Information Processing Letters 110 (2010) 474–480

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Whole mirror duplication-random loss model and pattern avoiding permutations

Jean-Luc Baril *, Rémi Vernay

LE2I UMR-CNRS 5158, Université de Bourgogne, B.P. 47 870, 21078 Dijon Cedex, France

ARTICLE INFO

Article history:
Received 29 June 2009
Received in revised form 10 February 2010
Accepted 20 April 2010
Available online 21 April 2010
Communicated by A.A. Bertossi

Keywords:
Algorithms
Combinatorial problems
Pattern avoiding permutation
Whole duplication-random loss model
Genome
Generating algorithm
Binary reflected Gray code

ABSTRACT

In this paper we study the problem of the whole mirror duplication-random loss model in terms of pattern avoiding permutations. We prove that the class of permutations obtained with this model after a given number p of duplications of the identity is the class of permutations avoiding the alternating permutations of length $2^p + 1$. We also compute the number of duplications necessary and sufficient to obtain any permutation of length n. We provide two efficient algorithms to reconstitute a possible scenario of whole mirror duplications from identity to any permutation of length n. One of them uses the well-known binary reflected Gray code (Gray, 1953) [10]. Other relative models are also considered.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction and notation

The well-known genome duplication consists in copying a part of the original genome inserted into itself, followed by the loss of one copy of each of the duplicated genes (see [2,9,11,14,17] for an explanation of different methods of duplication). From a formal point of view, a genome of n genes is represented by a permutation of length n. In a previous article, Chaudhuri et al. [7] investigated a variant called the tandem duplication-random loss model: the duplicated part (of size K) of the genome is inserted immediately after the original portion, followed by the loss procedure. This model comes from evolutionary biology where it has been applied to the vertebrate mitochondrial genomes. Chaudhuri et al. introduce a notion of distance between two genomes and they provide an algorithm to compute it efficiently for certain regions of the parameter space. Bouvel and Rossin [5] have also studied this model. They proved that the class of permutations obtained from the identity after p steps (of width K) is also a class of pattern avoiding permutations. More particularly, they investigate the restricted case of a *whole duplication* (*W-duplication* for short): the whole duplication consists in copying entirely the permutation on its right and the loss procedure consists to delete one of the two copies of each gene. Here, we give an example of the process of a W-duplication followed by the loss procedure on the permutation 123456:

$$123456 \rightsquigarrow 123456\underbrace{123456}_{\text{duplication}} \rightsquigarrow \underbrace{1\not2\not3\not4\not5\not6\not1\not2\not3\not4\not5\not6}_{\text{loss procedure}} \rightsquigarrow 124635.$$

So, they prove that the permutations obtained after p W-duplications is the class of permutations avoiding all minimal permutations with 2^p descents, minimal in the sense of pattern-involvement relation on permutations. Moreover, they computed the number of duplication-loss steps

* Corresponding author.
E-mail addresses: barjl@u-bourgogne.fr (J.-L. Baril), remi.vernay@u-bourgogne.fr (R. Vernay).

0020-0190/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ipl.2010.04.016

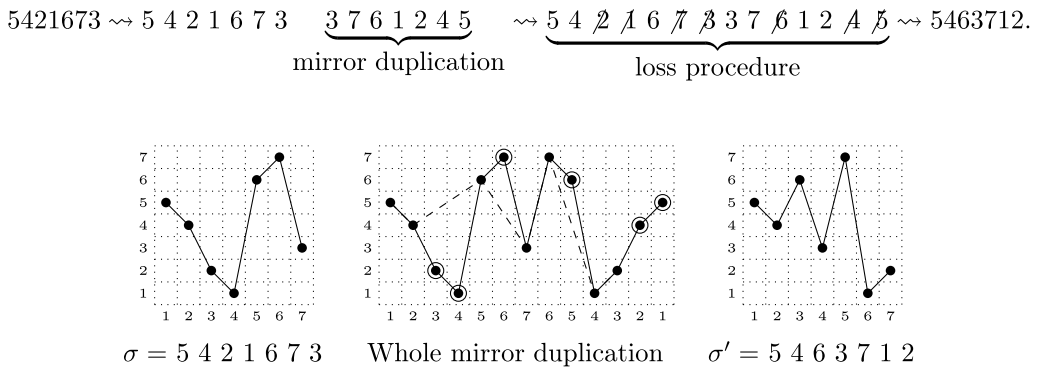


Fig. 1. A WM-duplication of the permutation 5421673. The encircled points are deleted by the loss procedure.

of width K necessary and sufficient to obtain any permutation. More recently, Bouvel and Pergola [4] showed a local and simpler characterization and several properties of the set of minimal permutations with 2^p descents.

In this article, we focus on the *whole mirror duplication-random loss model* (*WM-duplication* for short): it consists in copying the mirror of the permutation on its right followed by the loss procedure. This model very likely occurs in one half of eubacterial genomes, and possibly in most chromosomes [8,15]. Fig. 1 illustrates a *WM-duplication* for the permutation $\sigma = 5421673$:

$$5421673 \rightsquigarrow 5\,4\,2\,1\,6\,7\,3\ \underbrace{3\,7\,6\,1\,2\,4\,5}_{\text{mirror duplication}}$$
$$\rightsquigarrow \underbrace{5\,4\,\not{2}\,\not{1}\,6\,\not{7}\,\not{3}\,3\,7\,\not{6}\,1\,2\,\not{4}\,\not{5}}_{\text{loss procedure}}$$
$$\rightsquigarrow 5463712.$$

Let $\mathcal{S}_n$ denote the set of n-length *sequences* $s = s_1 s_2 \dots s_n$ of positive integers. The *mirror* of s is $\bar{s} = s_n s_{n-1} \dots s_1$. A *subsequence* of s is a sequence $s_{i_1} s_{i_2} \dots s_{i_m}$ for $1 \leqslant i_1 < i_2 < \cdots < i_m \leqslant n$. A subsequence is called a *substring* when the set $\{i_j,\ 1 \leqslant j \leqslant m\}$ is an interval, *i.e.* when the subsequence appears as consecutive elements in s. An *ascent* (resp. a *descent*) of s is any position i $(1 \leqslant i \leqslant n-1)$ with $s_i < s_{i+1}$ (resp. $s_i > s_{i+1}$). A *run up* (resp. *a run down*) of s is a substring (of length at least one) in which the elements are in increasing order (resp. decreasing order). More generally, a run up (or run down) is said to be *maximal* when it cannot be extended in a longer run up (resp. run down) in the sequence. We refer the reader to Rodney and Wilf [6] for results concerning the enumeration of permutations that have a given number of runs up and down. For instance, if $s = 5467312$ then position 4 is a descent, the substring 46 is a run up, and 467 is a maximal run up. Note that 5 is also a maximal run up. A *valley* of s is a substring which is a run down followed by a run up each of the length of at least two, *i.e.* a substring $s_k s_{k+1} \dots s_\ell$, $1 \leqslant k < \ell \leqslant n$, such that there is j $(k < j < \ell)$ verifying $s_k > s_{k+1} > \cdots > s_j$ and $s_j < s_{j+1} < \cdots < s_\ell$. A valley is *maximal* if the substring is maximal for this property. In the above example, the substrings 5467 and 7312 are the only maximal valleys of s. We denote by $\mathrm{val}(s)$ the number of maximal valleys in s, *i.e.* the cardinality of the set $\{j, s_j < \min\{s_{j-1}, s_{j+1}\}\}$. A recursive formula enumerating permutations with a given number of valleys can be found in [16].

On the other hand, a sequence s of length n is a *permutation* whenever each s_i is a distinct member of the $[n] = \{1, 2, \dots, n\}$. In the sequel, permutations will be denoted by Greek letters: $\sigma, \pi, \tau, \dots$. Let S_n be the set of all permutations of length n $(n \geqslant 1)$. In relation to the previous definition, any permutation σ contains at most $\lfloor \frac{n-1}{2} \rfloor$ valleys. A permutation $\sigma \in S_n$ is *alternating* if $\sigma_1 > \sigma_2 < \sigma_3 > \sigma_4 < \sigma_5 > \cdots$. In the literature [1], alternating permutations are also called *down-up* permutations and are enumerated by the Euler numbers (A000111 [19]). For instance, the permutation 324165 is alternating. Notice that an alternating permutation of length n contains exactly $\lfloor \frac{n-1}{2} \rfloor$ valleys.

A permutation π of length k, $k \leqslant n$, is a *pattern* of a permutation $\sigma \in S_n$ if there is a subsequence of σ which is order-isomorphic to π; *i.e.*, if there is a subsequence $\sigma_{i_1} \sigma_{i_2} \dots \sigma_{i_k}$ of σ (with $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$) such that $\sigma_{i_\ell} < \sigma_{i_m}$ whenever $\pi_\ell < \pi_m$. We write $\pi \prec \sigma$ to denote that π is a pattern of σ. A permutation σ that does not contain π as a pattern is said to *avoid* π. For example, $\sigma = 1423$ contains the patterns 132, 312 and 123; but σ avoids the pattern 321. The class of all permutations avoiding the patterns $\pi_1, \pi_2, \dots, \pi_k$ is denoted $S(\pi_1, \pi_2, \dots, \pi_k)$, and $S_n(\pi_1, \pi_2, \dots, \pi_k)$ denotes the set of permutations of length n avoiding $\pi_1, \pi_2, \dots$ and π_k. We also say that $S(\pi_1, \pi_2, \dots, \pi_k)$ is a class of pattern-avoiding permutations of basis $\{\pi_1, \pi_2, \dots, \pi_k\}$. A class $\mathcal{C}$ of permutations is *stable* for $\prec$ if, for any $\sigma \in \mathcal{C}$, for any $\pi \prec \sigma$, then we also have $\pi \in \mathcal{C}$. We now formulate a remark that is crucial for the present study.

Remark 1. If a class $\mathcal{C}$ of permutations is stable for $\prec$ then $\mathcal{C}$ is also a class of pattern avoiding permutations of basis $\mathcal{B} = \{\sigma \notin \mathcal{C}, \forall \pi \prec \sigma \text{ with } \pi \neq \sigma, \pi \in \mathcal{C}\}$.

The paper is organized as follow. In Section 2, we prove that the class of permutations obtained from the identity after a given number p of whole mirror duplications is the class of permutations with at most $2^{p-1} - 1$ valleys. This is the class of permutations avoiding the alternating permutations of length $2^p + 1$. Moreover, we obtain the length of a shortest path between any permutation and the identity. In Section 3, we yield two algorithms (and their

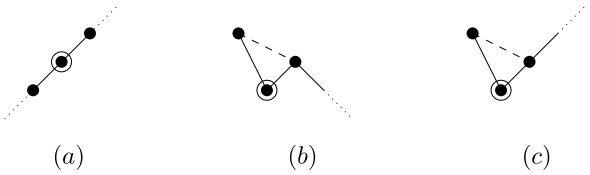


Fig. 2. The three non-isomorphic configurations in the proof of Lemma 1.

complexity) which construct such a shortest path. One of them uses an efficient algorithm for generating the well-known binary reflected Gray code [3,10]. In Section 4, we give some results about other models of duplications using W- and WM-duplications.

2. Pattern avoiding permutations and the mirror duplication-random loss model

In this section we study the WM-duplication random loss model in terms of pattern avoiding permutations. We establish that the class $\mathcal{C}(p)$ obtained from the identity after p WM-duplications is exactly the class of permutations avoiding all alternating permutations of length 2^p+1.

Lemma 1. *Let σ and π be two different permutations such that $\pi \prec \sigma$. Then σ contains at least many valleys as π does.*

Proof. It is sufficient to check the result for $\sigma \in S_n$ and $\pi \in S_{n-1}$ since a straightforward induction will complete the proof. Let $\sigma \in S_n$ and $\pi \in S_{n-1}$ such that $\pi \prec \sigma$, then π is order-isomorphic to a subsequence of σ obtained from σ by deleting only one entry of σ. We distinguish three non-isomorphic configurations illustrated in Fig. 2. More precisely, if we delete the encircled value in configuration (b), this reduces the number of valleys of σ. This does not occur for (a) and (c) where the number of valleys remains invariant after deletion of the encircled value. In all cases, a deletion of value in σ can not increase the number of valleys. □

Lemma 2. *A permutation obtained from the identity after a given number p of WM-duplications contains at most $2^{p-1}-1$ valleys.*

Proof. We obtain the proof by induction. The result holds for $p=1$; indeed a mirror duplication of the identity cannot create a valley. Now, let us assume that each permutation π obtained from the identity after $(p-1)$ mirror duplications contains at most $2^{p-2}-1$ valleys. Let σ be a permutation obtained from the identity after p mirror duplications. Then σ is obtained from a permutation π with $2^{p-2}-1$ valleys after exactly one mirror duplication. Therefore σ can be written as the concatenation of two subsequences of π and $\bar{\pi}$: *i.e.* $\sigma=\tau\tau'$ where τ (resp. τ') is a subsequence of π (resp. $\bar{\pi}$). According to Lemma 1, τ and τ' contains at most $2^{p-2}-1$ valleys. As the concatenation of τ and τ' can (eventually) create one valley between them, σ contains at most $2\cdot(2^{p-2}-1)+1=2^{p-1}-1$ valleys which achieves the induction. □

Theorem 1. *The class $\mathcal{C}(p)$ of permutations obtained from the identity after a given number p of mirror duplications is the class of permutations with at most $2^{p-1}-1$ valleys.*

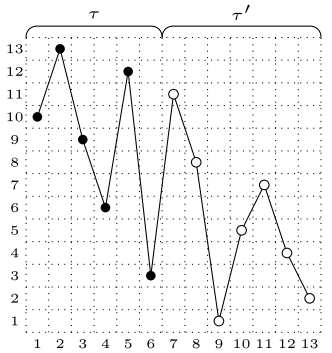
Proof. After considering Lemma 2, it suffices to prove that any permutation σ with at most $2^{p-1}-1$ valleys can be obtained from the identity after p mirror duplications. We proceed by induction on p. Indeed, let σ be a permutation with k valleys, $2^{p-2}-1<k\leqslant 2^{p-1}-1$. Then σ can be written $\sigma=\tau\tau'$ where τ corresponds to the longest prefix of σ containing exactly $2^{p-2}-1$ valleys and τ' the remaining suffix. We decompose the permutation $\tau=u_1d_1u_2d_2\dots u_\ell d_\ell$, where u_i and d_i are respectively runs up and down defined as follows: u_1 is the first run up excepted its top value; d_1 the run down just after u_1; u_2 the run up just after d_1 excepted its top value, and so on. Notice that u_1 can be empty which does not occur for d_ℓ. Remark that we have $\ell=2^{p-2}$ and thus $k<2\ell$. For example, $\tau=5421673$ has the decomposition: u_1 is empty, $d_1=5421$, $u_2=6$ and $d_2=73$. Let also $u_{\ell+1}d_{\ell+1}\dots u_kd_k\dots u_{2\ell}d_{2\ell}$ be the similar decomposition for τ'. In this decomposition, the runs u_i and d_i are empty for $i>k$. Now let us perform the following process: we sort in decreasing order the values that appear in d_ℓ or $u_{\ell+1}$ which creates a run down D_ℓ; we sort in increasing order the values in u_ℓ or $d_{\ell+1}$ which creates a run up U_ℓ; we construct the sequence $S=U_\ell D_\ell$. We sort in a decreasing order sequence $D_{\ell-1}$ the values in $d_{\ell-1}$ or $u_{\ell+2}$, and so on. At each step j, we insert the obtained ordered sequence U_j and D_j at the beginning of S. The permutation $S=U_1D_1\dots U_{\ell-1}D_{\ell-1}U_\ell D_\ell$ obtained at the end of this process contains at most $2^{p-2}-1$ valleys. See Fig. 3 for an example of construction of S. Thus, by induction, S can be obtained from the identity after $(p-1)$ mirror duplications. By construction σ is reached by one mirror duplication of S. Indeed, from any U_i (resp. D_i), $1\leqslant i\leqslant\ell$, we can reconstitute the corresponding u_i and $d_{2\ell-i+1}$ (resp. d_i and $u_{2\ell-i+1}$). Therefore σ can be constructed from the identity with p mirror duplications.

Notice that the permutation σ is decomposed in a partition into runs up and down u_j and d_j. We will use this decomposition in Section 3.2 to reconstitute a path of WM-duplications from the identity to σ. □

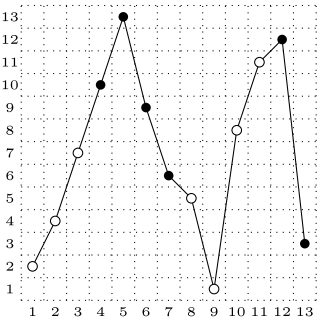
Corollary 1. *Let σ be a permutation and* $\mathrm{val}(\sigma)$ *the number of its valleys. In the mirror duplication model, $\lceil\log_2(\mathrm{val}(\sigma)+1)\rceil+1$ steps are necessary and sufficient in order to obtain σ from the identity permutation.*

Proof. Let p be the integer such that $2^{p-2}-1<\mathrm{val}(\sigma)\leqslant 2^{p-1}-1$, *i.e.* $p=\lceil\log_2(\mathrm{val}(\sigma)+1)\rceil+1$. According to Theorem 1, $p=\lceil\log_2(\mathrm{val}(\sigma)+1)\rceil+1$ steps are sufficient to obtain σ from the identity. It is also necessary since σ contains at least 2^{p-2} valleys which means that σ cannot be obtained from the identity in $(p-1)$ steps at most. □

In the next part we will provide algorithms in order to reconstitute a shortest path between the identity and a given permutation.



$\sigma = 10\ 13\ 9\ 6\ 12\ 3\ 11\ 8\ 1\ 5\ 7\ 4\ 2$



$S = 2\ 4\ 7\ 10\ 13\ 9\ 6\ 5\ 1\ 8\ 11\ 12\ 3$

Fig. 3. Decomposition $\sigma = \tau\tau'$ and the permutation S obtained after the process in the proof of Theorem 1. We have $\tau = 10\ 13\ 9\ 6\ 12\ 3$, $\tau' = 11\ 8\ 1\ 5\ 7\ 4\ 2$, $u_1 = 10$, $d_1 = 13\ 9\ 6$, u_2 is empty, $d_2 = 12\ 3$, u_3 is empty, $d_3 = 11\ 8\ 1$, $u_4 = 5$, $d_4 = 7\ 4\ 2$ and, $D_2 = 12\ 3$, $U_2 = 1\ 8\ 11$; $D_1 = 13\ 9\ 6\ 5$; $U_1 = 2\ 4\ 7\ 10$. Thus $S = U_1D_1U_2D_2 = 2\ 4\ 7\ 10\ 13\ 9\ 6\ 5\ 1\ 8\ 11\ 12\ 3$.

Theorem 2. *The class $\mathcal{C}(p)$ of permutations obtained after a given number p of WM-duplications is the class of permutations avoiding the alternating permutations of length $2^p + 1$.*

Proof. Indeed, the permutations obtained after p mirror duplications is stable for the relation $\prec$, *i.e.* if $\sigma \in \mathcal{C}(p)$ and $\pi \prec \sigma$ then $\pi \in \mathcal{C}(p)$ (see Lemma 1). Thus (see Remark 1), $\mathcal{C}(p)$ is also a class of pattern avoiding permutations $S(B)$ where B is the set of minimal (relatively to $\prec$) permutations σ that are not in $\mathcal{C}(p)$. Such a minimal permutation σ contains exactly 2^{p-1} valleys. Indeed, $\sigma \notin \mathcal{C}(p)$ means $\mathrm{val}(\sigma) \geqslant 2^{p-1}$ and any permutation with at least $2^{p-1} + 1$ valleys is not minimal since it contains a pattern $\pi \notin \mathcal{C}(p)$ with 2^{p-1} valleys. Moreover, the configuration $\sigma_{i-1} < \sigma_i < \sigma_{i+1}$ and $\sigma_{i-1} > \sigma_i > \sigma_{i+1}$ can not occur since if we delete σ_i we do not decrease the number of valleys. Also we necessarily have $\sigma_1 > \sigma_2$. Thus we deduce that σ is an alternating permutation with 2^{p-1} valleys and its length is $2^p + 1$. □

For example, $\mathcal{C}(1) = S(213, 312)$ and $\mathcal{C}(2) = S(21435, 31425, 41325, 32415, 42315, 21534, 31524, 51324, 32514, 52314, 41523, 51423, 42513, 52413, 43512, 53412)$.

We also obtain a more general result (see Theorem 3) for the class of permutations having at most p valleys. Notice that, with [16,12,13], a bivariate generating function for this class is:

$$\frac{1}{1-y}\left(1 - \frac{1}{y} + \frac{1}{y}\sqrt{y-1} \times \tan\left(x\sqrt{y-1} + \arctan\left(\frac{1}{\sqrt{y-1}}\right)\right)\right),$$

where the coefficient of $x^n y^k$ is the number of permutations of length n with at most k valleys.

Theorem 3. *The class of permutations having at most p valleys is the class of permutations avoiding the alternating permutations of length $2p + 3$.*

Proof. The proof is obtained *mutatis mutandis*. □

3. Algorithmic considerations

In this section, we present two algorithms that provide a possible scenario of whole mirror duplications and losses for any $\sigma \in S_n$. Both construct a shortest path of WM-duplications from identity to σ. The first algorithm reconstitutes the path in reverse order, *i.e.*, we start from σ and at each step we compute the predecessor of the current permutation until the identity. The second algorithm computes the successor of the current permutation by starting from the identity until σ. These two algorithms give (generally) two different paths. The second method is interesting in the sense where it uses the special structure of the reflected binary Gray code [10]. Moreover, in the case where we want to get the first (resp. last) elements of the path, Algorithm 2 (resp. Algorithm 1) will be more efficient. Finally, we discuss the complexity of these algorithms.

Notice that if we have a shortest path $\mathcal{P}$ of WM-duplications from identity to σ, we can easily obtain a shortest path of WM-duplications from σ^{-1} to the identity by applying σ^{-1} on each permutation of $\mathcal{P}$.

3.1. A path from $12\ldots n$ to $\sigma \in S_n$: Algorithm 1

Here we explain how we can establish an algorithm in order to construct a scenario of WM-duplications from identity to $\sigma \in S_n$ by reconstructing this path from σ, *i.e.* at each step we find the predecessor of the current permutation and the process finishes when the identity is reached. We partition the permutation σ as follow: $\sigma = u'_1 d'_1 u'_2 d'_2 \ldots u'_k d'_k$ where $u'_1 = \sigma_1 \ldots \sigma_{i_1}$ is the maximal run up containing the first entry σ_1; d'_k is the maximal run down containing σ_n; d'_1 is the maximal run down containing σ_{i_1+1}; u'_k is the maximal run up just before d'_k, *i.e.* the value just before d'_k in σ belongs to u'_k; we continue this process by alternating runs up and down until the permutation σ is entirely partitioned. Notice that this decomposition is not the same as we have done in the proof of Theorem 1. For i from 1 to $\ell = \lfloor \frac{k+1}{2} \rfloor$, we define by U_i (resp. D_i) the increasing (resp. decreasing) order sorting of u'_i and d'_{k-i+1} (resp. d'_i and u'_{k-i+1}). Let $\pi = U_1 D_1 U_2 D_2 \ldots U_\ell D_\ell$ be the concatenation of all sorted

Algorithm 1 The procedure BackPath producing a scenario of WM-duplications from $12\dots n$ to $\sigma \in S_n$.

procedure BackPath
while $\sigma \neq 12\dots n$ **do**
- $\pi \leftarrow$ *empty*
- Let $\sigma = u'_1 d'_1 u'_2 d'_2 \dots u'_k d'_k$ be the partition into runs up and runs down

for $i \leftarrow 1$ **to** $\lfloor \frac{k+1}{2} \rfloor$ **do**
- Sort in increasing order u'_i and d'_{k-i+1} and append the sorted substring on the right of π
- Sort in decreasing order d'_i and u'_{k-i+1} and append the sorted substring on the right of π

end for
- $\sigma \leftarrow \pi$

end while

substrings U_i and D_i where D_ℓ can be empty. Therefore π contains at most $(\ell - 1) = \lfloor \frac{k+1}{2} \rfloor - 1 = \lfloor \frac{\mathrm{val}(\sigma)+2}{2} \rfloor - 1 = \lfloor \frac{\mathrm{val}(\sigma)}{2} \rfloor$ valleys. This step requires $\mathcal{O}(n)$ computations since we can simultaneously detect and sort the runs up and down. By iterating this process with this new permutation π, Corollary 1 ensures that the procedure BackPath (see Algorithm 1) constructs a path from identity to σ in $(\lceil \log_2(\mathrm{val}(\sigma)+1) \rceil + 1)\mathcal{O}(n)$, *i.e.* $\mathcal{O}(n \cdot \log_2(n))$ in the worst case. For instance, this process applied to the permutation $\sigma = \mathbf{5}421\mathbf{6}73$ gives the path $\sigma \leftarrow \mathbf{357}6421 \leftarrow 1234567$ (the runs up are in boldface).

3.2. A path from $12\dots n$ to $\sigma \in S_n$: Algorithm 2

In this section, we construct a path of WM-duplications from $12\dots n$ to the permutation $\sigma \in S_n$, *i.e.* we start from the identity and at each step we find the successor of the current permutation; the process finishes when σ is reached. We decompose the permutation $\sigma = u_1 d_1 u_2 d_2 \dots u_k d_k \dots u_{2\ell} d_{2\ell}$ in runs up and down in the same way as we have done in the proof of Theorem 1. We recall this decomposition: u_i and d_i are respectively runs up and down defined as follows: u_1 is the first run up excepted its top value; d_1 the run down just after u_1; u_2 the run up just after d_1 excepted its top value, and so on. Notice that u_1 can be empty: such a decomposition is given as example in the proof of Theorem 1.

We now label each run in this decomposition with the reflected binary Gray code [10]: we can do this using a loopless algorithm introduced by Bitner, Ehrlich, and Reingold [3]. For instance, Fig. 4 shows such a labeling. The structure of the binary reflected Gray code B_n is crucial: $B_n = 0 \cdot B_{n-1} \cup 1 \cdot \overline{B_{n-1}}$ anchored by $B_1 = \{0, 1\}$ and where $\overline{B_n}$ is the list B_n considered in the reverse order. The first runs from u_1 to d_ℓ have 0 as their least significant bit, then those have 1. According to the proof of Theorem 1, we reconstitute the previous permutation π of σ by concatenating (on the left) the sorting in decreasing order D_ℓ of d_ℓ and $u_{\ell+1}$, the sorting in increasing order U_ℓ of u_ℓ and $d_{\ell+1}$, and so on.

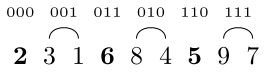
Conversely, at the step j of our algorithm we perform a WM-duplication of π by: (i) keeping in the first copy of π the elements labeled 0 on the jth least significant bit, and, (ii) keeping in the mirror the elements labeled 1 on the jth least significant bit (see Fig. 4 for an example). This step requires only $\mathcal{O}(n)$ computations.

Algorithm 2 The procedure Path producing a scenario of WM-duplications from $12\dots n$ to $\sigma \in S_n$.

procedure Path
- $\pi = 12\dots n$
- Partition $\sigma = u_1 d_1 u_2 d_2 \dots u_k d_k$ into runs up and runs down
- Label the runs up and runs down with the reflected binary Gray code (loopless algorithm [3])

for $j = 1$ **to** $1 + \lfloor \log_2(k-1) \rfloor$ **do**
- Make a WM-duplication step on π that keeps in the first copy of π exactly the elements whose label has 0 in its jth least significant bit.

end for



$$123456789 \to 245897631 \to 231679854 \to 231684597$$

Fig. 4. Labeling of $\sigma = 231684597$ with the binary reflected Gray code according to the runs up and down. Example of a path of WM-duplications from the identity permutation to σ.

By iterating this process, Corollary 1 ensures that the procedure Path (see Algorithm 2) constructs a path from the identity to σ in $(\lceil \log_2(\mathrm{val}(\sigma) + 1) \rceil + 1)\mathcal{O}(n)$, *i.e.* $\mathcal{O}(n \cdot \log_2(n))$ in the worst case. Remark that Algorithm 2 is more efficient than Algorithm 1 in the case where we want to get the first permutations of the path. An implementation of our algorithm is available on http://www.u-bourgogne.fr/jl.baril/id2perm.php.

4. Other models of duplication

In this section, we investigate two variants of the whole mirror duplication: 1) we do one WM-duplication of the identity followed by several W-duplications and, 2) we do one WM- or/and W-duplication of the identity followed by several W-duplications. For these two cases we provide a characterization of the class of permutations obtained after p duplications.

4.1. One WM-duplication followed by several W-duplications

We have seen in the first section that one WM-duplication of the identity gives the class of permutations without valley. Thus, we obtain by induction that the class of permutations obtained after one WM-duplication of the identity followed by $(p-1)$ W-duplications is included in the class of permutations with at most $2^{p-1} - 1$ valleys.

Theorem 4. *The class of permutations obtained after one WM-duplication of the identity followed by $(p-1)$ W-duplications is the class of permutations with at most $2^{p-1} - 1$ valleys.*

Proof. Theorem 1 induces that the class of permutations obtained after one WM-duplication of the identity is the class of permutations without valley. Thus the result is true for $p = 1$. Now we proceed by induction. Let us assume that the class of permutations obtained after one WM-duplication of the identity followed by $(p-1)$ W-duplications is the class of permutations with at most $2^{p-1} - 1$ valleys. Let σ be a permutation in this last class. If we apply one

W-duplication on σ, we obtain a permutation π with at most $2(2^{p-1}-1)+1 = 2^p - 1$ valleys. Now it suffices to prove that any permutation π with at most $2^p - 1$ valleys can be obtained after one W-duplication of a permutation σ with at most $2^{p-1} - 1$ valleys. Indeed, let π be a permutation with k valleys, $2^{p-1} - 1 < k \leqslant 2^p - 1$. As we have done in the proof of Theorem 2, we write $\pi = \tau\tau'$ where τ corresponds to the longest prefix of π containing exactly $2^{p-1} - 1$ valleys and τ' the remaining suffix. We decompose the permutation $\tau = u_1 d_1 u_2 d_2 \dots u_\ell d_\ell$, where u_i and d_i are respectively runs up and down defined as follows: u_1 is the first run up excepted its top value; d_1 is the run down just after u_1; u_2 is the run up just after d_1 excepted its top value, and so on. We necessarily have $\ell = 2^{p-1}$ and thus $k < 2\ell$. Let also $u_{\ell+1} d_{\ell+1} \dots u_k d_k \dots u_{2\ell} d_{2\ell}$ be the similar decomposition for τ'. Now let us perform the following process: we sort in increasing order the values that appear in u_1 or $u_{\ell+1}$ which creates a run up U_1; we sort in decreasing order the values in d_1 or $d_{\ell+1}$ which creates a run down D_1; we construct the sequence $S = U_1 D_1$. We sort in an increasing order sequence U_2 the values in u_2 or $u_{\ell+2}$, and so on. At each step j, we insert the obtained ordered sequences U_j and D_j on the right of S. The permutation $S = U_1 D_1 \dots U_{\ell-1} D_{\ell-1} U_\ell D_\ell$ obtained at the end of this process contains at most $2^{p-1} - 1$ valleys. By construction π is reached by one W-duplication of S since we can reconstitute u_i and $u_{\ell+i}$ from U_i (resp. d_i and $d_{\ell+i}$ from D_i). We conclude by induction. □

The following corollary is directly deduced from Theorem 4 with the same proof as for Theorem 2.

Corollary 2. *The class of permutations obtained after one WM-duplication of the identity followed by $(p-1)$ W-duplications is the class of permutations avoiding the alternating permutations of length $2^p + 1$.*

4.2. One W- or WM-duplication followed by several W-duplications

In this part, we perform one W- or WM-duplication of the identity followed by several W-duplications. Let us recall that an ascent (resp. a descent) of an n-length permutation σ is a position i, $1 \leqslant i \leqslant n-1$, such that $\sigma(i) < \sigma(i+1)$ (resp. $\sigma(i) > \sigma(i+1)$).

Theorem 5. *The class of permutations obtained after one W- or WM-duplication of the identity followed by p W-duplications is the union of the two classes*: (i) *permutations having at most $2^p - 1$ valleys and,* (ii) *permutations with at most $2^{p+1} - 1$ descents.*

Proof. After one W- or WM-duplication of the identity we obtain the union of the class of permutations without valley with the class of permutations with at most one descent. A straightforward induction on p allows us to obtain the result using Theorems 1 and 4. □

Corollary 3. *The class $\mathcal{C}'(p)$ of permutations obtained after one W- or WM-duplication of the identity followed by p W-duplications is the class of permutations avoiding the permutations of length $3 \cdot 2^p + 1$ having exactly 2^p valleys and 2^{p+1} descents.*

Proof. Let σ be a minimal permutation which is not in $\mathcal{C}'(p)$. The minimality of σ induces that σ does not contain any run up of size at least three. By contradiction, if it is the case then by deleting the second value of the run up, the obtained sequence is isomorphic to a permutation which is also not in $\mathcal{C}'(p)$. Also σ cannot begin by an ascent for the same reasons. The permutation σ necessarily has 2^p valleys and 2^{p+1} descents. Indeed, if σ contains at least $2^{p+1} + 1$ descents, it contains a run down of length at least three which contradicts the fact that σ is minimal. A simple calculation allows us to show that σ is necessarily of length $3 \cdot 2^p + 1$. □

For example, $\mathcal{C}'(0) = S(4132, 3142, 4312, 3241, 3214, 4231, 4213, 2143)$; we have computed the cardinality of the basis of $\mathcal{C}'(1)$, which is 720. In the same way, we have the more general corollary:

Corollary 4. *The class of permutations having at most p valleys and at most $2p+1$ descents is the class of permutations avoiding the permutations of length $3p+4$ having exactly $p+1$ valleys and $2p+2$ descents.*

Notice that the set of permutations of length $3p+4$ having exactly $p+1$ valleys and $2p+2$ descents is also the set of permutations with the same length, the same number of valleys and where every ascent is immediately preceded by a descent (which defines a valley). This set was studied by Shapiro et al. [18] (see also the sequence A101280 in [19]). Indeed, they enumerate the permutations of length n having k peaks and with the additional property that every ascent is immediately followed by a descent. Thus, the cardinality of our set is obtained when $n = 3p+4$ and $k = p+1$. For instance, the first cardinalities for $p = 0, 1, 2, 3, 4$ are $8, 720, 230144, 179266560, 277662253056$.

5. Further research directions

In this paper, the whole mirror duplication model is studied. However, many relative open problems still need to be considered. For example, let σ_1 and σ_2 be two permutations in S_n. Can one exhibit an efficient algorithm to compute the permutation π which minimizes the sum $d(\pi, \sigma_1) + d(\pi, \sigma_2)$ where $d(\pi, \sigma)$ is the minimum of WM-duplication steps required to transform π into σ? Can one characterize the class of avoiding permutations corresponding to the permutations obtained from the identity after one W-duplication followed by p WM-duplications? More generally, can we characterize the permutations obtained after p steps of either W- or WM-duplications? Can one obtain similar results with the tandem mirror duplication model: the mirror of the duplicated part (of size K) is inserted immediately after the original portion and we apply the loss procedure.

Acknowledgements

We would like to thank the anonymous referee and Vincent Vajnovszki for their constructive remarks which have greatly improved this paper.

References

[1] B. Bauslaugh, F. Ruskey, Generating alternating permutations lexicographically, BIT Numerical Mathematics 30 (1) (1990) 17–26.
[2] S. Bérard, A. Bergeron, C. Chauve, C. Paul, Perfect sorting by reversals is not always difficult, IEEE ACM Transactions on Computational Biology and Bioinformatics 4 (1) (2007) 4–16.
[3] J.R. Bitner, G. Ehrlich, E.M. Reingold, Efficient generation of the binary reflected Gray code and its applications, Communications of the ACM 19 (9) (1976) 517–521.
[4] M. Bouvel, E. Pergola, Posets and permutations in the duplication-loss model, Pure Math. Appl. 19 (2–3) (2008) 71–80.
[5] M. Bouvel, D. Rossin, A variant of the tandem duplication-random loss model of genome rearrangement, Theoretical Computer Science 410 (8–10) (2009) 847–858.
[6] E.R. Canfield, H.S. Wilf, Counting permutations by their alternating runs, Journal of Combinatorial Theory 115 (2008) 213–225.
[7] K. Chaudhuri, K. Chen, R. Mihaescu, S. Rao, On the tandem duplication-random loss model of genome rearrangement, in: Proceedings of the Seventeenth Annual ACM–SIAM Symposium on Discrete Algorithm, ACM, New York, NY, USA, 2006, pp. 564–570.
[8] H.D. Chen, W.L. Fan, S.G. Kong, H. Lee, B. Zheng, N. Zhou, Inverse symmetry in genomes and whole-genome inverse duplication, in: International Bioinformatics Workshop (IBW2008), Yunnan University, Kunming, Yunnan, China, 2008.
[9] M.C. Chen, R.C.T. Lee, Sorting by transpositions based on the first increasing substring concept, in: BIBE'04, Proceedings of the 4th IEEE Symposium on Bioinformatics and Bioengineering, IEEE Computer Society, Washington, DC, USA, 2004, p. 553.
[10] F. Gray, Pulse code communication, U.S. Patent, 2,632,058, 1953.
[11] M.T. Hallett, J. Lagergren, New algorithms for the duplication-loss model, in: Proceedings of the Fourth Annual International Conference on Computational Molecular Biology, ACM, New York, NY, USA, 2000, pp. 138–146.
[12] S. Kitaev, Introduction to partially ordered patterns, Discrete Applied Mathematics 155 (2007) 929–944.
[13] S. Kitaev, A. Pyatkin, On avoidance of V- and Λ-patterns in permutations, Ars Combinatoria (2010), in press.
[14] A. Labarre, New bounds and tractable instances for the transposition distance, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 3 (4) (2006) 380–394.
[15] R. Nussinov, Some indications for inverse DNA duplication, Journal of Theoretical Biology 95 (4) (1982) 783.
[16] R. Rieper, M. Zeleke, Valleyless sequences, ArXiv:math/0005180, 2000.
[17] D. Sankoff, Gene and genome duplication, Current Opinion in Genetics and Development 11 (2001) 681–684.
[18] L.W. Shapiro, W.-J. Woan, S. Getu, Runs, slides and moments, SIAM J. Algebraic and Discrete Methods 4 (461) (1983).
[19] N.J.A. Sloane, The on-line encyclopedia of integer sequences, http://www.research.att.com/~njas/sequences/.