can then select the most suitable trade-off based on their preferences. Owing to their population-based nature, multi-objective evolutionary algorithms (MOEAs) have become essential tools and are widely applied across various domains in practice [7, 38].

The mathematical runtime analysis of MOEAs began roughly two decades ago [20, 32, 45]. In recent years, this field has experienced significant growth, with studies focusing on well-known algorithms such as (G)SEMO, NSGA-II, NSGA-III, SMS-EMOA, and SPEA2 [4, 30, 40, 41, 44, 47]. Additionally, research has explored how MOEAs can be applied to combinatorial optimization problems like minimum spanning trees [6], subset selection problems [13] or submodular optimization [43]. A useful dichotomy of how to analyze such algorithms theoretically can be found in [11].

Another notable MOEA is the *Pareto Archived Evolution Strategy* (PAES) algorithm, originally proposed by Knowles and Corne [25] coming from the study of telecommunication problems. PAES is a (1+1) evolutionary strategy that performs mutation on a single solution, referred to as the *current solution*. To identify non-dominated solutions, the algorithm maintains an archive of previously found non-dominated solutions, enabling it to obtain many different trade-offs. Its size can be limited by a threshold L smaller than the Pareto front. When the archive is full, so-called *archivers* decide whether to accept mutated solutions as the new current solution and which solution to be removed in order to preserve diversity. As this algorithm maintains and mutates a single current solution instead of selecting a solution from a population, it is very simple and particularly well-suited for solving problems relying on multi-objective local search. For instance, in [42] it is proven that this algorithm using one-bit mutation beats SEMO on a problem called PATH by a factor of n . A further key advantage of PAES is its potential to scale better to many-objective problems compared to traditional MOEAs. This is primarily because it does not rely on selective pressure, which tends to deteriorate as the number of objectives increases and the Pareto front grows exponentially [40, 41, 47, 50]. Due to its strong empirical performance across a variety of benchmark problems [25], PAES has gained significant traction ([25] and [26] together have around 5000 citations).

However, PAES discards a mutated current solution if the archive already contains such a solution with the same fitness, which leads to some limitations. In [42] it is rigorously shown that it fails with overwhelming probability to find the entire Pareto front of the LOTZ benchmark, an easy pseudo-Boolean function. Additionally, it is shown in [25] that it performs empirically poorly on more complex pseudo-Boolean functions like LITZ, sLITZ and FRITZ, alternative multi-objective generalizations of the LOTZ benchmark. To address these shortcomings, Knowles and Liefooghe [28] proposed an enhanced version of PAES, called PAES-25. This updated algorithm has slightly different acceptance criteria. It adds solutions

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

FOGA '25, August 27–29, 2025, Leiden, Netherlands

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1859-5/2025/08
<https://doi.org/10.1145/3729878.3746618>

to the archive and accepts mutated current solutions as the new current solution, even if one with the same fitness already exists in the archive which is then replaced. This enables progress on fitness plateaus. Empirical results by Knowles and Liefooghe [28] suggest that PAES-25 not only progresses more quickly towards the Pareto front but also discovers a significantly larger fraction of the front on benchmarks such as LOTZ, LITZ, sLITZ and FRITZ compared to the standard PAES.

Despite these practical results, a rigorous theoretical understanding of PAES remains limited. Apart from [42], no formal runtime analysis existed for either the original PAES or its updated variant PAES-25. Consequently, it is still unclear under which conditions PAES-25 performs well, and what structural features of problems contribute to its success. Addressing these open questions could offer valuable guidance for designing more efficient variants of PAES-25 and advancing its applicability in practice.

Our contribution: In this paper we make a significant step towards such an understanding and provide the first runtime analysis of PAES-25 on different benchmark problems. Our three main contributions are detailed as follows.

Firstly, we analyze PAES-25 with one bit mutation (i.e. a single bit chosen uniformly at random is flipped) on m -LOTZ. We show that if the archive size is sufficiently large, at least equal to the size of the Pareto front and possibly larger depending on the archiving strategy, then the algorithm efficiently discovers all Pareto optimal search points. Specifically, the number of iterations required is $\Theta(n^3)$ for $m = 2$, $\Theta(n^3 \log^2 n)$ for $m = 4$, and $\Theta(n \cdot (2n/m)^{m/2} \log(n/m))$ for $m > 4$. The key insight is that the current solution reaches the Pareto front easily and then performs a random walk on the Pareto front which can be described by a suitable random walk on a grid graph, the product graph of the single path graph. All Pareto-optimal solutions are found if all nodes in the grid graph are visited, and estimating the cover time of product graphs is a well-known task (see for example [23]). To the best of our knowledge, these are the first known tight runtime bounds in the many-objective setting and surprisingly, if $m \geq 4$, PAES-25 achieves a strictly better asymptotic runtime than the best known upper bound of $O(n^{m+1})$ established for SMO on m -LOTZ. We also show that if the size of the archive is smaller than the size of the Pareto front, the underlying archiver spreads the solutions in the archive well across the Pareto front in many scenarios where many of them are restricted to LOTZ, the bi-objective variant of m -LOTZ.

Secondly, we investigate PAES-25 with standard bit mutation and show that if the size of the archive is at least $n + 1$, the archive contains every Pareto-optimal search point of LOTZ after $O(n^4)$ iterations in expectation. This is another novel approach, since, as far as we know, the only rigorously proven runtime bounds of PAES or variants relies on one-bit mutation.

Thirdly, we want to demonstrate also some limitations of PAES-25 and rigorously prove that, even if the archive size is $\Omega(n)$, the algorithm can only find $o(n)$ different Pareto-optimal solutions of OMM and COCZ after $2^{\Omega(n)}$ iterations. To achieve this, we will use the negative drift theorem from [34].

Related work: The mathematical runtime analysis of MOEAs began around 20 years ago. Initial work focused on simple algorithms like the simple evolutionary multi-objective optimizer (SEMO) [32]

or the global SEMO (GSEMO) [20]. Since then, the field has steadily advanced. The foundational runtime results for (G)SEMO were presented in [31], the journal version of [32], analyzing pseudo-Boolean functions like m -LOTZ and m -COCZ. There it was shown that (G)SEMO can find the entire Pareto front of m -COCZ (the m -objective analogues of the classic COCZ benchmark) and m -LOTZ in an expected number of $O(n^{m+1})$ fitness evaluations, for an even number of objectives. The upper bound for LOTZ naturally extends to more than two objectives while the bi-objective COCZ guarantees a runtime of $O(n^2 \log(n))$. Subsequent work by Bian et al. [2] improved the results for m -COCZ to $O(n^m)$ for $m > 4$ and to $O(n^3 \log n)$ in the special case $m = 4$. These results were further refined by [47], who showed that for m -COCZ and m -LOTZ the runtime depends only linearly in the size of the largest set of incomparable solutions, aside from small polynomial factors in the number m of objectives and n . That work also analyzed (G)SEMO on the m -OMM and m -OJZJ_k benchmarks (the m -objective variants of the bi-objective OMM and OJZJ_k). Most recently, Doerr et al. [16] presented the first runtime analysis of (G)SEMO on integer-valued search spaces and there are even lower runtime bounds of GSEMO on COCZ, OMM and OJZJ_k [15].

These insights motivated the rigorous analysis of more sophisticated MOEAs that are commonly used in practice. The most prominent among them is NSGA-II [12] which has received over 50000 citations. The first rigorous runtime analysis of this algorithm was conducted by Zheng et al. [51] on classical benchmark problems. This was followed by a series of results investigating its performance on a multimodal problem [17], approximation qualities in covering the Pareto front [48], robustness to noisy environments [8], lower bounds [18], the usefulness of crossover [9, 19], trap functions [10] and stochastic population update [4]. NSGA-II has also been studied in the context of combinatorial optimization problems, including the minimum spanning tree [6] and the subset selection problem [13]. Despite these successes, NSGA-II's performance is very limited to the bi-objective setting as shown in [49]. This limitations has recently led to the runtime analysis of other prominent MOEAs designed for many-objectives, such as NSGA-III [39–41, 46], SMS-EMOA [50], SPEA-2 [44] and a modified version of NSGA-II with a simple tie-breaking rule [14]. All of these algorithms are population based and, in each iteration, select parents for mutation uniformly at random in contrast to PAES and PAES-25. An additional property of PAES-25 is that it uses an archive to store non-dominated solutions. The study of the effects of such archives in multiobjective optimization was initiated only recently by Bian et al. [3]. They introduced an unbounded archive to drastically reduce the required population size of MOEAs for efficiently optimizing classical benchmark problems, resulting in significant speedups. Their analysis leverages crossover at boundary points to effectively cover the entire Pareto front.

However, for the archive-based algorithm PAES, to the best of our knowledge only one rigorous runtime analysis is known [42]. In that work, the pseudo-Boolean function PATH is introduced, where PAES with one-bit mutation has an expected runtime bound of $\Theta(n^2)$ which outperforms SEMO by a factor of n . However, it is also shown that PAES fails to find the complete Pareto front of the LOTZ benchmark with overwhelming probability. We are not aware of any other rigorous runtime results for the standard PAES

or its variant PAES-25. In this paper we show that PAES-25 is, in fact, able to find the Pareto front of LOTZ very efficiently.

2 PRELIMINARIES

Notation: For a finite set A we denote by $|A|$ its cardinality. Let \log be the logarithm to base e and $[k] := \{1, \dots, k\}$ for $k \in \mathbb{N}$. The number of ones in a bit string x is denoted by $|x|_1$. The number of leading ones in x , denoted by $\text{LO}(x)$, is the length of the longest prefix of x which contains only ones, and the number of trailing zeros in x , denoted by $\text{TZ}(x)$, is the length of the longest suffix of x containing only zeros respectively. For example, if $x = 1110101100000$, then $\text{LO}(x) = 3$ and $\text{TZ}(x) = 5$. This paper is about many-objective optimisation, particularly the maximisation of a discrete m -objective function $f(x) := (f_1(x), \dots, f_m(x))$ where $f_i : \{0, 1\}^n \rightarrow \mathbb{N}_0$ for each $i \in [m]$. When $m = 2$, the function is also called *bi-objective*. Denote by f_{\max} the maximum possible value of f in one objective, i.e. $f_{\max} := \max\{f_i(x) \mid x \in \{0, 1\}^n, i \in [m]\}$. Given two search points $x, y \in \{0, 1\}^n$, x *weakly dominates* y , denoted by $x \succeq y$, if $f_i(x) \geq f_i(y)$ for all $i \in [m]$, and x *(strictly) dominates* y , denoted by $x \succ y$, if one inequality is strict; if neither $x \succeq y$ nor $y \succeq x$ then x and y are *incomparable*. A set $S \subseteq \{0, 1\}^n$ is a *set of mutually incomparable solutions* with respect to f if all search points in S are incomparable. Each solution not dominated by any other in $\{0, 1\}^n$ is called *Pareto-optimal*. A mutually incomparable set of these solutions that covers all possible non-dominated fitness values is called a *Pareto-optimal set* of f .

We also need the following Theorem from [29] to describe the hitting time of unbiased random walks on the line, with one barrier. We will use this result for describing the dynamics of PAES-25 with standard bit mutation on LOTZ below.

THEOREM 2.1 (THEOREM 4.2 IN [29]). *Let $n \in \mathbb{N}$ and let $(X_t)_{t \in \mathbb{N}}$ be an integrable random process over $\{0, \dots, n\}$, and let $T := \inf\{t \in \mathbb{N} \mid X_t = n\}$. Suppose that there is a $\delta \in \mathbb{R}^+$ such that, for all $t < T$, we have the following conditions (variance, drift).*

- (1) $\text{Var}(X_{t+1} - X_t \mid X_0, \dots, X_t) \geq \delta$.
- (2) $\mathbb{E}(X_{t+1} - X_t \mid X_0, \dots, X_t) \geq 0$.

Then $\mathbb{E}(T) \leq n^2/\delta$.

Algorithm: The PAES-25 [28], a modification of PAES [25, 26] is displayed in Algorithm 1. Initially, a search point s from $\{0, 1\}^n$ is sampled uniformly at random, and the archive A_0 is set to $\{s\}$. Then, as long as the termination criterion is not met, a copy s' of s is created and mutation is applied on s' (i.e. each bit is flipped uniformly at random with probability $1/n$ in case of standard bit mutation and one bit is flipped chosen uniformly at random in case of one-bit mutation) to generate a new solution c . The solution s is referred to as the *current solution*, and c as the *candidate solution*. Then, if the candidate solution c weakly dominates a point in the archive A_t , all points weakly dominated by c are removed from A_t and c is added to A_t . This stands in strict contrast to the standard PAES which adds c only to the archive and accepts it as a new current solution if it is neither weakly dominated by s nor by any point in the archive. The current solution for the next iteration then becomes c . Conversely, if c is dominated by a member of the archive, it is discarded. If c is not weakly dominated by any search point x in the archive A_t (that is, it is incomparable to all search points in A_t),

Algorithm 1: PAES-25 with archive size L and archiver X for maximizing $f : \{0, 1\}^n \rightarrow \mathbb{N}_0^m$

```

1 Initialize  $s \sim \text{Unif}(\{0, 1\}^n)$ 
2 Set  $t = 0$  and create set  $A_0 = \{s\}$  as an archive
3 while termination criterion not met do
4   Create  $s'$  as a copy of  $s$ 
5   Create  $c$  by mutation on  $s'$ 
6   if  $c$  weakly dominates any element in  $A_t$  then
7     Remove all  $x \in A_{t+1}$  weakly dominated by  $c$ 
8      $A_{t+1} = A_t \cup \{c\}$ ,  $s = c$ 
9   else if there is  $x \in A_t$  dominating  $c$  then
10     $A_{t+1} = A_t$ 
11  else
12    if  $|A_t| < L$  then
13       $A_{t+1} = A_t \cup \{c\}$ ,  $s = c$ 
14    else
15      if  $c$  is accepted by the archiver  $X$  then
16        Choose  $y \in A_t$  according to  $X$ 
17         $A_{t+1} = (A_t \setminus \{y\}) \cup \{c\}$ ,  $s = c$ 
18   $t = t + 1$ 

```

it is added to A_t if the archive size is smaller than L . Otherwise, if the archive is full, an *archiver* is responsible for deciding whether c will be accepted into A_t and becomes the new current solution. The archiver's primary role is to maintain *diversity* in the archive [28]. If the archiver rejects c , the archive A_t remains unchanged, and c is discarded. Otherwise, c is added to the archive, becomes the new current solution s for the next iteration, and another point $y \in A_t$, selected by the archiver, is removed from the archive. Note that the archive always contains pairwise incomparable search points and if $A_t = P$ for a Pareto-optimal set in one iteration, then $A_t = P$ for all future iterations. In this paper we examine three archivers which are also discussed in Knowles and Liefooghe [28], but with minor modifications to enable progress on fitness plateaus even when the archive is full. Other archivers, such as *crowding distance* archiver (NSGA-II archiver), the SPEA2 archiver or the ε -dominance archiver, exhibit behavior that closely resembles that of one of the three considered in this paper (see for example [36]).

Adaptive Grid Archiver (AGA) [26]: It covers the search space with a *grid* and associates solutions from the archive to *cells* as follows. Let the values $a \geq f_{\max}$ and $\ell \in \mathbb{N}$ be given. By bisecting the interval $[0, a]$ of each objective axis ℓ times into disjoint intervals $[0, a/2^\ell], \dots, [(2^\ell - 2)a/2^\ell, (2^\ell - 1)a/2^\ell], [(2^\ell - 1)a/2^\ell, a]$, the hypercube $[0, a]^m$ is divided into $2^{\ell \cdot m}$ cells. The location of an objective vector $v = (v_1, \dots, v_m) \in [0, a]^m$ in the objective space is then described by a binary string of length $2^{\ell \cdot m}$, where the decimal value $0 \leq k < 2^\ell - 1$ of the j -th block of 2^ℓ elements from the left indicates the interval $I = [ka/2^\ell, (k+1)a/2^\ell] \subset [0, a]$ such that $v_j \in I$. If $k = 2^\ell - 1$ then $I = [ka/2^\ell, (k+1)a/2^\ell]$. Two solutions x, y belong to the same cell if their corresponding objective values $f(x)$ and $f(y)$ are represented by the same binary string of length $2^{\ell \cdot m}$. The entire grid is formed by these cells. This archiver always accepts the candidate solution c in Line 15 in Algorithm 1 (i.e. if c

is incomparable to all points $x \in A_t$ and A_t is full). Then, a solution distinct from c from a grid cell containing the most search points is selected uniformly at random and removed from A_t (in Line 15 in Algorithm 1). Ties with respect to such cells are broken uniformly at random.

Hypervolume Archiver (HVA) [24, 27]: For a set $P \subset \{0, 1\}^n$, the *hypervolume* with respect to a fitness function $f: \{0, 1\}^n \rightarrow \mathbb{R}^m$ and a reference point $h \in \mathbb{R}^m$ (preset by the user at the beginning) is defined as

$$\text{HV}(P) := \mathcal{L} \left(\bigcup_{x \in P} \{v \in \mathbb{R}^m \mid \text{for all } i \in [m] : h_i \leq v_i \wedge v_i \leq f_i(x)\} \right)$$

where \mathcal{L} denotes the Lebesgue measure. The *hypervolume contribution* of a point $x \in A_t \cup \{c\}$ is the value $\text{HV}(A_t \cup \{c\}) - \text{HV}((A_t \cup \{c\}) \setminus \{x\})$ (see for example [1, 52] for theoretical investigations of the hypervolume or [5] for an efficient algorithm to compute minimal hypervolume contributions for subsets of $\{0, 1\}^n$). If c contributes smallest to $\text{HV}(A_t \cup \{c\})$ as the only point from $A_t \cup \{c\}$, it is not accepted. Otherwise, it is added to A_t , and a different point with the smallest hypervolume contribution to $\text{HV}(A_t \cup \{c\})$ is removed from A_t where ties are broken randomly.

Multi-Level Grid Archiver (MGA) [33, 37]: The multi-level grid archiver (MGA) can be considered as a hybrid approach, combining elements from both the adaptive grid archiver (AGA) and the ε -Pareto archiver (see, for example, [36] for the latter). The MGA uses a hierarchical family of *boxes* with respect to a *coarseness level* $b \geq 0$ over the objective space. For an objective vector $v := (v_1, \dots, v_m)$ we define the *box index vector* at coarseness level b as $v^{\text{box}} := (v_1^{\text{box}}, \dots, v_n^{\text{box}})$ where $v_i^{\text{box}} := \lfloor v_i \cdot 2^{-b} \rfloor$. Then, when a candidate solution c is created, the smallest level b is determined such that two box index vectors of solutions from $A_t \cup \{c\}$ are comparable. The candidate solution c is rejected if its corresponding box index vector is the only one that is weakly dominated by another element from A_t at this level b . Otherwise, c is added to A_t , and an arbitrary solution x from A_t distinct from c that is weakly dominated by another element from $(A_t \setminus \{x\}) \cup \{c\}$ at level b is removed.

In all three schemes, in case of a tie between the candidate solution c and another solution $z \in A_t$ in the archiver (e.g., when they have the same smallest hypervolume contribution in case of HVA), the candidate solution c is always accepted becoming the new current solution. This approach reflects a common strategy and is useful for exploring plateaus. It is a minor difference to some versions displayed in [27], where a solution x is not necessarily added to A_t if there is already a point in A_t with the same fitness as x . In such a case the candidate solution is rejected.

A major difference to the classical MOEAs studied before where a parent is chosen according to some probability distribution from a *population* is that only the current solution is selected for mutation and this solution is either the current solution of the previous iteration if the candidate solution was rejected, or the solution which is added to the archive becoming the new current solution if the candidate solution was accepted in the previous iteration.

3 RUNTIME ANALYSIS OF PAES-25 WITH ONE-BIT MUTATION ON m -LOTZ

We recall the m -LEADINGONES TRAILING ZEROES function (m -LOTZ for short) for many objectives. In m -LOTZ(x), we divide x in $m/2$ many blocks x^j of equal length such that $x = (x^1, \dots, x^{m/2})$ and in each block we count the LO-value and the TZ-value. This can be formalized as follows.

Definition 3.1 (Laumanns et al. [30]). Let m be divisible by 2 and let the problem size be a multiple of $m/2$. Then the m -objective function m -LOTZ is defined by m -LOTZ: $\{0, 1\}^n \rightarrow \mathbb{N}_0^m$ as

$$m\text{-LOTZ}(x) = (f_1(x), f_2(x), \dots, f_m(x))$$

with

$$f_k(x) = \begin{cases} \sum_{i=1}^{2n/m} \prod_{j=1}^i x_{j+n(k-1)/m}, & \text{if } k \text{ is odd,} \\ \sum_{i=1}^{2n/m} \prod_{j=i}^{2n/m} (1 - x_{j+n(k-2)/m}), & \text{else,} \end{cases}$$

for all $x = (x_1, \dots, x_n) \in \{0, 1\}^n$.

A Pareto-optimal set of m -LOTZ is

$$\{1^{i_1} 0^{2n/m-i_1} \dots 1^{i_{m/2}} 0^{2n/m-i_{m/2}} \mid i_1, \dots, i_{m/2} \in \{0, \dots, 2n/m\}\}$$

which we denote by P which coincides with its set of Pareto-optimal search points. The cardinality of this set is $(2n/m + 1)^{m/2}$. In [41] it is shown that the size of a set of mutually incomparable solutions of m -LOTZ is at most $(2n/m + 1)^{m-1}$. In the next lemma it is shown that the size of this set can be indeed asymptotically larger than the size of the Pareto front for a wide range of values m . For reasons of space, its proof is presented in the appendix.

LEMMA 3.2. Let S be a maximum cardinality set of mutually incomparable solutions for $f := m$ -LOTZ. Then $|S| = n + 1$ if $m = 2$ and

$$\frac{(2n/m + 1)^{m-1}}{4(m-2)^{m/2-1}} \leq |S| \leq (2n/m + 1)^{m-1}$$

if $m \geq 4$.

In this section we analyze PAES-25 with one-bit mutation on m -LOTZ and establish tight bounds on the expected runtime to fill the archive with the full Pareto set P . A key part of the analysis involves proving that the potential function $W_t := \sum_{i=1}^{m/2} (\text{LO}(s^i) + \text{TZ}(s^i))$, where s denotes the current solution in iteration t , is non-decreasing. Observe that $0 \leq W_t \leq n$, and s is Pareto-optimal if and only if $W_t = n$. Consequently, once a Pareto-optimal solution is found, it remains Pareto-optimal in all subsequent iterations. This monotonicity of the potential function W_t also enables the derivation of an expected runtime bound for reaching the Pareto front.

LEMMA 3.3. Consider PAES-25 with one-bit mutation and any archiver optimizing m -LOTZ for any $m \leq n$ divisible by 2 and archive size at least one. Let $W_t := \sum_{i=1}^{m/2} (\text{LO}(s^i) + \text{TZ}(s^i))$. Then $W_{t+1} \geq W_t$. Furthermore, the expected number of iterations until the current solution is Pareto-optimal is $O(n^2)$.

PROOF. The current solution s either remains unchanged if the candidate solution c is rejected or it is updated if the candidate solution is accepted. Suppose we flip bit $i \in [2n/m]$ in block ℓ . If $i \in \{1, \dots, \text{LO}(s^\ell)\}$, then $\text{LO}(c^\ell) < \text{LO}(s^\ell)$ and all LO- and TZ-values

of all other blocks remain unchanged. Similarly, if $i \in \{2n/m - \text{TZ}(s^\ell) + 1, \dots, 2n/m\}$, then $\text{TZ}(c^\ell) < \text{TZ}(s^\ell)$, and again, all LO- and TZ-values of all other blocks remain unchanged. In both cases, c is dominated by s and c is therefore discarded which implies that s remains unchanged yielding $W_{t+1} = W_t$. Further, suppose that $i \in \{\text{LO}(s^\ell) + 2, \dots, 2n/m - \text{TZ}(s^\ell) - 1\}$. Then all LO- and TZ-values remain unchanged and hence $W_{t+1} = W_t$ no matter if s is updated. If $i = \text{LO}(s^\ell) + 1$ a zero is flipped to one and if $i = 2n/m - \text{TZ}(s^\ell)$ a one is flipped to zero. Hence, in the first case, $\text{LO}(c^\ell) > \text{LO}(s^\ell)$ while $\text{TZ}(s^\ell)$ decreases only by at most one. In the second case $\text{TZ}(s^\ell) > \text{TZ}(c^\ell)$ while $\text{LO}(s^\ell)$ decreases only by at most one. Hence, the current solution is updated yielding $W_{t+1} \geq W_t$, particularly $W_{t+1} > W_t$ if s is not Pareto-optimal. All these cases together show $W_{t+1} \geq W_t$. Flipping one specific bit requires $O(n)$ iteration in expectation. Repeating this at most n times (as long as $W_t < n$), derives the desired runtime bound concluding the proof. \square

We now analyze how the current solution explores the Pareto front. When the current solution generates a new Pareto-optimal candidate c , it is accepted as the new current solution and added to the archive A_t if L is sufficiently high. If there already exists a solution $y \in A$ with $f(y) = f(c)$, then y is replaced by c in the archive. In both scenarios, the candidate c becomes the current solution for the next iteration. This mechanism enables the current solution to perform a random walk on the Pareto front P . We will demonstrate that this walk can be modeled as a specific type of random walk on an m -dimensional grid graph.

To this end, we need some more notation. For an undirected connected graph G , a *simple random walk* on $G = (V, E)$ is a random process in which, at each step, an edge is chosen uniformly at random from the set of edges incident to the current node, and the walk transitions to the next node along that edge. We denote by $T_{G,v}$ the time it takes for the simple random walk on G to visit all nodes, starting at v . The *cover time* $E(T_{G,v})$ with respect to $v \in V$ is the expected time it takes for the simple random walk on G to visit all nodes, starting at v .

For two undirected graphs $G := (V_1, E_1)$ and $H := (V_2, E_2)$, the *product graph* $G \times H$ is the graph with node set $V_1 \times V_2$ and edge set E where $\{(u_1, u_2), (v_1, v_2)\} \in E$ if and only if $u_1 = v_1$ and $(u_2, v_2) \in E_2$ or $u_2 = v_2$ and $(u_1, v_1) \in E_1$. We write $G^2 := G \times G$ and inductively $G^k := G \times G^{k-1}$ for $k \geq 3$. For example, if G is a *path graph*, that is its vertices can be labeled as v_1, \dots, v_n with edges $\{v_i, v_{i+1}\}$ for $i \in [n-1]$, then the product graph G^k is the *grid graph* (V^k, E^k) of dimension k and length n . Hereby, two $u, w \in V^k$ are adjacent if and only if there are $j, \ell \in [k]$ such that $w_j = v_\ell$ and $u_j \in \{v_{\ell-1}, v_{\ell+1}\}$, and $u_i = w_i$ for all $i \in \{1, \dots, n\} \setminus \{j\}$. We need the following result from Jonasson [23] about the cover time of the product graph G^k . Surprisingly, this depends only asymptotically on the *average degree* $d_G := 2|E|/n$ of the underlying graph G and the number of nodes in G^k , and not even on the choice of the starting node $v \in G^k$.

THEOREM 3.4 ([23], THEOREM 1.2). *Let $G = (V, E)$ be a connected graph with n vertices and average degree d_G . Then the cover time $E(T_{G,v})$ for a simple random walk on G^k is $\Theta(d_G n^2 \log^2(n))$ if $k = 2$ and $\Theta(d_G n^k \log(n^k))$ otherwise for every $v \in G^k$.*

For the remainder of this section, we mention the cover time without referring to the starting node, by convention. If G is the path graph as mentioned above, then $d_G = 2(n-1)/n \in \Theta(1)$ and hence, the cover time on G^k is $\Theta(n^2 \log^2(n))$ if $k = 2$ and $\Theta(n^k \log(n^k))$ for $k > 2$. Now we are ready to prove the desired runtime bound of PAES-25 with 1-bit mutation on m -LOTZ.

THEOREM 3.5. *Consider PAES-25 with one-bit mutation, an arbitrary archiver and archive size $L \geq (2n/m + 1)^{m-1}$, or with the AGA and archive size $L \geq (2n/m + 1)^{m/2}$, optimizing m -LOTZ for any $m \leq n$ where m is divisible by 2. Then the expected number of iterations until $A_t = P$ is $\Theta(n^3)$ if $m = 2$, $\Theta(n^3 \log^2(n))$ if $m = 4$ and $\Theta(n \cdot (2n/m)^{m/2} (\log(n/m)))$ if $m > 4$.*

PROOF. By Lemma 3.3 the first Pareto-optimal search point is created in expected $O(n^2)$ iterations. From this time on, the current solution always maintains Pareto-optimality by Lemma 3.3. At first we assume that the archiver is arbitrary and hence, $L \geq (2n/m + 1)^{m-1}$. The candidate solution c is always added to A_t becoming the new current solution s while possibly removing another point from A_t : If A_t is not full or there is already an $x \in A_t$ with the same fitness as c , then it is obvious. Otherwise, c dominates another $x \in A$ (since $L \geq |S|$ by Lemma 3.2 where S is a maximum set of incomparable solutions) and x is removed from A_t . Hence, we have to compute the expected time until s attained all Pareto-optimal search points. To this end, we define the following graph $G = (V, E)$ and model the trajectory of s by a random walk on G . Let $V := \{(v_1, \dots, v_{m/2}) \mid v_i \in \{0, \dots, 2n/m\} \text{ for every } i \in [m/2]\}$ and $E := \{(v, w) \mid v, w \in V, |v - w| = 1\}$ where $| \cdot |$ denotes the Euclidean distance in the objective space. Then $G = \tilde{G}^{m/2}$ where \tilde{G} is the path graph with $2n/m + 1$ nodes. Then the current solution $s = (s^1, \dots, s^{m/2})$ visits $v := (v_1, \dots, v_{m/2})$ if it is Pareto-optimal and for each $i \in [m/2]$ we have that $\text{LO}(s^i) = v_i$. The transition probability between two adjacent nodes in the graph $\tilde{G}^{m/2}$ is $1/n$, as for a specific block i the $\text{LO}(s^i)$ increases or decreases by one with probability $1/n$, while the $\text{LO}(s^j)$ -values for $j \neq i$ remain unchanged. The probability that the node v is not left is $1 - |\text{adj}(v)|/n$ where $\text{adj}(v)$ is the number of adjacent nodes to v . Let t_0 be the first iteration when s is Pareto-optimal. In iteration t_0 , the random walk begins at $v \in V$, where $\text{LO}(s^i) = v_i$ for each $i \in [m/2]$. A mutation of s into c corresponds to a transition from the node $v = (\text{LO}(s^1), \dots, \text{LO}(s^{m/2}))$ to $w = (\text{LO}(c^1), \dots, \text{LO}(c^{m/2}))$ if c is also Pareto-optimal. The time (from iteration t_0) to obtain $A = P$ corresponds to the time T of the random walk on $G = \tilde{G}^{m/2}$ described above to visit all nodes, starting at v . Note that the random walk is not simple, as for every node v there is a certain probability to stay at v . We asymptotically determine $E(T)$. Let $M \leq T$ be the number of moves the walk makes on $\tilde{G}^{m/2}$ to visit all nodes (i.e. idle steps where nodes are not left are ignored in M). Then M is also the number of steps a *simple random walk* has to make on the grid graph $\tilde{G}^{m/2}$ to visit all nodes. Denote by X the waiting time to leave a node with the most neighbors and by Y the waiting time to leave a node with the least ones. Then $E(X) = n/(2m)$ and $E(Y) = n/m$ as $\text{adj}(v) \in \{m, \dots, 2m\}$. Note that M can be regarded as independent to X and Y and therefore $E(X) \cdot E(M) \leq E(T) \leq E(Y) \cdot E(M)$ which implies (since the number of nodes in $\tilde{G}^{m/2}$ is $(2n/m + 1)^{m/2}$)

$$E(T) = \Theta(n/m \cdot (2n/m + 1)^{m/2} \log((2n/m + 1)^{m/2}))$$

$$= \Theta(n \cdot (2n/m)^{m/2} (\log(n/m)))$$

if $m \geq 6$ and

$$\begin{aligned} E(T) &= \Theta(n/m \cdot (2n/m + 1)^{m/2} \log^2(2n/m + 1)) \\ &= \Theta(n^3 \log^2(n)) \end{aligned}$$

if $m = 4$. If $m = 2$ we have that $\tilde{G}^{m/2}$ is a path graph and therefore $E(X) = E(Y) = \Theta(1/n)$ and $E(T) = \Theta(n^3)$ since $E(M) = \Theta(n^2)$ (by Theorem 1 in [22]). This concludes the proof for arbitrary archivers with size L of at least $(2n/m + 1)^{m-1}$. In the case of AGA, any solution that is incomparable to all solutions in A_t is always accepted as the new candidate solution. Consequently, once the Pareto front has been reached, the same random walk performed by the current solution as above eliminates any non Pareto-optimal search points from the archive by visiting all Pareto-optimal solutions. In a second phase of visiting all nodes again, if necessary, it fills the archive with Pareto-optimal solutions. This shows that the stated expected tight runtime bound also holds in this case since every Pareto-optimal search point needs to be visited by the current solution at most twice, but at least once. \square

For other MOEAS like (G)SEMO only an upper bound of $O(n^{m+1})$ is shown on m -LOTZ (see [30, 47]). Theorem 3.5 shows that we beat this upper runtime bound by a factor of $n^2/\log^2(n)$ if $m = 4$ and $(nm/2)^{m/2}/(\log(n/m))$ if $m > 4$. To the best of our knowledge, these are the first tight runtime bounds of an MOEA in the many-objective setting.

For the remainder of this section we show that PAES-25 with one-bit mutation can efficiently create a well-distributed archive on the Pareto front of m -LOTZ when using AGA, HVA and MGA, assuming that the archive size L is smaller than the size of the Pareto front. In the case of HVA and MGA, we restrict our analysis to the bi-objective setting.

AGA: PAES-25 with AGA efficiently distributes all solutions in A_t very evenly across the cells that contain Pareto-optimal search points, stated as follows.

LEMMA 3.6. *Consider PAES-25 with one-bit mutation, AGA and archive size $L < (2n/m + 1)^{m/2}$ optimizing m -LOTZ for any $m \leq n$ where m is divisible by 2. Then after expected $O(n^3)$ iterations if $m = 2$, $O(n^3 \log^2(n))$ iterations if $m = 4$ and $O(n \cdot (2n/m)^{m/2} (\log(n/m)))$ iterations if $m > 4$ we obtain the following distribution of solutions from A_t across all cells that contain Pareto-optimal search points. For any cell containing k Pareto-optimal solutions from A_t , the following holds:*

- (1) *Every other cell containing at least k distinct Pareto-optimal solutions from $\{0, 1\}^n$ contains at least $k - 1$ Pareto-optimal solutions from A_t .*
- (2) *Every cell containing $\ell \leq k - 1$ Pareto-optimal solutions from $\{0, 1\}^n$ contains ℓ Pareto-optimal solutions from A_t .*

PROOF. With Lemma 3.3 we can assume that $A_t \subset P$. The current solution performs the same random walk as in the proof of Theorem 3.5. Hence after $\Theta(n^3)$ iterations if $m = 2$, $\Theta(n^3 \log^2(n))$ iterations if $m = 4$ and $\Theta(n \cdot (2n/m)^{m/2} (\log(n/m)))$ iterations if $m > 4$ in expectation, the current solution has visited all Pareto-optimal fitness vectors. From this time on, A_t consists only of Pareto-optimal solutions since dominated solutions are removed from A_t

(see Algorithm 1). Finally, to fill and spread A_t on the Pareto front, we repeat the process a second time to visit all Pareto-optimal fitness vectors again. Suppose that after the final iteration in this second phase there exists a cell C containing k Pareto-optimal solutions from A_t , and another cell C' that contains at most $\ell \leq k - 2$ Pareto-optimal solutions from A_t , but at least $\ell + 1$ Pareto-optimal solutions from P . Then, during an earlier iteration, at least one solution from $P \cap C'$ must have been removed from A_t when cell C' had exactly $\ell + 1$ Pareto-optimal solutions from A_t (since all Pareto-optimal solutions have been visited after the second phase). At this time t , A_t was full and every other cell (including C) contained at most $\ell + 1 < k$ solutions from A_t . But AGA always removes Pareto-optimal solutions from A_t contained in cells with the largest number of Pareto-optimal solutions from A_t . Hence, for any iteration $t' \geq t$, once a cell is filled with an $(\ell + 2)$ -th Pareto optimal search point from A_t , one of these $(\ell + 2)$ points is removed. Consequently, each cell contains at most $\ell + 1 < k$ solutions from A_t , after iteration t' which is a contradiction proving properties (1) and (2). \square

HVA: We show that PAES-25 with HVA and reference point $h = (-1, -1)$ efficiently distributes the archive A_t across the Pareto front, resulting in a high hypervolume. We restrict our analysis to the case $m = 2$, as for $m > 2$, there may be non-dominated points in A_t which are not Pareto-optimal even if a Pareto-optimal point is already found. This complicates the analysis since non Pareto-optimal points in A_t may have a greater hypervolume contribution than newly generated Pareto-optimal points from the random walk, potentially causing the algorithm to reject new Pareto-optimal solutions. At first we establish some basic properties and compute the hypervolume of A_t under the assumption that all points in A_t are Pareto-optimal and there are no *holes*, meaning that for each $x, y \in A_t$ and $z \in P$ with $\text{LO}(x) < \text{LO}(z) < \text{LO}(y)$, we also have $z \in A_t$. We also derive a formula for the case $A_t = P$, which is the maximum possible hypervolume for an archive A_t .

LEMMA 3.7. *Let $A_t \subset P$ be without a hole. Let $a := \min\{\text{LO}(x) \mid x \in A_t\}$ and $b := \max\{\text{LO}(x) \mid x \in A_t\}$. Then we have*

$$\text{HV}(A_t) = (n+1)(b+1) - \frac{a(a+1)}{2} - \frac{b(b+1)}{2}$$

with respect to the reference point $h = (-1, -1)$. If $d := b - a$ is fixed then the minimum hypervolume is attained when $a = 0$ or $b = n$ which is $\text{HV}(A_t) = (d+1)(n+1 - d/2)$.

PROOF. Note that $f(x) = (a, n - a)$ for a Pareto-optimum x with $\text{LO}(x) = a$ and hence, x dominates all y with $-1 \leq f_1(y) \leq a$ and $-1 \leq f_2(y) \leq n - a$ which contribute $(a+1)(n-a+1)$ to $\text{HV}(A_t)$. For $a < i \leq b$ all search points z with $i-1 < f_1(z) \leq i$ and $-1 \leq f_2(z) \leq n-i$ are dominated by a search point $x \in A_t$ with $\text{LO}(x) = i > a$ (since $\text{TZ}(x) = n-i$), but not by a search point $y \in A_t$ with $\text{LO}(y) < i$. This contributes $n-i+1$ to $\text{HV}(A_t)$. Therefore, we obtain

$$\begin{aligned} \text{HV}(A_t) &= (a+1)(n-a+1) + \sum_{i=a+1}^b (n-i+1) \\ &= (a+1)(n-a+1) + \sum_{i=0}^b (n-i+1) - \sum_{i=0}^a (n-i+1) \end{aligned}$$

$$\begin{aligned}
&= (a+1)(n-a+1) + (n+1)(b-a) + \sum_{i=1}^a i - \sum_{i=1}^b i \\
&= (n+1)b - a(a+1) + n+1 + \frac{a(a+1)}{2} - \frac{b(b+1)}{2}
\end{aligned}$$

proving the first formula. Now suppose that $d = b - a$ is fixed. If a increases by one then b also increases by one. Therefore, $HV(A_t)$ increases by

$$\begin{aligned}
&(n+1)(b+2) - \frac{(a+1)(a+2)}{2} - \frac{(b+1)(b+2)}{2} \\
&- (n+1)(b+1) + \frac{a(a+1)}{2} + \frac{b(b+1)}{2} \\
&= n+1 - (a+1) - (b+1) = n - a - b - 1.
\end{aligned}$$

Hence, if $b \geq n - a$ the hypervolume is smallest if $b = n$ and $a = n - d$ which is

$$\begin{aligned}
&(n+1)^2 - \frac{(n-d)(n-d+1)}{2} - \frac{n(n+1)}{2} \\
&= (n+1)^2 - \frac{(n-d)(n+1)}{2} + \frac{d(n+1-d-1)}{2} - \frac{n(n+1)}{2} \\
&= (n+1) \left(n+1 - \frac{n-d}{2} + \frac{d}{2} - \frac{n}{2} \right) - \frac{d(d+1)}{2} \\
&= (n+1)(d+1) - \frac{d(d+1)}{2} = (d+1) \left(n+1 - \frac{d}{2} \right).
\end{aligned}$$

Similarly, if a decreases by one then b decreases also by one and the hypervolume increases by $a + b - n - 1$. If $b < n - a$ then the hypervolume is smallest if $a = 0$ and $b = d$ which is

$$(n+1)(d+1) - \frac{d(d+1)}{2} = (d+1) \left(n+1 - \frac{d}{2} \right)$$

proving the result. \square

We now show that PAES-25 with HVA and one-bit mutation on LOTZ effectively distributes the archive across the Pareto front. To this end, we set $a_t := \min\{\text{LO}(x) \mid x \in A_t\}$, $b_t := \max\{\text{LO}(x) \mid x \in A_t\}$ and $d_t := b_t - a_t$. Call a Pareto-optimal z hole if $a_t < \text{LO}(z) < b_t$ and $z \notin A_t$.

LEMMA 3.8. Consider PAES-25 with one-bit mutation, HVA with $h = (-1, -1)$, and archive size $L < n+1$ optimizing LOTZ. Then after expected $O(n^3)$ iterations the following holds.

- (1) For any given archive size L with $L + \lceil L/2 \rceil \leq n+2$, all solutions in A_t are distributed such that $d_t = L + \lceil L/2 \rceil - 2$. Furthermore, there are $\lceil L/2 \rceil - 1$ holes and two of them are not neighbored.
- (2) For any given archive size L with $L + \lceil L/2 \rceil \leq n+2$, a hypervolume of at least

$$(L + \lceil L/2 \rceil - 1) \cdot \left(n+1 - \frac{L + \lceil L/2 \rceil - 2}{2} \right) - \lceil L/2 \rceil + 1$$

is attained. If $L = cn$ for some $0 < c \leq 1$ this corresponds to at least a fraction of $3c(1 - 3c/4)(1 - o(1))$ of the total hypervolume $HV(P)$. If instead $L + \lceil L/2 \rceil > n+2$, then a hypervolume of $(n+2)(n+1)/2 - \max\{n+1-L, 0\}$ is achieved which is also the maximum possible in this case.

PROOF. With Lemma 3.3 we can assume that $A_t \subset P$. For $L = 1$ the lemma holds since $HV(A_t) \geq n+1$. Suppose that $L \geq 2$. As long as solutions $x, y, z \in A_t$ with $\text{LO}(x) = \ell$, $\text{LO}(y) = \ell+1$, and

$\text{LO}(z) = \ell+2$ exist, the solution y contributes a hypervolume of one which is smallest possible. Therefore, if the candidate solution c is incomparable to all solutions in A_t , its contribution cannot be smaller than that of y , and it is added to A_t . As long as there are fewer than $\lfloor (d_t+1)/3 \rfloor$ holes, such x, y, z can be found, and the current solution conducts the same random walk as in the proof of Theorem 3.5 and a new created Pareto-optimal solution is always added to A_t (possibly replacing another solution with the same fitness). Hence, after $O(n^3)$ iterations in expectation, the current solution will have visited all Pareto-optimal points if it does not get stuck. The latter can only happen if the archive is full and there are at least $\lfloor (d_t+1)/3 \rfloor$ holes. The latter implies $L \leq (d_t+1) - \lfloor (d_t+1)/3 \rfloor$ yielding

$$\begin{aligned}
L + \lceil L/2 \rceil - 2 &\leq \frac{3L}{2} - \frac{3}{2} \leq \frac{3(d_t+1)}{2} - \frac{3\lfloor (d_t+1)/3 \rfloor}{2} - \frac{3}{2} \\
&\leq \frac{3(d_t+1)}{2} - \frac{3((d_t+1)/3 - 2/3)}{2} - \frac{3}{2} \\
&= \frac{3(d_t+1) - (d_t+1) + 2}{2} - \frac{3}{2} = d_t + \frac{1}{2}
\end{aligned}$$

and therefore $L + \lceil L/2 \rceil - 2 \leq d_t$.

Suppose that $L + \lceil L/2 \rceil \leq n+2$. Then in $O(n^3)$ iterations in expectation an iteration t^* is reached when A_{t^*} is full and $L + \lceil L/2 \rceil - 2 = d_{t^*} = b_{t^*} - a_{t^*}$ since $L + \lceil L/2 \rceil - 2 > d_t$ for all iterations t before t^* . The number of search points x in P with $a_{t^*} \leq \text{LO}(x) \leq b_{t^*}$ is $d_{t^*} + 1$ and hence, there are $d_{t^*} + 1 - L = \lceil L/2 \rceil - 1$ holes. Note that in all iterations t before t^* only solutions from A_t are removed which have a hypervolume contribution of at most one to $HV(A_t)$ (since there is $\ell \in \{0, \dots, n-2\}$ and x, y, z with $\text{LO}(x) = \ell$, $\text{LO}(y) = \ell+1$ and $\text{LO}(z) = \ell+2$) and hence, two holes in A_{t^*} are never neighbored. This proves property (1). By Lemma 3.7 we see at iteration $t := t^*$

$$HV(A_t) \geq (L + \lceil L/2 \rceil - 1) \left(n+1 - \frac{L + \lceil L/2 \rceil - 2}{2} \right) - \lceil L/2 \rceil + 1$$

since plugging a hole with an additional search point would increase $HV(A_t)$ by one. If $L = cn$ for some $0 < c < 1$ this is at least a fraction of

$$\begin{aligned}
&\frac{(cn + \lceil cn/2 \rceil - 1) \cdot \left(n+1 - \frac{cn + \lceil cn/2 \rceil - 2}{2} \right) - \lceil L/2 \rceil + 1}{(n+2)(n+1)/2} \\
&\geq \frac{(3cn/2 - 1) \cdot (n+3/2 - 3cn/4) - \lceil L/2 \rceil + 1}{(n+2)(n+1)/2} \\
&\geq \frac{3c/2 \cdot (1 - 3c/4)n^2(1 - o(1))}{(n+2)(n+1)/2} = 3c(1 - 3c/4)(1 - o(1)).
\end{aligned}$$

Suppose $L + \lceil L/2 \rceil > n+2$. This implies $L + L/2 + 0.5 \geq n+3$ and hence $L \geq 2(n+2.5)/3 = 2n/3 + 5/3$. The number of holes is at most $d+1-L \leq d+1 - (2n/3 + 5/3) \leq d+1 - (2d/3 + 5/3) = d/3 - 2/3 < \lfloor (d+1)/3 \rfloor$. Hence, c_t performs the same random walk as in the proof of Theorem 3.5 where only search points with hypervolume contribution of one to $HV(A_t)$ are removed from A_t . Hence, after $O(n^3)$ iterations in expectation we have $a = 0$, $b = n$ and there are $k := \max\{n+1-L, 0\}$ holes. Plugging such a hole with an additional search point increases $HV(A_t)$ by one, and hence, we see with Lemma 3.7

$$HV(A_t) = (n+1) \left(n+1 - \frac{n}{2} \right) - k = \frac{(n+1)(n+2)}{2} - k$$

proving the result. \square

For instance, Lemma 3.8 shows that when $h = (-1, -1)$, a fraction of approximately 15/16 of the total hypervolume of P can be attained by A_t , if its size is $n/2$, since $3/2 \cdot (1-3/8)(1-o(1)) \approx 15/16$.

MGA: We also show that PAES-25 with MGA efficiently creates an archive A_t , in which any two points $x, y \in A_t$ have different box index vectors with respect to a certain level $b > 0$, where the level b depends on the archive size. The smaller the archive size, the larger the corresponding level b . A higher level b leads to a finer distribution of solutions across the Pareto front. As for HVA, we restrict our analysis to the bi-objective case, since in higher dimensions, non Pareto-optimal vectors may dominate Pareto-optimal ones at level $b > 0$. This can result in a newly discovered Pareto-optimal solution not being added to A_t if the archive is full, thereby complicating the analysis. We start with some elementary properties.

LEMMA 3.9. *Let $n = 2^k \cdot \ell - 1$ where $k \in \mathbb{N}_0$ and ℓ is odd. Then the following holds:*

- (1) *Two distinct solutions $x, y \in P$ either have identical or mutually incomparable box index vectors at coarseness level $j \in \{0, \dots, k\}$.*
- (2) *There are exactly $(\ell + 1)/2$ incomparable box index vectors at coarseness level $k + 1$ and exactly $2^{k-j} \ell$ incomparable box index vectors at coarseness level $j \in \{0, \dots, k\}$.*

PROOF. (1): Let $x \in P$ be a Pareto-optimal solution. We find $a \in \{0, \dots, 2^{k-j} \cdot \ell - 1\}$ and $b \in \{0, \dots, 2^j - 1\}$ with

$$\begin{aligned} \text{LOTZ}(x) &= (2^j \cdot a + b, n - 2^j \cdot a - b) \\ &= (2^j \cdot a + b, 2^k \ell - 2^j \cdot a - (b + 1)). \end{aligned}$$

The corresponding box index vector at level j is $(a, 2^{k-j} \ell - a - 1)$. These are identical for the same a , but incomparable for different a .

(2): The components of a box index vector at coarseness level $k + 1$ have a range of $\{0, \dots, (\ell - 1)/2\}$ and the box index vectors of search points x with $(\text{LO}(x), \text{TZ}(x)) \in \{(0, n), (2^{k+1}, n - 2^{k+1}), \dots, (\ell - 1)/2 \cdot 2^{k+1}, n - (\ell - 1)/2 \cdot 2^{k+1})\}$ at coarseness level $k + 1$ are $\{(0, (\ell - 1)/2), (1, (\ell - 1)/2 - 1), \dots, ((\ell - 1)/2, 0)\}$ and therefore, pairwise incomparable. The components of a box index vector at coarseness level $j \in \{0, \dots, k\}$ have a range of $\{0, \dots, 2^{k-j} \ell - 1\}$ and the box index vectors of search points x with $(\text{LO}(x), \text{TZ}(x)) \in \{(0, n), (2^j, n - 2^j), \dots, (2^k \ell - 2^j, n - (2^k \ell - 2^j))\}$ at coarseness level j are $\{(0, 2^{k-j} \cdot \ell - 1), (1, 2^{k-j} \cdot \ell - 2), \dots, (2^{k-j} \cdot \ell - 1, 0)\}$ and therefore, pairwise incomparable. \square

LEMMA 3.10. *Consider PAES-25 with one-bit mutation and archive size $L \leq n$ optimizing LOTZ. Let $n = 2^k \cdot \ell - 1$ where $k \in \mathbb{N}_0$ and ℓ is odd. Then after expected $O(n^3)$ iterations the following holds. If $L \leq (\ell + 1)/2$, all solutions in A_t have mutually incomparable box index vectors at coarseness level $k + 1$. If $(\ell + 1)/2 < L \leq \ell$, all solutions in A_t have mutually incomparable box index vectors at coarseness level k and if $2^{j-1} \ell < L \leq 2^j \ell$ for $j \in [k]$ all solutions in A_t have mutually incomparable box index vectors at coarseness level $k - j$.*

PROOF. With Lemma 3.3 we can assume that $A_t \subset P$. Then, Lemma 3.9 (1) shows that the current solution s is always added to A_t if Algorithm 1 makes a decision at coarseness level j . It is also added if Algorithm 1 makes a decision at coarseness level $k + 1$ and there is a solution $x \in A_t$ whose box index vector is weakly dominated by the box index vector of a solution from $A_t \cup \{c\}$

at coarseness level $k + 1$. Here c is the candidate solution created from s by mutation. As long as one of these two happens, the current solution s performs the same random walk as in the proof of Theorem 4.3. Suppose that $L \leq (\ell + 1)/2$. Note that there are at least L many search points with pairwise incomparable box index vectors at coarseness level $k + 1$ by Lemma 3.10. These are found by the random walk after expected $O(n^3)$ iterations since it is sufficient that the current solution visits all Pareto-optimal points. The same happens if $(\ell + 1)/2 < L \leq \ell$ with coarseness level k instead of $k + 1$ and if $2^{j-1} \ell < L \leq 2^j \ell$ with coarseness level $k - j$ instead of $k + 1$ for $j \in [k - 1]$. This concludes the proof. \square

4 RUNTIME ANALYSIS OF PAES-25 WITH STANDARD BIT MUTATION ON LOTZ

In this section, we investigate the performance of PAES-25 on the bi-objective LOTZ benchmark using standard bitwise mutation. The experimental results reported in [28] indicate that PAES-25 with standard bit mutation requires more time to discover a constant fraction of Pareto-front than PAES-25 with one-bit mutation. This slower progress arises because the current solution may not remain Pareto-optimal even after a point on the Pareto front has been found. For instance, a two-bit flip can transform a Pareto-optimal solution into a new, non-dominated one (with respect to the archive A_t) by simultaneously increasing the LO-value and decreasing the TZ-value by more than one. As a result, the search dynamics under standard bit mutation are more complex than under one-bit mutation, and the resulting random walk may involve non-Pareto-optimal solutions. Rather than analyzing this random walk in detail, we instead focus on the hypervolume $\text{HV}(A_t)$ of the archive A_t . We show that $\text{HV}(A_t)$ cannot decrease in any iteration of PAES-25 with respect to any reference point h satisfying $h_i \leq 0$ for each $i \in [m]$. To maintain generality, we present our result for an arbitrary fitness function $f : \{0, 1\}^n \rightarrow \mathbb{N}_0^m$.

LEMMA 4.1. *Let $f : \{0, 1\}^n \rightarrow \mathbb{N}_0^m$ be a fitness function and S be a maximum set of mutually incomparable solutions with respect to f . Let t be any iteration of PAES-25 with current archive A_t with size of at least $|S|$ on f . Then $\text{HV}(A_{t+1}) \geq \text{HV}(A_t)$ with respect to any reference point h with $h_i \leq 0$ for all $i \in [m]$. Particularly, if the candidate solution c dominates an $x \in A_t$, $\text{HV}(A_{t+1}) > \text{HV}(A_t)$.*

PROOF. Note that a search point will only be removed from A_t if a candidate solution c is generated that is added to A_t . If A_t is full and the candidate solution c is added to A_t , c weakly dominates an $x \in A_t$ (since $|A_t| \geq |S|$). If there is $x \in A_t$ with $f(x) = f(c)$ then $\text{HV}(A_{t+1}) = \text{HV}(A_t)$. Suppose not and let $B \subset A_t$ be the set of search points dominated by c . Then $\text{HV}(A_{t+1}) = \text{HV}((A_t \setminus B) \cup \{c\}) = \text{HV}(A_t \cup \{c\})$. Let $v := f(c)$. Since any search point z with $v_i - 1/2 \leq f_i(z) \leq v_i$ for every $i \in [m]$ is weakly dominated by c , but not weakly dominated by any $y \in A_t$, we have that $\text{HV}(A_t \cup \{c\}) > \text{HV}(A_t)$. \square

Using the maximum possible hypervolume of an archive A_t with respect to the reference point $h = (-1, -1)$, we can estimate the expected number of iterations required for the current solution to reach the Pareto front when PAES-25 optimizes LOTZ.

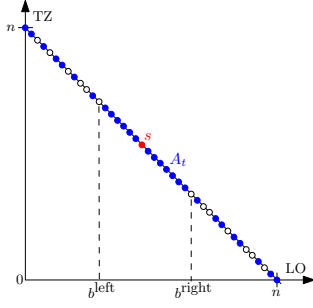


Figure 1: Illustration of b^{left} and b^{right} for the case $b^{\text{left}} \neq -\infty$ and $b^{\text{right}} \neq \infty$. The current solution s quite in the middle is depicted in red and the other Pareto-optimal points from the archive A_t are represented by the remaining filled points in blue. The white points with black outlines are the holes.

LEMMA 4.2. *Consider PAES-25 with standard bit mutation optimizing m -LOTZ for $m = 2$. Then for every possible A_t the current solution s is Pareto-optimal in expected $O(n^3)$ iterations.*

PROOF. Suppose that $c \notin P$. By flipping one specific bit to increase the LO or TZ value of s while the other bits remain unchanged (happening with probability at least $1/(en)$) one can create a solution c which (strictly) dominates s . Since A_t contains no solution dominating s , and dominance is transitive, c is not weakly dominated by any solution in A_t , and is therefore added to A_t which becomes the current solution for the next iteration. By Lemma 4.1 the hypervolume $HV(A_t)$ increases. By Lemma 3.7 we obtain (for $a = 0$ and $b = n$) $HV(P) = (n+2)(n+1)/2 = O(n^2)$ and $HV(A_t)$ cannot decrease by Lemma 4.1. Hence, the expected number of iterations until the current solution is Pareto-optimal is $O(n^3)$. \square

We will also show that, in $O(n^3)$ expected time, a new Pareto-optimal search point is added to the archive, even when a Pareto-optimum already exists in A_t . However, establishing this result requires a different proof technique than used in Lemma 4.2, as the current solution may be also Pareto-optimal. This insight is key to deriving the overall runtime of PAES-25 on LOTZ.

THEOREM 4.3. *Consider PAES-25 with standard bit mutation and archive size at least $n+1$ optimizing m -LOTZ for $m = 2$. Then the expected number of iterations until some Pareto-optimal search point x with $x \notin A_t$ is added to A_t is $O(n^3)$. This implies that the expected number of iterations until $A_t = P$ is at most $O(n^4)$.*

PROOF. By Lemma 4.2 the expected number of iterations until $s \in P$ is $O(n^3)$. Now suppose that $s \in P$ and not the complete set P of Pareto-optimal individuals is contained in A_t . Consider $b^{\text{left}} := b^{\text{left}}(s) := \sup\{\ell \in \{0, \dots, n-1\} \mid \ell < \text{LO}(s) \text{ and there is no } x \in A_t \cap P \text{ with } \text{LO}(x) = \ell\}$ and $b^{\text{right}} := b^{\text{right}}(s) := \inf\{\ell \in \{1, \dots, n\} \mid \text{LO}(s) < \ell \text{ and there is no } x \in A_t \cap P \text{ with } \text{LO}(x) = \ell\}$. Then b^{left} denotes the largest LO value among the individuals from $P \setminus A_t$ that have a smaller LO value than s . If all such individuals with a smaller LO value than s are already present in A_t , we have $b^{\text{left}} = -\infty$.

Similarly, b^{right} denotes the smallest LO value among the individuals from $P \setminus A_t$ that have a larger LO value than s (see also

Figure 1). If there is no such individual with a larger LO value than s , then $b^{\text{right}} = \infty$. As long as $b^{\text{left}} + 1 \leq \text{LO}(s) \leq b^{\text{right}} - 1$, the current solution s remains Pareto-optimal (since a solution y with $b^{\text{left}} + 1 \leq \text{LO}(y) \leq b^{\text{right}} - 1$ is dominated by the Pareto-optimal $z \in P_t$ with $\text{LO}(z) = \text{LO}(y)$) and the LO(s)-value of the current solution s performs a random walk on the discrete interval $[b^{\text{left}} + 1, b^{\text{right}} - 1] \cap \{0, \dots, n\}$.

Since s remains Pareto-optimal during this process, it is updated in each iteration when the candidate solution c is also Pareto-optimal. Note also that $b^{\text{left}} \neq -\infty$ or $b^{\text{right}} \neq \infty$ since $A_t \neq P$.

Denote an iteration as a *failure* if one of the following two conditions is met.

- (1) The value of $\text{LO}(s)$ changes by at least four.
- (2) The value $\text{LO}(s)$ exits $I := [b^{\text{left}} + 1, b^{\text{right}} - 1]$, without resulting in a Pareto-optimal individual x with $x \notin A_t$.

In the following we show that in $O(n^3)$ iterations in expectation a failure occurs or a Pareto-optimal search point x with $x \notin A_t$ is added to A_t . To this end, it is sufficient to show that in expected $O(n^3)$ iterations the value $\text{LO}(s)$ exits I or a failure occurs. The aim is to apply Theorem 2.1. However, there may be a negative drift away from the borders of I as the example $b^{\text{right}} = \infty$, $b^{\text{left}} = 0$ and $s = 10^{n-1}$ shows: Creating $c = 1^2 0^{n-2}$ happens with the same probability as creating $c = 1^j 0^{n-j}$ for $j \geq 3$. Hence, this theorem is not straightforwardly applicable. However, this negative drift vanishes, if no failure occurs and $s \in [b^{\text{left}} + 4, b^{\text{right}} - 4]$, allowing the theorem to apply for proving the following claim.

CLAIM 1. *After $O(n^3)$ iterations in expectation a failure occurs, or $\text{LO}(s) \notin [b^{\text{left}} + 4, b^{\text{right}} - 4]$ is satisfied for every possible $b^{\text{left}} < b^{\text{right}}$ with $b^{\text{left}} \neq -\infty$ or $b^{\text{right}} \neq \infty$, and for every possible Pareto-optimal starting point s_0 with $\text{LO}(s_0) \in [b^{\text{left}} + 4, b^{\text{right}} - 4]$.*

PROOF OF CLAIM. We may assume that no failure occurs, as any failure would only accelerate the achievement of the goal outlined in this claim. Without loss of generality assume that $\text{LO}(s) \in [b^{\text{left}} + 4, b^{\text{right}} - 4]$ for all $0 \leq t \leq T$ where $T := \inf\{t \geq 0 \mid \text{LO}(s) \notin [b^{\text{left}} + 4, b^{\text{right}} - 4]\}$. For $j \in \{-3, \dots, 3\}$ denote by p_j the probability that $\text{LO}(s)$ changes by j (which also means that s stays Pareto-optimal). Then $p_{-j} = p_j$ and $p_j = \Theta(1/n^j)$ since it is required to flip j specific bits. Let $X_t := \text{LO}(s)$ where s denotes the current solution in iteration t . We consider three cases.

Case 1: Suppose that $b^{\text{right}} \neq \infty$ and $b^{\text{left}} \neq -\infty$ and let $d := (b^{\text{right}} + b^{\text{left}})/2$. Let $Y_t := |X_t - d|$. Then we have

$$E(Y_{t+1} - Y_t \mid X_0, \dots, X_t) = E(|X_{t+1} - d| - |X_t - d| \mid X_0, \dots, X_t).$$

If $X_t \geq d$ we obtain

$$\begin{aligned} &= E(|X_{t+1} - d| - X_t + d \mid X_0, \dots, X_t) \\ &= E(X_{t+1} - X_t \mid X_0, \dots, X_t) = \sum_{j=-3}^3 j \cdot p_j = 0. \end{aligned}$$

On the other hand, if $X_t < d$,

$$\begin{aligned} &E(|X_{t+1} - d| - |X_t - d| \mid X_0, \dots, X_t) \\ &\geq E(-X_{t+1} + d + X_t - d \mid X_0, \dots, X_t) \\ &= E(X_t - X_{t+1} \mid X_0, \dots, X_t) = 0. \end{aligned}$$

Further,

$$\text{Var}(Y_{t+1} - Y_t \mid Y_0, \dots, Y_t) \geq E((Y_{t+1} - Y_t)^2 \mid Y_0, \dots, Y_t) \geq \delta$$

for $\delta := 1/(en)$ since flipping a specific bit (the first zero from the left if $X_t \geq d$ to increase X_t and otherwise the first one from the right to decrease X_t) changes Y_t by one. This happens with probability at least $(1 - 1/n)^{n-1} \cdot 1/n \geq 1/(en)$. So both conditions of Theorem 2.1 are satisfied and we obtain $E(T') \leq n^2/\delta = O(n^3)$ for

$$T' := \inf \left\{ t \geq 0 \mid Y_t > \frac{b^{\text{right}} - b^{\text{left}}}{2} - 4 \right\} :$$

We see that $b^{\text{left}} + 4 \leq X_t \leq b^{\text{right}} - 4$ iff $(b^{\text{left}} - b^{\text{right}})/2 + 4 \leq X_t - d \leq (b^{\text{right}} - b^{\text{left}})/2 - 4$ iff $Y_t \leq (b^{\text{right}} - b^{\text{left}})/2 - 4$. This also yields $T = T'$.

Case 2: Suppose that $b^{\text{right}} \neq \infty$ and $b^{\text{left}} = -\infty$. Then we have $T = \inf \{t \geq 0 \mid X_t > b^{\text{right}} - 4\}$. If $X_t \geq 3$ we obtain $E(X_{t+1} - X_t \mid X_0, \dots, X_t) = \sum_{j=3}^3 j \cdot p_j = 0$ since $3 \leq X_t \leq b^{\text{right}} - 4$. Further, if $X_t = j$ for $j \in \{0, 1, 2\}$, then X_t cannot decrease by more than j and we obtain $E(X_{t+1} - X_t \mid X_0, \dots, X_t) = \sum_{i=-j}^j i \cdot p_i + \sum_{i=j+1}^3 i \cdot p_i > 0$. As above, $\delta := \text{Var}(X_{t+1} - X_t \mid X_0, \dots, X_t) = \Omega(1/n)$. So with Theorem 2.1 we obtain $E(T) \leq n^2/\delta = O(n^3)$ since $b^{\text{right}} - 4 \in O(n)$.

Case 3: Suppose that $b^{\text{right}} = \infty$ and $b^{\text{left}} \neq -\infty$. Let $Y_t := n - X_t$. Then $T = \inf \{t \geq 0 \mid X_t < b^{\text{left}} - 4\} = \inf \{t \geq 0 \mid Y_t > n - b^{\text{left}} + 4\}$, and similar as in Case 2, $E(Y_{t+1} - Y_t \mid Y_0, \dots, Y_t) \geq 0$ and $\text{Var}(Y_{t+1} - Y_t \mid Y_0, \dots, Y_t) = \Omega(1/n)$. Therefore, $E(T) = O(n^3)$ with Theorem 2.1. All the above cases conclude the proof of Claim 1. ■

We show in the next claim that $\text{LO}(s)$ even exits $I := [b^{\text{left}} + 1, b^{\text{right}} - 1]$ in expected $O(n^3)$ iterations if no failure occurs. This is because, if $\text{LO}(s) \notin [b^{\text{left}} + 4, b^{\text{right}} - 4]$, then $\text{LO}(s)$ only needs to be changed by three specific one-bit flips in the next three iterations when the fitness of s changes.

CLAIM 2. *After $O(n^3)$ iterations in expectation a failure occurs or $\text{LO}(s) \notin [b^{\text{left}} + 1, b^{\text{right}} - 1]$ is satisfied for every possible $b^{\text{left}} < b^{\text{right}}$ with $b^{\text{left}} \neq -\infty$ or $b^{\text{right}} \neq \infty$, and for every possible starting point s_0 with $\text{LO}(s_0) \in [b^{\text{left}} + 1, b^{\text{right}} - 1]$.*

PROOF OF CLAIM. As in the proof of the claim before, we may assume that no failure occurs. Denote an iteration as *profitable* if a candidate solution c is created with $\text{LO}(c) = \text{LO}(s) + 1$ if $\text{LO}(s) > b^{\text{right}} - 4$ or with $\text{LO}(c) = \text{LO}(s) - 1$ if $\text{LO}(s) < b^{\text{left}} + 4$. Note that if $\text{LO}(s) \notin [b^{\text{left}} + 4, b^{\text{right}} - 4]$, then also $\text{LO}(c) \notin [b^{\text{left}} + 4, b^{\text{right}} - 4]$ if t is profitable. With Claim 1 we can assume that $\text{LO}(s) \notin [b^{\text{left}} + 4, b^{\text{right}} - 4]$. Suppose that $\text{LO}(s) > b^{\text{right}} - 4$ (i.e. $b^{\text{right}} \neq \infty$). To exit the interval $[b^{\text{left}} + 1, b^{\text{right}} - 1]$, it is sufficient that the next $j = b^{\text{right}} - \text{LO}(s) \leq 3$ iterations where the fitness of s changes (i.e. the candidate solution c with $f(c) \neq f(s)$ becomes the current solution for the next iteration) are profitable. In the following we estimate the probability that an iteration where the fitness of s changes is profitable. The probability that the fitness of s changes is at most $2/n$ since it is required to either flip the first zero from the left (to increase $\text{LO}(s)$ but decrease $\text{TZ}(s)$) or the last one from the left (to decrease $\text{LO}(s)$ but increase $\text{TZ}(s)$). The probability to perform a specific one-bit flip is $1/n \cdot (1 - 1/n)^{n-1} \geq 1/(en)$. Hence, as long as $s \in [b^{\text{left}} + 1, b^{\text{right}} - 1]$, an iteration where the fitness changes is profitable with probability at least $1/(en) \cdot 1/(2n) = 1/(2e) = \Omega(1)$ and hence, the next $j \leq 3$ iterations where the

fitness changes are profitable with probability $\Omega(1)$. If this happens, s exits I . By symmetry, the same also holds if $\text{LO}(s) < b^{\text{left}} + 4$.

Denote the next j iterations during which the fitness of s changes, as a *trial*, and call the trial *successful* if the interval $[b^{\text{left}} + 1, b^{\text{right}} - 1]$ is exited by s . Note that the expected number of iterations until s exits I is $O(n)$ if the first trial is successful since the expected waiting time for a fitness change of s is $O(n)$. If the first trial is not successful, we repeat the arguments above (possibly those included in Claim 1). Since the probability that the first trial is successful is $\Omega(1)$, we see that the expected number of repetitions until a successful trial occurs is $O(1)$. This proves Claim 2. ■

Claim 2 also shows that in expected $O(n^3)$ iterations a Pareto-optimal search point x with $x \notin A_t$ is added to A_t or a failure occurs. Now we estimate the probability that within $O(n^3)$ iterations a failure occurs. The probability that $\text{LO}(s)$ changes by at least four if s is Pareto-optimal is $O(1/n^4)$ since it is required to flip a specific block of four ones or four zeros. By a union bound, the probability is $o(1)$ that such a change occurs in $O(n^3)$ iterations. Further, we argue that a new Pareto-optimal solution is created with probability $\Omega(1)$ when $\text{LO}(s)$ exits the interval $[b^{\text{left}} + 1, b^{\text{right}} - 1]$. Let j be the smallest possible Hamming distance between s and a Pareto-optimal solution x not already in A_t . Then the probability that $c = x$ is at least $1/(en^j)$ while s exits the interval with probability at most $2/n^j$ since it is required to either increase $\text{LO}(s)$ by at least j or decrease $\text{LO}(s)$ by at least j . Otherwise, s remains Pareto-optimal. Hence, with probability $\Omega(1)$ the candidate solution c is a new Pareto-optimal solution if s exits I . This yields that a failure does not occur with probability $\Omega(1)$ by a union bound on both possible failures within the $O(n^3)$ iterations from Claim 2 above. If a failure occurs within these $O(n^3)$ iterations we repeat the above arguments, including those that s reaches the Pareto front and then exiting the interval $[b^{\text{left}} - 1, b^{\text{right}} + 1]$. The expected number of repetitions is $O(1)$ concluding the proof. □

5 PAES-25 FAILS ON OMM AND COCZ

In this section, we demonstrate that PAES-25 using either one-bit mutation or standard bit mutation performs poorly on other pseudo-Boolean benchmarks such as OMM or COCZ. This inefficiency arises from the update mechanism. The current solution is replaced by a candidate solution s whenever s is added to the archive A_t and A_t is not full. Our theoretical results and proofs naturally extend to the many-objective setting if the number of objectives is constant. Recall that OMM [21] is defined as $\text{OMM}(x) = (\sum_{i=1}^n x_i, \sum_{i=1}^n (1 - x_i))$, where the first component counts the number of ones in x , and the second counts the number of zeros. The function COCZ [30] is defined as $\text{COCZ}(x) = (\sum_{i=1}^n x_i, \sum_{i=1}^{n/2} x_i + \sum_{i=n/2+1}^n (1 - x_i))$. Here, the first component again counts the number of ones, while the second component counts the number of ones in the first half and the number of zeros in the second half of the bit string.

THEOREM 5.1. *Consider PAES-25 with standard bit mutation or one-bit mutation and archive size at least $n + 1$ optimizing OMM. Let $0 < \alpha \leq 1/2$ be a constant. Then there is a constant $\beta > 0$ such that with probability $1 - e^{-\Omega(n)}$ the value $|s|_1$ where s is the current solution does not exit the interval $[n/2 - \alpha n, n/2 + \alpha n]$ within $2^{\beta n}$ iterations. Hence, with probability $1 - e^{-\Omega(n)}$ the fraction of the Pareto front contained in A_t is $o(n)$ after $2^{\Omega(n)}$ iterations.*

PROOF. By a classical Chernoff bound, we have that $n/2 - \alpha n/2 \leq s \leq n/2 + \alpha n/2$ with probability $1 - e^{-\Omega(n)}$ in the first iteration. Suppose that this happens. Note that a new created candidate solution is either non-dominated or has the same fitness value as a solution $x \in A_t$. In both cases it is added to A_t . Hence, the current solution is updated in each iteration no matter how many bits in s are flipped. Denote by X_t the number of ones in s in iteration t . Then we obtain $E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b) \leq b/n - (n - b)/n = -(n - 2b)/n = -\alpha$ for $a := n/2 - \alpha n$ and $b := n/2 + \alpha n/2$. Hence by the negative drift theorem (see for example Theorem 2.4.9 in [35]) there is a constant $\beta > 0$ such that $X_t < a$ within the next $2^{\beta n}$ iterations with probability $e^{-\Omega(n)}$. By changing the roles of ones and zeros we also see that $X_t > n/2 + \alpha n$ within $2^{\beta n}$ iterations with probability $e^{-\Omega(n)}$. By a union bound we obtain that $n/2 - \alpha n \leq X_t \leq n/2 + \alpha n$ within the next $2^{\beta n}$ iterations with probability $1 - e^{-\Omega(n)}$. \square

This negative result directly transfers to OJZJ $_k$ for $2 \leq k \leq n/2 - cn$ for a constant $0 < c < 1/2$ since it coincides with OMM if $n/2 - cn \leq |s|_1 \leq n/2 + cn$ up to the additive vector (k, k) . We formulate a similar result for COCZ. For a bit string $x \in \{0, 1\}^n$ denote by x^1 its first half and by x^2 its second if 2 divides n .

THEOREM 5.2. *Suppose that 2 divides n . Consider PAES-25 with standard bit mutation or one-bit mutation and archive size at least $n/2 + 1$ optimizing COCZ. Let $0 < \alpha \leq 1/4$ be a constant. Then there is a constant $\beta > 0$ such that $|s^2|_1$ does not exit the interval $[n/4 - \alpha n, n/4 + \alpha n]$ within $2^{\beta n}$ iterations with probability $1 - e^{-\Omega(n)}$. Hence, with probability $1 - e^{-\Omega(n)}$ the fraction of the Pareto front contained in A_t is $o(n)$ after $2^{\Omega(n)}$ iterations.*

PROOF. We can assume that $\alpha < 1/4$ since $|s^2|_1$ never exits the interval $[0, n/2]$. By a classical Chernoff bound, we have that $n/4 - \alpha n/2 \leq |s^2|_1 \leq n/4 + \alpha n/2$ with probability $1 - e^{-\Omega(n)}$. Suppose that this happens. Let B be the event that no bit is flipped in the first half of the bit string during mutation. If B occurs, a new created candidate solution is either non-dominated or has the same fitness value as a solution $x \in A_t$ and we are in the same situation as in the proof of Theorem 5.1. Denote by X_t the number of ones in the second half of the current solution in iteration t . Then we obtain for $a := n/4 - \alpha n$ and $b := n/4 - \alpha n/2$ also $E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b; B) \leq b/n - (n/2 - b)/n = -\alpha$. Further we obtain $E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b; \bar{B}) \leq n^{-\Omega(n)}$: If the offspring was not accepted then the latter becomes zero. Suppose that s is accepted by either flipping k ones and ℓ zeros or flipping ℓ ones and k zeros for $k, \ell \leq n/4 - \alpha n$ with $k < \ell$ in the second half. The probability for the first event is larger than the second and hence, conditioned on the union of both events yields $E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b; \bar{B}) \leq 0$. Hence, by the law of total probability, $E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b; \bar{B}) \leq 0$ if at most $n/4 - \alpha n$ many ones and zeros are flipped. Note that the probability to flip more than $n/4 - \alpha n$ ones or zeros is $n^{-\Omega(n)}$ (since $n/4 - \alpha n = \Omega(n)$) which can increase X_t by at most $n/2 - a = n/4 + \alpha n/2 = O(n)$. Hence, in total again by the law of total probability $E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b; \bar{B}) \leq n^{-\Omega(n)}$. Again, applying the law of total probability yields $E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b) \leq E(X_t - X_{t+1} \mid X_0, \dots, X_t; a < X_t < b; B)P(B) + n^{-\Omega(n)} \leq -\alpha/(2e)$ due to $P(B) = (1 - 1/n)^{n/2} \geq 1/e$ for

n sufficiently large. Hence, again by the negative drift theorem, there is a constant $\beta > 0$ such that $X_t < a$ within the next $2^{\beta n}$ iterations with probability $e^{-\Omega(n)}$. By changing the roles of ones and zeros we also see that $X_t > n/4 + \alpha n$ within $2^{\beta n}$ iterations with probability $e^{-\Omega(n)}$. By a union bound we obtain that $n/4 - \alpha n \leq X_t \leq n/4 + \alpha n$ within the next $2^{\beta n}$ iterations with probability $1 - e^{-\Omega(n)}$. \square

6 CONCLUSIONS

In this paper, we rigorously analyzed the performance of PAES-25 on the classical benchmark functions m -LOTZ, COCZ, and OMM. Notably, PAES-25 with one-bit mutation performs surprisingly well on m -LOTZ, and we were able to derive tight runtime bounds in the many-objective setting. To the best of our knowledge, these are the first tight bounds established for an MOEA on a classical benchmark problem. Further, our analysis reveals new insights into the behavior of local search-based algorithms on classical pseudo-Boolean benchmarks. In particular, we showed that the time of PAES-25 with one-bit mutation finding the Pareto front of m -LOTZ is, up to a factor of n/m , asymptotically the same as the cover time of a simple random walk on an $m/2$ -dimensional grid graph with $(2n/m)^{m/2}$ nodes, where the latter is well known. Further, PAES-25 with one-bit mutation achieves a good distribution of the archive across the Pareto front when the Adaptive Grid Archiver (AGA) is used, even when the archive size is smaller than the size of the Pareto front. In the bi-objective setting, good distributions for small archive sizes are also achieved when PAES-25 uses the Hypervolume based Archiver (HVA) or the Multi-level Grid Archiver (MGA). A further contribution is that we derived upper runtime bounds for PAES-25 on LOTZ for standard bit mutation. A major challenge in this analysis is that the current solution does not necessarily remain on the Pareto front. We also demonstrated that PAES-25 has still some limitations. It does not perform well on COCZ and OMM, since it covers only a sublinear fraction of the Pareto front in $e^{\Omega(n)}$ iterations with probability $1 - e^{-\Omega(n)}$.

We hope that the techniques developed mark a significant step forward in understanding the behavior of PAES-25, while also shedding light on its strengths and limitations. We are confident that our analysis carries over to other MOEAs with similar working principles as the PAES-25 like only mutating a current solution instead of choosing a parent uniformly at random. To deepen this understanding, future work could focus on more variants of PAES-25 which follow a $(\mu + \lambda)$ or (μ, λ) scheme to possibly overcome the difficulties with OMM or COCZ. Another research direction is to extend our analysis to more complex benchmark problems, such as FRITZ, LITZ or sLITZ proposed in [28], where PAES-25 has also been shown to perform empirically well. We also hope that the gained deeper theoretical insights into the dynamic of PAES-25 will provide valuable guidance for practitioners, helping to develop refined versions of the algorithm, especially for problems with non-flat fitness landscapes or featuring many local optima.

REFERENCES

- [1] Anne Auger, Johannes Bader, Dimo Brockhoff, and Eckart Zitzler. 2012. Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications. *Theoretical Computer Science* 425 (2012), 75–103.
- [2] Chao Bian, Chao Qian, and Ke Tang. 2018. A general approach to running time analysis of multi-objective evolutionary algorithms. In *International Joint*

- Conference on Artificial Intelligence (IJCAI 2018)*. ijcai.org, 1405–1411.
- [3] Chao Bian, Shengjie Ren, Miqing Li, and Chao Qian. 2024. An archive can bring provable speed-ups in multi-objective evolutionary algorithms. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI 2024)*. ijcai.org, 6905 – 6913.
 - [4] Chao Bian, Yawen Zhou, Miqing Li, and Chao Qian. 2023. Stochastic population update can provably be helpful in multi-objective evolutionary algorithms. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI 2023)*. 5513–5521.
 - [5] Karl Bringmann and Tobias Friedrich. 2010. An Efficient Algorithm for Computing Hypervolume Contributions. *Evolutionary Computation* 18, 3 (2010), 383–402.
 - [6] Sacha Cerf, Benjamin Doerr, Benjamin Hebras, Yakob Kahane, and Simon Wietheger. 2023. The First Proven Performance Guarantees for the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) on a Combinatorial Optimization Problem. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2023)*. ijcai.org, 5522–5530.
 - [7] C Coello, D Van Veldhuizen, and G Lamont. 2013. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer US.
 - [8] Duc-Cuong Dang, Andre Opris, Bahare Salehi, and Dirk Sudholt. 2023. Analysing the Robustness of NSGA-II under Noise. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'23)*. ACM Press, 642–651.
 - [9] Duc-Cuong Dang, Andre Opris, and Dirk Sudholt. 2024. Crossover can guarantee exponential speed-ups in evolutionary multi-objective optimisation. *Artificial Intelligence* 330 (2024), 104098.
 - [10] Duc-Cuong Dang, Andre Opris, and Dirk Sudholt. 2024. Illustrating the Efficiency of Popular Evolutionary Multi-Objective Algorithms Using Runtime Analysis. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2024)*. ACM Press, 484–492.
 - [11] Duc-Cuong Dang, Andre Opris, and Dirk Sudholt. 2024. Level-Based Theorems for Runtime Analysis of Multi-objective Evolutionary Algorithms. In *Proceedings of the International Conference on Parallel Problem Solving from Nature (PPSN XVIII)*. Springer, Berlin, Heidelberg, 246–263.
 - [12] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6 (2002), 182–197.
 - [13] Renzhong Deng, Weijie Zheng, Mingfeng Li, Jie Liu, and Benjamin Doerr. 2024. Runtime Analysis for State-of-the-Art Multi-objective Evolutionary Algorithms on the Subset Selection Problem. In *Proceedings of the International Conference on Parallel Problem Solving from Nature (PPSN XVIII)*. Springer, 264–279.
 - [14] Benjamin Doerr, Tudor Ivan, and Martin S. Krejca. 2025. Speeding Up the NSGA-II With a Simple Tie-Breaking Rule. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2025)*. AAAI Press, 26964–26972.
 - [15] Benjamin Doerr, Martin Krejca, and Andre Opris. 2025. Tight Runtime Guarantees From Understanding the Population Dynamics of the GSEMO Multi-Objective Evolutionary Algorithm. arXiv:2505.01266 [cs.NE] to appear at IJCAI 2025.
 - [16] Benjamin Doerr, Martin S. Krejca, and Günter Rudolph. 2025. Runtime Analysis for Multi-Objective Evolutionary Algorithms in Unbounded Integer Spaces. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2025)*, 26955–26963.
 - [17] Benjamin Doerr and Zhongdi Qu. 2022. A First Runtime Analysis of the NSGA-II on a Multimodal Problem. In *Proceedings of the International Conference on Parallel Problem Solving from Nature (PPSN XVII)*. Springer, 399–412.
 - [18] Benjamin Doerr and Zhongdi Qu. 2023. From Understanding the Population Dynamics of the NSGA-II to the First Proven Lower Bounds. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2023)*. AAAI Press, 12408–12416.
 - [19] Benjamin Doerr and Zhongdi Qu. 2023. Runtime Analysis for the NSGA-II: Provable Speed-Ups From Crossover. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2023)*. AAAI Press, 12399–12407.
 - [20] Oliver Giel. 2003. Expected runtimes of a simple multi-objective evolutionary algorithm. In *Congress on Evolutionary Computation, CEC 2003*. IEEE, 1918–1925.
 - [21] Oliver Giel and Per Kristian Lehre. 2010. On the Effect of Populations in Evolutionary Multi-Objective Optimisation. *Evolutionary Computation* 18, 3 (2010), 335–356.
 - [22] Satoshi Ikeda, Izumi Kubo, and Masafumi Yamashita. 2009. The hitting and cover times of random walks on finite graphs using local degree information. *Theoretical Computer Science* (2009), 94–100.
 - [23] Johan Jonasson. 2000. An upper bound on the cover time for powers of graphs. *Discrete Mathematics* 222, 1 (2000), 181–190.
 - [24] Joshua Knowles. 2002. *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. Ph. D. Dissertation. The University of Reading, United Kingdom.
 - [25] Joshua Knowles and David Corne. 1999. The Pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimisation. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99*. 98–105.
 - [26] Joshua Knowles and David Corne. 2000. Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy. *Evolutionary Computation* 8, 2 (2000), 149–172.
 - [27] Joshua Knowles and David Corne. 2003. Properties of an adaptive archiving algorithm for storing nondominated vectors. *Transactions of Evolutionary Computation* (2003), 100–116.
 - [28] Joshua Knowles and Arnaud Liefooghe. 2025. PAES-25: Local Search, Archiving, and Multi/Many-Objective Pseudo-Boolean Functions. In *Evolutionary Multi-Criterion Optimization*. Springer, 60–73.
 - [29] Timo Kötzing. 2024. Theory of Stochastic Drift. arXiv:2406.14589 [math.PR]
 - [30] Marco Laumanns, Lothar Thiele, and Eckart Zitzler. 2004. Running Time Analysis of Multiobjective Evolutionary Algorithms on Pseudo-Boolean Functions. *IEEE Transactions on Evolutionary Computation* 8, 2 (2004), 170–182.
 - [31] Marco Laumanns, Lothar Thiele, and Eckart Zitzler. 2004. Running time analysis of multiobjective evolutionary algorithms on pseudo-Boolean functions. *IEEE Transactions on Evolutionary Computation* 8 (2004), 170–182.
 - [32] Marco Laumanns, Lothar Thiele, Eckart Zitzler, Emo Welzl, and Kalyanmoy Deb. 2002. Running time analysis of multi-objective evolutionary algorithms on a simple discrete optimization problem. In *Parallel Problem Solving from Nature, PPSN 2002*. Springer, 44–53.
 - [33] Marco Laumanns and Rico Zenklusen. 2011. Stochastic convergence of random search methods to fixed size Pareto front approximations. *European Journal of Operational Research* 213, 2 (2011), 414–421.
 - [34] Johannes Lengler. 2020. Drift Analysis. In *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*. Springer, 89–131.
 - [35] Johannes Lengler. 2020. A General Dichotomy of Evolutionary Algorithms on Monotone Functions. *IEEE Transactions on Evolutionary Computation* 24, 6 (2020), 995–1009.
 - [36] Manuel López-Ibáñez, Joshua Knowles, and Marco Laumanns. 2011. On Sequential Online Archiving of Objective Vectors. In *Evolutionary Multi-Criterion Optimization*. Springer Berlin Heidelberg, 46–60.
 - [37] Manuel López-Ibáñez, Joshua Knowles, and Marco Laumanns. 2011. On sequential online archiving of objective vectors. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Nature, 46–60.
 - [38] Soivi Luukkainen, Helle W. van den Maagdenberg, Michael T.M. Emmerich, and Gerard J.P. van Westen. 2023. Artificial Intelligence in Multi-Objective Drug Design. *Current Opinion in Structural Biology* 79 (2023), 102537.
 - [39] Andre Opris. 2025. A First Runtime Analysis of NSGA-III on a Many-Objective Multimodal Problem: Provable Exponential Speedup via Stochastic Population Update. arXiv:2505.01256 [cs.NE] to appear at IJCAI 2025.
 - [40] Andre Opris. 2025. A Many Objective Problem Where Crossover is Provably Indispensable. In *Proceedings of the International Joint Conference on Artificial Intelligence (AAAI 2025)*. AAAI Press, 27108–27116.
 - [41] Andre Opris, Duc-Cuong Dang, Frank Neumann, and Dirk Sudholt. 2024. Runtime Analyses of NSGA-III on Many-Objective Problems. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'24)*. ACM Press, 1596–1604.
 - [42] Xue Peng, Xiaoyun Xia, Weizhi Liao, and Zhanwei Guo. 2018. Running time analysis of the Pareto archived evolution strategy on pseudo-Boolean functions. *Multimedia Tools Application* 77, 9 (2018), 11203–11217.
 - [43] Chao Qian, Yang Yu, Ke Tang, Xin Yao, and Zhi-Hua Zhou. 2019. Maximizing submodular or monotone approximately submodular functions by multi-objective evolutionary algorithms. *Artificial Intelligence* 275 (2019), 279–294.
 - [44] Shengjie Ren, Chao Bian, Miqing Li, and Chao Qian. 2024. A first running time analysis of the Strength Pareto Evolutionary Algorithm 2 (SPEA2). In *Proceedings of the International Conference of Parallel Problem Solving from Nature (PPSN XVIII)*. Springer, 295–312.
 - [45] Dirk Thierens. 2003. Convergence time analysis for the multi-objective counting ones problem. In *Evolutionary Multi-Criterion Optimization, EMO 2003*. Springer, 355–364.
 - [46] Simon Wietheger and Benjamin Doerr. 2023. A Mathematical Runtime Analysis of the Non-dominated Sorting Genetic Algorithm III (NSGA-III). In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI 2023*. ijcai.org, 5657–5665.
 - [47] Simon Wietheger and Benjamin Doerr. 2024. Near-Tight Runtime Guarantees for Many-Objective Evolutionary Algorithms. In *Proceedings of the International Conference of Parallel Problem Solving from Nature (PPSN XVIII)*. Springer, 153–168.
 - [48] Weijie Zheng and Benjamin Doerr. 2022. Better approximation guarantees for the NSGA-II by using the current crowding distance. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2022)*. ACM Press, 611–619.
 - [49] Weijie Zheng and Benjamin Doerr. 2024. Runtime Analysis for the NSGA-II: Proving, Quantifying, and Explaining the Inefficiency for Many Objectives. *IEEE Transactions on Evolutionary Computation* 28, 5 (2024), 1442–1454.
 - [50] Weijie Zheng and Benjamin Doerr. 2024. Runtime Analysis of the SMS-EMOA for Many-Objective Optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2024)*. AAAI press, 20874–20882.
 - [51] Weijie Zheng, Yufei Liu, and Benjamin Doerr. 2022. A First Mathematical Runtime Analysis of the Non-dominated Sorting Genetic Algorithm II (NSGA-II). In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2022)*. AAAI Press, 10408–10416.
 - [52] Eckart Zitzler. 1999. *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. Shaker Verlag GmbH, Germany.