

The Hierarchy of Hereditary Sorting Operators

Vít Jelínek^{*1}, Michal Opler^{†2}, and Jakub Pekárek¹

¹Charles University, Czechia

²Czech Technical University in Prague, Czechia

Abstract

We consider the following general model of a sorting procedure: we fix a hereditary permutation class \mathcal{C} , which corresponds to the operations that the procedure is allowed to perform in a single step. The input of sorting is a permutation π of the set $[n] = \{1, 2, \dots, n\}$, i.e., a sequence where each element of $[n]$ appears once. In every step, the sorting procedure picks a permutation σ of length n from \mathcal{C} , and rearranges the current permutation of numbers by composing it with σ . The goal is to transform the input π into the sorted sequence $1, 2, \dots, n$ in as few steps as possible.

Formally, for a hereditary permutation class \mathcal{C} and a permutation π of $[n]$, we say that \mathcal{C} *can sort π in k steps*, if the inverse of π can be obtained by composing k (not necessarily distinct) permutations from \mathcal{C} . The *\mathcal{C} -sorting time* of π , denoted $\text{st}(\mathcal{C}; \pi)$, is the smallest k such that \mathcal{C} can sort π in k steps; if no such k exists, we put $\text{st}(\mathcal{C}; \pi) = +\infty$. For an integer n , the *worst-case \mathcal{C} -sorting time*, denoted $\text{wst}(\mathcal{C}; n)$, is the maximum of $\text{st}(\mathcal{C}; \pi)$ over all permutations π of $[n]$.

This model of sorting captures not only classical sorting algorithms, like insertion sort or bubble sort, but also sorting by series of devices, like stacks or parallel queues, as well as sorting by block operations commonly considered, e.g., in the context of genome rearrangement.

Our goal is to describe the possible asymptotic behavior of the function $\text{wst}(\mathcal{C}; n)$, and relate it to structural properties of \mathcal{C} . As the main result, we show that any hereditary permutation class \mathcal{C} falls into one of the following five categories:

- $\text{wst}(\mathcal{C}; n) = +\infty$ for n large enough,
- $\text{wst}(\mathcal{C}; n) = \Theta(n^2)$,
- $\Omega(\sqrt{n}) \leq \text{wst}(\mathcal{C}; n) \leq O(n)$,
- $\Omega(\log n) \leq \text{wst}(\mathcal{C}; n) \leq O(\log^2 n)$, or
- $\text{wst}(\mathcal{C}; n) = 1$ for all $n \geq 2$.

In addition, we characterize the permutation classes in each of the five categories.

^{*}Supported by the Czech Science Foundation under the grant agreement no. 23-04949X.

[†]Supported by the Czech Science Foundation Grant no. 22-19557S.

1 Introduction

Sorting is, needless to stress, a central concept in computer science. The aim of this paper is to study a general way to formalize the notion of sorting, based upon the concept of hereditary permutation classes. In this approach, we fix a hereditary class \mathcal{C} of permutations, which corresponds to the operations that we are allowed to perform in a single step of sorting. Given an input sequence of n distinct values, itself viewed as a permutation, we aim to sort it into an increasing sequence by a sequence of steps, where in each step we compose the current sequence with an element of \mathcal{C} . The goal is to minimize the number of steps for a given input, and to determine the worst-case number of steps \mathcal{C} needs on inputs of size n . This last quantity, which we call the *worst-case sorting time* of the class \mathcal{C} , is the main focus of this paper.

Our formalism is general enough to encompass standard sorting procedures, like sorting by adjacent transpositions (“bubble-sort”) or sorting by insertions (“insertion sort”), as well as parallel procedures like Habermann’s “parallel neighbor sort” [21] (a.k.a. “odd-even sort”), which allows performing any number of adjacent swaps in a single step.

Our formalism also models sorting procedures which are described via operations not on single elements, but on blocks of consecutive elements, e.g. block reversals or block transpositions. Such block-based sorting occurs, e.g., in the well-known toy problem known as “pancake sorting” [11, 19], where the goal is to sort the input by a sequence of prefix reversals. Block operations are also the backbone of the sorting procedures studied in the context of genome rearrangement research (see, e.g., [7, 14, 22, 24], or the surveys in [9, 18]).

Finally, we point out that our sorting formalism is closely related to the concept of sorting networks, an approach which dates back to the works of Knuth [25] and Tarjan [27] in late 1960s and early 1970s, respectively. A sorting network is an acyclic directed graph whose every vertex represents a “sorting device” (e.g., a stack, a queue, a double-ended queue, or a more general “ \mathcal{C} -container” considered by Albert et al. [3]). An input sequence enters into the network from a designated source, its elements can move along the edges and be rearranged by the devices at the vertices, with the goal that the elements reach a designated sink in ascending order. Since the action of most sorting devices considered in the literature can be modelled by a hereditary permutation classes, our sorting formalism can be seen as a serial sorting network (i.e., the network is a path) in which all the devices are of the same type.

The topic of sorting networks continues to be a major focus of research [10, 12, 13, 17], but we note that most of its previous results address the question of how many permutations can be sorted by a given fixed network and what these permutations look like, while our own focus is on finding the shortest serial network that sorts every permutation of a given length. Additionally, some authors consider sorting networks where the operation of the devices is restricted by the properties of the input sequence, e.g., by assuming that a stack may only hold a monotone sequence of values. These restricted devices no longer fit into our own formalism.

The aim of our paper is neither to improve the analysis of any specific sorting algorithm, nor to propose a new specific algorithm; instead, our goal is to answer the general question of which functions can occur as worst-case sorting times of hereditary

permutation classes. Accordingly, we disregard constant multiplicative factors and focus on asymptotic behavior. Our results show that at this level of detail, the possible sorting times can be classified into a hierarchy with gaps between individual levels. Disregarding trivial extreme cases, a worst-case sorting time of a hereditary class is either $\Theta(n^2)$, or a function between $O(n)$ and $\Omega(\sqrt{n})$, or $O(\log^2 n)$, and for each of these cases we can provide a structural characterization of the corresponding hereditary classes.

We remark that the assumption of heredity is substantial in our setting. If we were to consider sorting with an arbitrary set of generators, the sorting-time problem would be essentially reduced to the problem of determining the diameter of a Cayley graph in the symmetric group. This is a well-known group-theoretic open problem, with a conjecture of Babai and Seress [6] stating that the diameter is at most polynomial.

Let us now introduce the necessary definitions to state our results properly.

Notation, terminology, and the main result

A *permutation* of size n is a sequence $\pi = (\pi(1), \pi(2), \dots, \pi(n))$ in which each number from the set $[n] = \{1, 2, \dots, n\}$ appears exactly once. We let \mathcal{S}_n be the set of permutations of size n , $\mathcal{S}_{\leq n}$ the set of permutations of size at most n , and \mathcal{S} the set $\bigcup_{n \geq 1} \mathcal{S}_n$. We say that a permutation $\pi \in \mathcal{S}_n$ *contains* a permutation $\sigma \in \mathcal{S}_k$ if π contains a subsequence of length k whose elements are in the same relative order as the elements of σ , or more formally, if there are indices $1 \leq i_1 < i_2 < \dots < i_k \leq n$ such that $\pi(i_a) < \pi(i_b) \iff \sigma(a) < \sigma(b)$ for every $a, b \in [k]$. If π does not contain σ , we say it *avoids* σ .

A *hereditary permutation class* (or henceforth just *class*) is a set $\mathcal{C} \subseteq \mathcal{S}$ such that for every $\pi \in \mathcal{C}$ and every σ contained in π , σ is also in \mathcal{C} . For a permutation class \mathcal{C} , we let \mathcal{C}_n denote the set $\mathcal{C} \cap \mathcal{S}_n$.

For a pair of permutations $\sigma, \pi \in \mathcal{S}_n$, their *composition* $\sigma \circ \pi$ is the permutation $(\sigma(\pi(1)), \sigma(\pi(2)), \dots, \sigma(\pi(n)))$. For a pair of permutation classes \mathcal{A} and \mathcal{B} , we write $\mathcal{A} \circ \mathcal{B}$ for $\{\alpha \circ \beta; \alpha \in \mathcal{A}, \beta \in \mathcal{B}\}$, $\mathcal{A}^{\circ k}$ for the k -fold composition $\mathcal{A} \circ \mathcal{A} \circ \dots \circ \mathcal{A}$, and \mathcal{A}^* for $\bigcup_{k \geq 1} \mathcal{A}^{\circ k}$. Note that $\mathcal{A} \circ \mathcal{B}$, $\mathcal{A}^{\circ k}$ and \mathcal{A}^* are again permutation classes.

The *inverse* of a permutation $\pi \in \mathcal{S}_n$ is the permutation $\pi^{-1} \in \mathcal{S}_n$ such that $\pi^{-1} \circ \pi$ is the identity permutation $(1, 2, \dots, n)$.

Recall that the *\mathcal{C} -sorting time* of π , denoted $\text{st}(\mathcal{C}; \pi)$, is the smallest k such that π^{-1} is in $\mathcal{C}^{\circ k}$, or $+\infty$ if no such k exists. For an integer n , the *worst-case \mathcal{C} -sorting time*, denoted $\text{wst}(\mathcal{C}; n)$, is the maximum of $\text{st}(\mathcal{C}; \pi)$ over all permutations π of $[n]$.

We are now ready to recall the main result of this paper.

Theorem 1.1. *Every permutation class \mathcal{C} falls into one of the following five categories:*

- $\text{wst}(\mathcal{C}; n) = +\infty$ for n large enough,
- $\text{wst}(\mathcal{C}; n) = \Theta(n^2)$,
- $\Omega(\sqrt{n}) \leq \text{wst}(\mathcal{C}; n) \leq O(n)$,

- $\Omega(\log n) \leq \text{wst}(\mathcal{C}; n) \leq O(\log^2 n)$, or
- $\text{wst}(\mathcal{C}; n) = 1$ for all $n \geq 2$.

The rest of this paper is devoted to the proof of this theorem. After we introduce some more terminology and review some relevant facts about permutation classes (Section 2), we will proceed by characterizing the classes with infinite sorting time in Section 3. This is easily done, as it more or less directly follows from previous results on composition-closed permutation classes. We then show, in Section 4, that if the sorting time is finite, it is at most $O(n^2)$. Next, in Section 5, we establish the gap between quadratic and linear sorting times. Our main device is a numerical parameter called *reduced inversion number* $\text{rin}(\pi)$, which we introduce. We show that a class \mathcal{C} has $\text{wst}(\mathcal{C}; n) = \Omega(n^2)$ whenever $\text{rin}(\cdot)$ is bounded on \mathcal{C} , and $\text{wst}(\mathcal{C}; n) = O(n)$ otherwise.

The most difficult part of the proof appears in Section 6, where we show that there are twelve (or four, after we account for symmetries) minimal classes with sorting time $O(\log^2 n)$, while any class that does not contain one of these twelve as a subclass has sorting time $\Omega(\sqrt{n})$. The proof of this lower bound combines a new structural result on classes avoiding the twelve ‘polylog’ subclasses, an argument showing that sorting with such a class can be encoded into a specific graph which has a drawing on a surface of not-too-large genus with not-too-many crossings, and a bound on the tree-width of a graph that can be drawn on a surface of a given genus with a given number of crossings.

The short Section 7 contains the proof that any class not containing all permutations has sorting time $\Omega(\log n)$. This is a straightforward counting argument using the Marcus–Tardos theorem. We conclude in Section 8 with open problems, chief of which is to close the gaps between $\Omega(\sqrt{n})$ and $O(n)$, and between $\Omega(\log n)$ and $O(\log^2 n)$ implied by Theorem 1.1. In fact, for all the classes whose sorting time we can establish asymptotically tightly, the sorting time is either $+\infty$, $\Theta(n^2)$, $\Theta(n)$, $\Theta(\log n)$ or 1.

2 Permutation classes

We let $\text{Av}(\pi^1, \pi^2, \dots, \pi^m)$ denote the class of permutations that avoid all the permutations π^1, \dots, π^m . We also use the special notation \Box for the class $\text{Av}(21)$ of all the increasing permutations, and \sqsupset for the class $\text{Av}(12)$ of all the decreasing permutations. Let ι_k denote the identity permutation $(1, 2, \dots, k)$, and let δ_k be the decreasing permutation $(k, k-1, \dots, 1)$.

It is often convenient to represent a permutation $\pi \in \mathcal{S}_n$ by its *diagram*, which is the point set $D(\pi) = \{(i, \pi(i); i \in [n])\}$.

Apart from the inverse π^{-1} introduced before, we will need several other operations on a permutation $\pi = (\pi(1), \dots, \pi(n))$. Its *reverse*, denoted π^r , is the permutation $(\pi(n), \pi(n-1), \dots, \pi(1))$, its *complement* π^c is the permutation $(n+1-\pi(1), n+1-\pi(2), \dots, n+1-\pi(n))$. Finally, the *flip* of π , denoted π^f , is defined as $\pi^f = ((\pi^r)^{-1})^r$.

In terms of permutation diagrams, the reverse corresponds to reflection over a vertical line, the complement corresponds to reflection over a horizontal line, the

inverse corresponds to reflection over the diagonal line $y = x$, and flip corresponds to reflection over the line $y = n + 1 - x$.

For a set of permutations \mathcal{C} , we write $\mathcal{C}^r = \{\pi^r; \pi \in \mathcal{C}\}$, $\mathcal{C}^c = \{\pi^c; \pi \in \mathcal{C}\}$, $\mathcal{C}^{-1} = \{\pi^{-1}; \pi \in \mathcal{C}\}$ and $\mathcal{C}^f = \{\pi^f; \pi \in \mathcal{C}\}$. Note that if \mathcal{C} is a class, then so are \mathcal{C}^r , \mathcal{C}^c , \mathcal{C}^{-1} and \mathcal{C}^f .

For a pair of permutations $\alpha \in \mathcal{S}_k$ and $\beta \in \mathcal{S}_\ell$, their *direct sum* $\alpha \oplus \beta$ is the permutation $(\alpha(1), \alpha(2), \dots, \alpha(k), k + \beta(1), k + \beta(2), \dots, k + \beta(\ell))$, and their *skew sum* $\alpha \ominus \beta$ is the permutation $(\ell + \alpha(1), \ell + \alpha(2), \dots, \ell + \alpha(k), \beta(1), \beta(2), \dots, \beta(\ell))$. We let $\bigoplus^m \alpha$ denote the direct sum $\alpha \oplus \alpha \oplus \dots \oplus \alpha$ with m summands, and $\bigominus^m \alpha$ is defined similarly. For a set of permutations \mathcal{C} , its *sum-closure* $\bigoplus \mathcal{C}$ is the set of all the permutations that can be obtained as a direct sum of a finite number of (not necessarily distinct) elements of \mathcal{C} . Similarly, we define the *skew-closure* $\bigominus \mathcal{C}$. If \mathcal{C} is a class, then so are $\bigoplus \mathcal{C}$ and $\bigominus \mathcal{C}$.

Before we proceed, let us collect several simple observations related to sorting times of various classes.

Observation 2.1. *For any two permutations σ and π of size n , we have $(\sigma \circ \pi)^{-1} = \pi^{-1} \circ \sigma^{-1}$ and $(\sigma \circ \pi)^f = \pi^f \circ \sigma^f$. Consequently, for any permutation class \mathcal{C} and any permutation τ , $\text{st}(\mathcal{C}; \tau) = \text{st}(\mathcal{C}^{-1}; \tau^{-1}) = \text{st}(\mathcal{C}^f; \tau^f)$, and in particular, $\text{wst}(\mathcal{C}; n) = \text{wst}(\mathcal{C}^{-1}; n) = \text{wst}(\mathcal{C}^f; n)$.*

Observation 2.2. *If \mathcal{A} and \mathcal{B} are permutation classes such that $\mathcal{A} \subseteq \mathcal{B}^{\circ k}$ and π is any permutation, then $\text{st}(\mathcal{B}; \pi) \leq k \cdot \text{st}(\mathcal{A}; \pi)$, and hence $\text{wst}(\mathcal{B}; n) \leq k \cdot \text{wst}(\mathcal{A}; n)$.*

Observation 2.3. *Let \mathcal{C} be a class containing \boxtimes as a subclass. Then $\mathcal{C}^r = \mathcal{C} \circ \boxtimes \subseteq \mathcal{C}^{\circ 2}$, and in particular, by Observation 2.2, $\text{wst}(\mathcal{C}; n) \leq 2\text{wst}(\mathcal{C}^r; n)$. Similarly, $\text{wst}(\mathcal{C}; n) \leq 2\text{wst}(\mathcal{C}^c; n)$.*

Griddings and grid classes

We say that a set P of points in the plane is in *general position* if no two points of P are on the same horizontal or vertical line. We say that two finite point sets P and Q in general position are *order-isomorphic*, if there is a bijection from P to Q which preserves both the left-to-right and the bottom-to-top order of the points. Note that any finite set P of points in general position is order-isomorphic to a permutation diagram of a unique permutation α ; we then say that P *induces a copy of* α .

Let $\pi \in \mathcal{S}_n$ be a permutation with a permutation diagram $D(\pi)$. A $k \times \ell$ *gridding* of π is a set of $k + 1$ vertical lines v_0, v_1, \dots, v_k numbered left to right and $\ell + 1$ horizontal lines h_0, h_1, \dots, h_ℓ numbered bottom to top, such that $D(\pi)$ is inside the rectangle bounded by v_0, v_k, h_0 and h_ℓ , and no point of $D(\pi)$ lies on any of the $k + \ell$ lines of the gridding. For $i \in [k]$ and $j \in [\ell]$, the *cell* (i, j) of the gridding, denoted is the rectangle bounded by the lines v_{i-1}, v_i, h_{j-1} and h_j . A permutation together with its gridding is referred to as a *gridded permutation*. If π is a gridded permutation and X a cell of its gridding, we let $\pi[X]$ denote the set $D(\pi) \cap X$, and the permutation induced by this set is simply referred to as the subpermutation of π induced by X .

A *gridding matrix* is a matrix M whose every entry is a permutation class; see Figure 1 for an example. For consistency with the Cartesian coordinates we use for

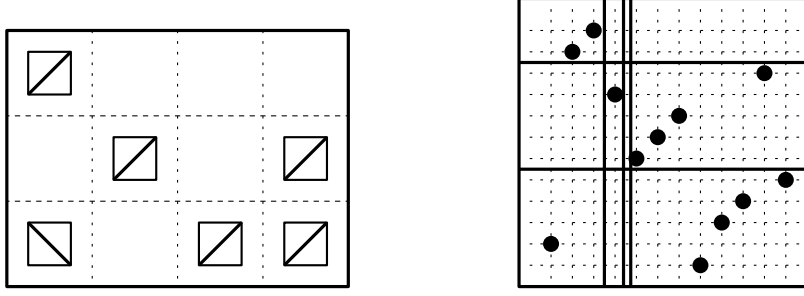


Figure 1: Left: a 4×3 gridding matrix M ; by convention, we assume that the empty cells represent the permutation class that only contains the empty permutation. Right: an example of an M -gridded permutation, represented by its diagram.

permutation diagrams, we will assume that rows in a matrix are numbered bottom to top, and we let $M_{i,j}$ denote the entry in column i and row j of the matrix M . Suppose M is a gridding matrix with k columns and ℓ rows. An M -gridding of a permutation π is a $k \times \ell$ gridding with the property that for every $(i, j) \in [k] \times [\ell]$, the cell (i, j) of the gridding induces in π a permutation belonging to the class $M_{i,j}$. The *grid class* determined by M , denoted $\text{Grid}(M)$, is the set of those permutations π that admit an M -gridding.

An important special case of grid classes are the so-called *monotone juxtapositions*, whose gridding matrices have shape 2×1 or 1×2 , and each of their two entries is \square or \boxtimes . In particular, there are eight monotone juxtapositions.

2.1 Some notable classes

We will now introduce several permutation classes that will play an important role in our results. See Figure 2 for some examples and Table 1 for a complete list.

The class of *layered permutations*, denoted \mathcal{L} , is the sum-closure of the class \boxtimes . Equivalently, it can be defined as $\text{Av}(231, 312)$. We remark that sorting by layered permutations is equivalent by sorting with a series of so-called *pop-stacks*.

The *Fibonacci class*, denoted \mathcal{F} , is the sum-closure of the set $\{1, 21\}$, or equivalently the class of 321-avoiding layered permutations. Its name refers to the fact that $|\mathcal{F}_n|$ is the n -th Fibonacci number. Sorting with the Fibonacci class \mathcal{F} corresponds to sorting where in each step we perform parallel transpositions of disjoint adjacent pairs of elements. This includes, as a special case, an algorithm known as odd-even sort (also known as parallel neighbor-sort), introduced and analyzed by Habermann [21], who showed that the algorithm requires n rounds in the worst case. This implies that $\text{wst}(\mathcal{F}; n) = O(n)$. This is asymptotically tight since, e.g., any permutation π with $\pi(n) = 1$ requires at least $n - 1$ \mathcal{F} -steps to be sorted. Thus, $\text{wst}(\mathcal{F}; n) = \Theta(n)$.

The *rotation class* is the class $\mathcal{R} := \{\iota_a \ominus \iota_b; a, b \in \mathbb{N}_0\}$. Note that for $\rho \in \mathcal{R}_n$ and $\pi \in \mathcal{S}_n$, $\pi \circ \rho$ is a cyclic rotation of the sequence π . We will also consider the class $\mathcal{RR} := \mathcal{R} \cup \mathcal{R}^r$, corresponding to rotations and their reversals. Both these classes are composition closed, and in particular have infinite sorting time, as they cannot sort any permutation that does not belong to the class.

The *parallel block transposition class*, denoted \mathcal{PBT} , is the sum-closure of \mathcal{R} . To our knowledge, this class has not been considered before, but it emerges naturally in

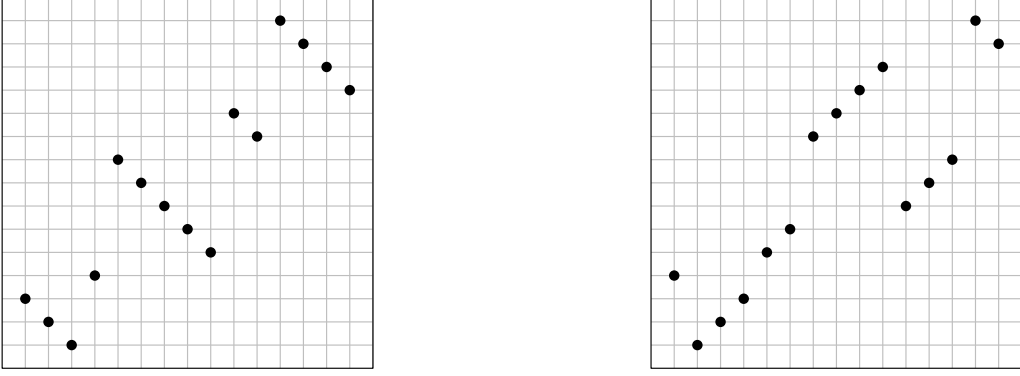


Figure 2: From left to right: diagram of a layered permutation, diagram of a permutation from the Fibonacci class \mathcal{F} , and a diagram of a permutation from the class \mathcal{PBT} .

our results as one of the twelve minimal classes with polylogarithmic sorting time.

The class $\mathcal{C}_{\text{Bub}} := \{\iota_a \oplus 21 \oplus \iota_b; a, b \in \mathbb{N}_0\} \cup \square$ corresponds to sorting by adjacent transpositions. This includes, e.g., the classical bubble sort algorithm. It can be easily seen that $\text{wst}(\mathcal{C}_{\text{Bub}}; n) = \binom{n}{2}$. A slight modification yields the class $\mathcal{T} := \{\iota_a \oplus 21 \oplus \iota_b; a, b \in \mathbb{N}_0\} \cup \{1 \ominus \iota_c \ominus 1; c \in \mathbb{N}_0\} \cup \square$, which corresponds to cyclic adjacent transpositions, where positions 1 and n are treated as adjacent.

The class $k\text{-Fringe} := \{\pi \oplus \iota_a \oplus \sigma; a \in \mathbb{N}_0, |\sigma| \leq k, |\pi| \leq k\}$ contains permutations in which only the prefix of length k and the suffix of length k can be permuted. We also define $k\text{-RFringe} := k\text{-Fringe} \cup (k\text{-Fringe})^r$.

The grid class $\mathcal{C}_{\text{Pan}} = \text{Grid} \left(\begin{smallmatrix} & \square \\ \square & \end{smallmatrix} \right)$ corresponds to sorting where in each step, we reverse the prefix of the permutation being sorted. This is known in the literature as *pancake sorting*. It is well known that $\text{wst}(\mathcal{C}_{\text{Pan}}; n) = \Theta(n)$. It is easy to see that the same asymptotic estimate also applies to $\mathcal{C}_{\text{Pan}}^r$, $\mathcal{C}_{\text{Pan}}^c$ and $\mathcal{C}_{\text{Pan}}^f$.

The grid class

$$\mathcal{C}_{\text{Ins}} = \text{Grid} \left(\begin{smallmatrix} & \square \\ \bullet & \\ & \square \end{smallmatrix} \right)$$

corresponds to sorting where in each step, we take the first element of the input and insert it to an arbitrary position i , while moving the elements in positions $2, 3, \dots, i$ one step to the left. This corresponds to the classical insertion sort algorithm. The standard application of the algorithm can sort any input in $n-1$ steps, by maintaining the invariant that after k steps, the final $k+1$ elements are in increasing order; hence, $\text{wst}(\mathcal{C}_{\text{Ins}}; n) \leq n-1$. This bound is tight, since any permutation π with $\pi(n) = 1$ requires $n-1$ steps to be sorted.

3 Classes that cannot sort

Let us say that a class \mathcal{C} *cannot sort*, if $\text{wst}(\mathcal{C}; n) = +\infty$ for some n . Note that this is equivalent to $\mathcal{C}_n^* \neq \mathcal{S}_n$. Moreover, if \mathcal{C}_n^* is different from \mathcal{S}_n for some n , then for any $m \geq n$, \mathcal{C}_m^* is different from \mathcal{S}_m , since \mathcal{C}^* is a permutation class.

Notation	Description
\mathcal{L}	$\oplus \boxtimes$
\mathcal{F}	$\oplus \{1, 12\} = \mathcal{L} \cap \text{Av}(321)$
\mathcal{R}	$\text{Grid} \begin{pmatrix} \boxtimes & \\ & \boxtimes \end{pmatrix}$
\mathcal{RR}	$\mathcal{R} \cup \mathcal{R}^r$
\mathcal{PBT}	$\oplus \mathcal{R}$
\mathcal{C}_{Bub}	$\text{Grid} \begin{pmatrix} & & \boxtimes \\ & \bullet & \\ \boxtimes & & \bullet \end{pmatrix}$

Notation	Description
\mathcal{T}	$\mathcal{C}_{\text{Bub}} \cup \text{Grid} \begin{pmatrix} \bullet & & \\ & \boxtimes & \\ & & \bullet \end{pmatrix}$
$k\text{-Fringe}$	$\text{Grid} \begin{pmatrix} & & \mathcal{S}_{\leq k} \\ & \boxtimes & \\ \mathcal{S}_{\leq k} & & \end{pmatrix}$
$k\text{-RFringe}$	$k\text{-Fringe} \cup k\text{-Fringe}^r$
\mathcal{C}_{Pan}	$\text{Grid} \begin{pmatrix} & \boxtimes \\ \boxtimes & \end{pmatrix}$
\mathcal{C}_{Ins}	$\text{Grid} \begin{pmatrix} & \boxtimes \\ \bullet & \\ & \boxtimes \end{pmatrix}$

Table 1: A summary of relevant permutation classes

When seeking to characterize the maximal classes that cannot sort, we may restrict our attention to the classes \mathcal{C} that are *composition closed*, that is, those that satisfy $\mathcal{C} = \mathcal{C}^*$. Composition closed permutation classes have been characterized by Atkinson and Beals [4]. Translating their results into our terminology, we obtain the following proposition.

Proposition 3.1. *A permutation class \mathcal{C} cannot sort if and only if it satisfies, for some $k \in \mathbb{N}$, one of the following two conditions:*

1. \mathcal{C} is a subclass of $\mathcal{RR} \cup \mathcal{S}_{\leq k}$, or
2. \mathcal{C} is a subclass of $k\text{-RFringe}$.

4 From infinity to $O(n^2)$

Let us now focus on showing that any class that can sort has at most quadratic sorting time. Our key concept will be the notion of *peg class*, which is a grid class with a gridding matrix whose every row and column has exactly one nonempty cell, and each nonempty cell is either the singleton class $\{1\}$ or one of the two monotone classes \boxtimes and \boxminus .

Peg classes have been previously studied mostly in the context of enumeration [2, 23, 22]. Indeed, it is known that a permutation class has polynomial growth rate if and only if it is a finite union of peg classes. For our purposes, we will use the following result.

Fact 4.1 (Homberger–Vatter [22, Theorem 3.1]). *If a permutation class \mathcal{C} does not contain as a subclass any monotone juxtaposition, nor \mathcal{F} nor \mathcal{F}^r , then \mathcal{C} is a finite union of peg classes.*

$$\begin{aligned}
\mathcal{C}_a &= \text{Grid} \begin{pmatrix} \bullet & & \\ & \bullet & \\ & & \boxplus \end{pmatrix}, \mathcal{C}_b = \text{Grid} \begin{pmatrix} \bullet & & \\ & \boxplus & \\ & & \bullet \end{pmatrix}, \mathcal{C}_c = \text{Grid} \begin{pmatrix} & & \bullet \\ \bullet & & \\ & \boxplus & \end{pmatrix}, \\
\mathcal{C}_d &= \text{Grid} \begin{pmatrix} & & \boxplus \\ & \bullet & \\ \boxplus & & \bullet \end{pmatrix}, \mathcal{C}_e = \text{Grid} \begin{pmatrix} \bullet & & \\ & \bullet & \\ & & \boxplus \end{pmatrix}.
\end{aligned}$$

Figure 3: Five peg classes that can sort in quadratic time.

It is not hard to see that any monotone juxtaposition has worst-case sorting time of order $O(\log n)$; we later prove this in Proposition 6.11. Also for the Fibonacci class \mathcal{F} , we have already pointed out that $\text{wst}(\mathcal{F}; n) = \Theta(n)$, and we may easily observe that $\text{wst}(\mathcal{F}^r; n) = O(n)$.

Thus, if a class \mathcal{C} has at least quadratic $\text{wst}(\mathcal{C}; n)$, it cannot contain any monotone juxtaposition and any symmetry of \mathcal{F} as a subclass. Thus, by Fact 4.1, such a class \mathcal{C} can be written as a finite union of *peg classes*.

Lemma 4.2. *Let \mathcal{C} be a peg class. If \mathcal{C} can sort, then $\text{wst}(\mathcal{C}; n) = O(n^2)$.*

Proof. If the gridding matrix of \mathcal{C} contains a cell equal to \boxplus as well as a cell equal to \boxminus , then \mathcal{C} contains $\mathcal{C}_{\text{Pan}} = \boxminus \oplus \boxplus$ or one of its symmetries as a subclass, and with the help of Observation 2.3 we easily conclude that $\text{wst}(\mathcal{C}; n) = O(n)$.

Suppose then that at most one of the two classes \boxminus and \boxplus occurs in the gridding of \mathcal{C} . If \boxminus occurs, then by Observation 2.3, $\text{wst}(\mathcal{C}; n) \leq 2\text{wst}(\mathcal{C}^r; n)$, and \mathcal{C}^r is a peg class whose gridding matrix contains \boxplus . It is therefore enough to deal with peg classes whose gridding matrix contains \boxplus but not \boxminus . We assume from now on that \mathcal{C} is such a peg class.

Let M be the gridding matrix of the class \mathcal{C} , and assume that M has been chosen to have the smallest possible number of rows and columns. In particular, this means that if $M_{i,j} = \boxplus$ for some i and j , then either the cell $M_{i,j}$ is in the last column or last row of M , or the cell $M_{i+1,j+1}$ is empty. If not, then we could delete column $i+1$ and row $j+1$ from M to obtain a smaller gridding matrix representing the same class \mathcal{C} . By the same reasoning, if $M_{i,j} = \boxplus$ for some $i > 1$ and $j > 1$, then $M_{i-1,j-1}$ is empty.

Consider now the five peg classes from Figure 3. We claim that all these classes, and therefore also their inverses and flips, have at most quadratic sorting time.

Consider the class \mathcal{C}_a , which corresponds to sorting where in every step we either move the first element of the input to the last position, shifting the remaining elements one step forward, or we move the first two elements to the end, while simultaneously changing their relative position. We may easily see that by a sequence of at most n steps of \mathcal{C}_a , we may simulate a single step of the class \mathcal{F} of parallel adjacent transpositions on an input of size n . As we pointed out in Section 2, $\text{wst}(\mathcal{F}; n) = O(n)$. It follows that $\text{wst}(\mathcal{C}_a; n) = \text{wst}(\mathcal{C}_a^{-1}; n) = \text{wst}(\mathcal{C}_a^f; n) = O(n^2)$. The same argument also applies to the remaining classes \mathcal{C}_b , \mathcal{C}_c , \mathcal{C}_d and \mathcal{C}_e .

Consider again the peg class \mathcal{C} from our lemma, with its gridding matrix M . We will show that \mathcal{C} either contains, as a subclass, one of the classes $\mathcal{C}_a, \dots, \mathcal{C}_e$ from Figure 3 or its inverse or flip. We know that M contains at least one entry equal to \square , so let us fix one such entry $M_{i,j}$. We now let M_{TL} , M_{TR} , M_{BL} and M_{BR} denote the submatrices of M formed by the cells strictly to the top-left, top-right, bottom-left and bottom-right of the cell $M_{i,j}$, respectively; for instance M_{BL} is formed by the intersection of the leftmost $i-1$ columns and bottommost $j-1$ rows of M .

If both M_{TL} and M_{BR} contain a non-empty cell, then \mathcal{C} contains \mathcal{C}_b as a subclass, so suppose this is not the case, and assume, without loss of generality, that M_{BR} only contains empty cells. If the grid class of M_{TL} contains 21, then \mathcal{C} contains \mathcal{C}_a and we are done, so we assume from now on that M_{TL} generates a (possibly empty) subclass of \square . It follows that at least one of M_{BL} and M_{TR} must contain a nonempty cell, otherwise \mathcal{C} would be a subclass of \mathcal{R} and it could not sort.

Suppose without loss of generality that M_{TR} contains at least as many \square -cells as M_{BL} , and in case neither M_{TR} nor M_{BL} contains any \square -cell, suppose that M_{TR} has at least one nonempty cell. Let $M_{i+1,k}$ be a nonempty cell in the column of M right next to $M_{i,j}$. Note that k is at least $j+2$, since M_{BR} only has empty cells and if the cell $M_{i+1,j+1}$ were nonempty, we could combine it with $M_{i,j}$ contradicting the minimality of M .

Let $M_{\ell,j+1}$ be a nonempty cell of M in the row above $M_{i,j}$. If $M_{\ell,j+1}$ is inside M_{TL} , then \mathcal{C} contains \mathcal{C}_c as a subclass, so suppose that $M_{\ell,j+1}$ is inside M_{TR} . If at least one of the two cells $M_{i+1,k}$ and $M_{\ell,j+1}$ is equal to \square , then \mathcal{C} contains either \mathcal{C}_c^f or $(\mathcal{C}_c^f)^{-1}$ as a subclass and we are done. Suppose that this is not the case. Since \mathcal{C} is not a subclass of $k\text{-Fringe}$ for any fixed k , there must be either a nonempty cell in M_{TL} or a cell equal to \square in M_{TR} . In the former case, \mathcal{C} contains as a subclass either \mathcal{C}_c or \mathcal{C}_e^f . In the latter case, it contains either \mathcal{C}_d or a class symmetric to \mathcal{C}_c as a subclass.

In all the cases considered, we found that \mathcal{C} contains a class symmetric to one of $\mathcal{C}_a, \dots, \mathcal{C}_e$ as a subclass, proving that $\text{wst}(\mathcal{C}; n) = O(n^2)$ as claimed. \square

Lemma 4.2 shows that a peg class that can sort has at most quadratic sorting time. We now turn our attention to classes that can be expressed as unions of two or more peg classes. The goal is to show that even if the individual peg classes cannot sort, as long as their union can sort, it has at most quadratic sorting time.

Lemma 4.3. *Assume that a class \mathcal{C} is a subclass of $\mathcal{RR} \cup \mathcal{S}_{\leq k}$ for some k , but is not a subclass of $k\text{-RFringe}$ for any k , and that a class \mathcal{D} is a subclass of $k\text{-RFringe}$ for some k , but not a subclass of $\mathcal{RR} \cup \mathcal{S}_{\leq k}$ for any k . Then $\text{wst}(\mathcal{C} \cup \mathcal{D}; n) = O(n^2)$.*

Proof. By assumption, \mathcal{C} contains infinitely many permutations that belong to \mathcal{RR} but not to $0\text{-RFringe} = \text{Av}(21) \cup \text{Av}(12)$. Any such permutation has the form $\iota_a \ominus \iota_b$ or $\delta_a \oplus \delta_b$ for some $a, b > 0$. It follows that \mathcal{C} contains as a subclass the peg class $\text{Grid} \left(\begin{smallmatrix} \bullet & \square \\ \bullet & \end{smallmatrix} \right)$ or one of its symmetries.

Similarly, the class \mathcal{D} contains, for some k , infinitely many permutations belonging to $k\text{-RFringe}$ but not to \mathcal{RR} . From this, we easily deduce that \mathcal{D} contains as a subclass the peg class $\text{Grid} \left(\begin{smallmatrix} \bullet & \square \\ \bullet & \bullet \end{smallmatrix} \right)$ or one of its symmetries.

We may then easily check that any permutation from the class \mathcal{F}_n can be obtained by composing $O(n)$ permutations from $\mathcal{C} \cup \mathcal{D}$, implying that $\text{wst}(\mathcal{C} \cup \mathcal{D}; n) = O(n^2)$. \square

Combining the previous arguments, we reach the main result of this section.

Proposition 4.4. *A class \mathcal{C} that can sort satisfies $\text{wst}(\mathcal{C}; n) = O(n^2)$.*

Proof. Suppose that \mathcal{C} can sort. If \mathcal{C} contains a monotone juxtaposition or any symmetry of the class \mathcal{F} as a subclass, then its worst-case sorting time is $O(n)$. Otherwise, \mathcal{C} can be written as a union of finitely many peg classes $\mathcal{C}_1, \dots, \mathcal{C}_t$, as shown by Albert et al. and Homberger–Vatter [2, 22]. If any of the peg classes \mathcal{C}_i can sort, then $\text{wst}(\mathcal{C}; n) \leq \text{wst}(\mathcal{C}_i; n) = O(n^2)$ by Lemma 4.2. Suppose then that none of the peg classes $\mathcal{C}_1, \dots, \mathcal{C}_t$ can sort, and in particular, by Proposition 3.1, each \mathcal{C}_i is either a subclass of $\mathcal{RR} \cup \mathcal{S}_{\leq k}$ or a subclass of $k\text{-RFringe}$ for some k .

If all of $\mathcal{C}_1, \dots, \mathcal{C}_t$ are subclasses of $k\text{-RFringe}$, or all of them are subclasses of $\mathcal{RR} \cup \mathcal{S}_{\leq k}$, then \mathcal{C} cannot sort by Proposition 3.1. Otherwise there is a \mathcal{C}_i which is not a subclass of $\mathcal{RR} \cup \mathcal{S}_{\leq k}$ for any k , and a \mathcal{C}_j which is not a subclass of $k\text{-RFringe}$ for any k , and by Lemma 4.3, $\text{wst}(\mathcal{C}; n) \leq \text{wst}(\mathcal{C}_i \cup \mathcal{C}_j; n) = O(n^2)$. \square

5 From quadratic to linear

Recall that \mathcal{T} denotes the class of cyclic adjacent transpositions. To show there is a gap between quadratic and linear worst-case sorting times, and to describe the classes with quadratic time, we need to introduce another concept: the *reduced inversion number* of a permutation π , denoted $\text{rin}(\pi)$, is the smallest integer $k \in \mathbb{N}_0$ such that π belongs to $\mathcal{T}^{\circ k} \circ \mathcal{RR}$.

The reduced inversion number $\text{rin}(\pi)$ is inspired by the (better-known) *inversion number* of a permutation π , which can be defined as the smallest k such that π belongs to $\mathcal{C}_{\text{Bub}}^{\circ k}$. Observe that the inversion number is an upper bound for the reduced inversion number.

Lemma 5.1. *For any $\alpha, \beta \in \mathcal{RR}_n$ and any $\pi \in \mathcal{S}_n$, $\text{rin}(\alpha \circ \pi \circ \beta) = \text{rin}(\pi)$.*

Proof. First, notice that $\mathcal{RR} \circ \mathcal{T} = \mathcal{T} \circ \mathcal{RR}$; in other words, a permutation obtained by composing a rotation (with optional reversal) with a cyclic transposition can also be obtained by performing a cyclic transposition first followed by a rotation (with optional reversal), and vice versa. It is also easy to see that $\mathcal{RR} \circ \mathcal{RR} = \mathcal{RR}$, that is, the class \mathcal{RR} is closed under compositions.

Suppose now that we are given α, β and π as above, and let $k := \text{rin}(\pi)$. Let σ be the permutation $\alpha \circ \pi \circ \beta$. Then π belongs to $\mathcal{T}^{\circ k} \circ \mathcal{RR}$, and hence σ belongs to $\mathcal{RR} \circ \mathcal{T}^{\circ k} \circ \mathcal{RR}^{\circ 2} = \mathcal{T}^{\circ k} \circ \mathcal{RR}$. This shows that $\text{rin}(\sigma) \leq \text{rin}(\pi)$. Conversely, π is equal to $\alpha^{-1} \circ \sigma \circ \beta^{-1}$, and since α^{-1} and β^{-1} both belong to \mathcal{RR} , we may repeat the above argument to show that $\text{rin}(\pi) \leq \text{rin}(\sigma)$. We conclude that $\text{rin}(\pi) = \text{rin}(\sigma)$, as claimed. \square

Lemma 5.2. *For any $\sigma, \pi \in \mathcal{S}_n$,*

$$\text{rin}(\pi \circ \sigma) \leq \text{rin}(\pi) + \text{rin}(\sigma).$$

Proof. Let $k := \text{rin}(\sigma)$ and $\ell := \text{rin}(\pi)$. Then $\sigma \circ \pi$ belongs to the class $\mathcal{T}^{\circ k} \circ \mathcal{RR} \circ \mathcal{T}^{\circ \ell} \circ \mathcal{RR} = \mathcal{T}^{\circ(k+\ell)} \circ \mathcal{RR}$, hence $\text{rin}(\sigma \circ \pi) \leq k + \ell$. \square

For a class \mathcal{C} and $n \in \mathbb{N}$, let us define $\text{rin}(\mathcal{C}; n) := \max\{\text{rin}(\sigma); \sigma \in \mathcal{C}_n\}$. As a direct consequence of Lemma 5.2, we obtain the following relationship between sorting time and reduced inversion number.

Corollary 5.3. *For any class \mathcal{C} and any $\pi \in \mathcal{S}_n$,*

$$\text{st}(\mathcal{C}; \pi) \geq \frac{\text{rin}(\pi)}{\text{rin}(\mathcal{C}; n)}.$$

To make use of this corollary, we need to show that there are permutations $\pi \in \mathcal{S}_n$ of large $\text{rin}(\pi)$.

Lemma 5.4. *We have $\text{rin}(\mathcal{S}; n) = \Theta(n^2)$.*

Proof. A permutation π of length n has inversion number at most $\binom{n}{2}$, and hence $\text{rin}(\mathcal{S}; n) \leq \binom{n}{2}$. For a lower bound, we first introduce some auxiliary notions. For two numbers $i, j \in [n]$, we define their *cyclic distance* $\text{cd}(i, j)$ by

$$\text{cd}(i, j) = \min\{|i - j|, n - |i - j|\},$$

and for a permutation $\pi \in \mathcal{S}_n$, define its *total cyclic distance* $\text{tcd}(\pi)$ by

$$\text{tcd}(\pi) = \text{cd}(\pi(1), \pi(n)) + \sum_{i=1}^{n-1} \text{cd}(\pi(i), \pi(i+1)).$$

Observe that for any $\rho \in \mathcal{RR}$ and $\pi \in \mathcal{S}_n$, both $\text{tcd}(\pi \circ \rho)$ and $\text{tcd}(\rho \circ \pi)$ are equal to $\text{tcd}(\pi)$. Observe also that for any $\sigma \in \mathcal{T}$, we have $\text{tcd}(\sigma \circ \pi) \leq \text{tcd}(\pi) + 4$. Consequently, given that the identity permutation ι_n satisfies $\text{tcd}(\iota_n) = n$, we see that for any $\pi \in \mathcal{T}^{\circ k}$ we have $\text{tcd}(\pi) \leq n + 4k$. Combining these observations, we see that any $\pi \in \mathcal{S}_n$ satisfies $\text{tcd}(\pi) \leq n + 4\text{rin}(\pi)$.

To prove the desired lower bound $\text{rin}(\mathcal{S}, n) = \Omega(n^2)$, it is enough to find a permutation $\pi \in \mathcal{S}_n$ with $\text{tcd}(\pi) = \Omega(n^2)$. An example can be the permutation $\pi = (1, \lfloor n/2 \rfloor + 1, 2, \lfloor n/2 \rfloor + 2, 3, \lfloor n/2 \rfloor + 3, \dots)$ obtained by interleaving the two sequences $(1, 2, 3, \dots, \lfloor n/2 \rfloor)$ and $(\lfloor n/2 \rfloor + 1, \lfloor n/2 \rfloor + 2, \dots, n)$. \square

Combining Corollary 5.3 and Lemma 5.4, we obtain the following bound.

Corollary 5.5. *For any class \mathcal{C} ,*

$$\text{wst}(\mathcal{C}; n) \geq \Omega\left(\frac{n^2}{\text{rin}(\mathcal{C}; n)}\right).$$

For a gridding matrix M with m rows and n columns, the *reversal* of M , denoted M^r , is the gridding matrix obtained by replacing each individual entry $\mathcal{C} = M_{i,j}$ with its reversal \mathcal{C}^r , and then reversing the order of columns of the whole matrix M ; formally, if $M_{i,j} = \mathcal{C}$, then $M_{n-i+1,j}^r = \mathcal{C}^r$. Note that the permutations in the class $\text{Grid}(M^r)$ are precisely the reverses of those in $\text{Grid}(M)$.

$$\mathcal{C}_{\text{Ins}} = \text{Grid} \begin{pmatrix} \bullet & & \square \\ & & \\ & \square & \end{pmatrix}, \quad \mathcal{C}_1 = \text{Grid} \begin{pmatrix} \bullet & & \\ & \square & \\ & & \square \end{pmatrix}, \quad \mathcal{C}_2 = \text{Grid} \begin{pmatrix} \square & & \\ & \bullet & \\ & & \square \end{pmatrix}$$

Figure 4: Three classes with congruent gridding matrices, all with linear sorting time

Let M be a gridding matrix. The *column rotation* of M is the operation that transforms M into the matrix M' of the same size, where the first column of M is equal to the last column of M' , and for each $i > 1$ the i -th column of M is equal to the $(i-1)$ -th column of M' . The *row rotation* of M is defined analogously. We say that a gridding matrix M is *congruent* to a matrix M' , if M can be transformed into M' by a sequence of column rotations, row rotations and reversals. Note that if M and M' are congruent, then for any $\pi \in \text{Grid}(M)$ there are permutations $\rho, \rho' \in \mathcal{RR}$ such that $\rho \circ \pi \circ \rho'$ belongs to $\text{Grid}(M')$. In particular, $\text{rin}(\text{Grid}(M); n) = \text{rin}(\text{Grid}(M'); n)$.

Lemma 5.6. *For any peg class \mathcal{C} , either $\text{rin}(\mathcal{C}; n)$ is bounded, or $\text{wst}(\mathcal{C}; n) = O(n)$.*

Proof. Let \mathcal{C} be a peg class with a gridding matrix M , and suppose that $\text{rin}(\mathcal{C}; n)$ is unbounded. In particular, M contains at least one entry equal to \square or to \sqcap . Let M' be a gridding matrix congruent to M whose top-right entry is equal to \square .

Suppose that M' contains an entry equal to \sqcap . Then $\text{Grid}(M')$ contains, as a subclass, the peg class $\mathcal{C}_{\text{Pan}} = \text{Grid} \begin{pmatrix} & \square \\ \sqcap & \end{pmatrix}$, and \mathcal{C} contains one of the symmetries of this class. Since all the symmetries of \mathcal{C}_{Pan} have linear worst-case sorting time, we conclude that $\text{wst}(\mathcal{C}; n) = O(n)$.

Suppose then that M' does not contain \sqcap . Let $M'_{i,j}$ be an entry of M' equal to \square . If M' contains a nonempty entry strictly to the left and above $M'_{i,j}$, then $\text{Grid}(M')$ contains \mathcal{C}_{Ins} as a subclass. It follows that \mathcal{C} contains as a subclass one of the three classes from Figure 4 or one its symmetries. Noting that $\mathcal{C}_{\text{Ins}} \subseteq \mathcal{C}_1^{\circ 2}$ and $\mathcal{C}_{\text{Ins}} \subseteq \mathcal{C}_2^{\circ 2}$, and applying Observation 2.3 if needed, we easily conclude that $\text{wst}(\mathcal{C}; n) = O(n)$.

An analogous argument applies when M' has a nonzero entry to the right and below $M'_{i,j}$.

Finally, suppose that no entry of M' equal to \square has any nonzero entry to the left and above, or to the right and below of itself. It follows that all the entries of M' equal to \square appear on the main diagonal, and M' is a block-diagonal matrix whose every block consists either of a single entry \square , or a submatrix that has no entry equal to \square . It then easily follows that $\text{rin}(\text{Grid}(M'); n)$ is bounded, and $\text{rin}(\mathcal{C}; n)$ is bounded as well. \square

Proposition 5.7. *For a class \mathcal{C} , the following are equivalent:*

1. $\text{wst}(\mathcal{C}; n) = o(n^2)$,
2. $\text{wst}(\mathcal{C}; n) = O(n)$, and
3. $\text{rin}(\mathcal{C}; n)$ is unbounded as $n \rightarrow \infty$.

Proof. By Corollary 5.5, if $\text{rin}(\mathcal{C}; n)$ is bounded then $\text{wst}(\mathcal{C}; n) = \Omega(n^2)$. This shows the implication $1 \implies 3$.

To prove $3 \implies 2$, suppose now that $\text{rin}(\mathcal{C}; n)$ is unbounded for a class \mathcal{C} . If \mathcal{C} contains as a subclass a monotone juxtaposition or its inverse or if it contains the Fibonacci class \mathcal{F} or its reverse, then clearly $\text{wst}(\mathcal{C}; n) = O(n)$. On the other hand, if \mathcal{C} contains no such subclass, then it can be expressed as a union of finitely many peg classes $\mathcal{C}_1, \dots, \mathcal{C}_k$. Since $\text{rin}(\mathcal{C}; n)$ is unbounded, there must be an $i \in [k]$ such that $\text{rin}(\mathcal{C}_i; n)$ is unbounded. By Lemma 5.6, $\text{wst}(\mathcal{C}_i; n) = O(n)$, and therefore $\text{wst}(\mathcal{C}; n) = O(n)$. This proves the implication $3 \implies 2$.

The implication $2 \implies 1$ is trivial. \square

6 From $\Omega(\sqrt{n})$ to polylog

The most difficult part of our argument deals with the gap between sorting times $O(\log^2 n)$ and $\Omega(\sqrt{n})$. The criterion distinguishing the two types of classes is rather simple to state, though. For ease of notation, let \mathfrak{X} be the set of permutation classes that contains the eight monotone juxtapositions and the four classes \mathcal{L} , \mathcal{L}^r , \mathcal{PBT} and \mathcal{PBT}^r . We say that a permutation class \mathcal{C} is \mathfrak{X} -avoiding if it does not contain any class from \mathfrak{X} as a subclass, otherwise we say that it is \mathfrak{X} -containing.

Theorem 6.1. *Any \mathfrak{X} -avoiding class \mathcal{C} satisfies $\text{wst}(\mathcal{C}; n) = \Omega(\sqrt{n})$, while any \mathfrak{X} -containing class \mathcal{C} satisfies $\text{wst}(\mathcal{C}; n) = O(\log^2 n)$.*

The rest of this section is devoted to a detailed proof of Theorem 6.1. We first focus on showing that every \mathfrak{X} -avoiding class \mathcal{C} satisfies $\text{wst}(\mathcal{C}; n) = \Omega(\sqrt{n})$.

An *inversion* in a permutation π is a decreasing subsequence of length 2, i.e., a pair $(\pi(i), \pi(j))$ such that $i < j$ and $\pi(i) > \pi(j)$. The first ingredient in the proof of Theorem 6.1 is a structural result characterizing permutation classes admitting a linear bound on the number of inversions.

Proposition 6.2. *For a permutation class \mathcal{C} , the following statements are equivalent:*

1. *Neither \mathcal{R} nor \boxtimes is a subclass of \mathcal{C} .*
2. *There is a $k \in \mathbb{N}$ such that every permutation in \mathcal{C} avoids both δ_k and $\iota_k \ominus \iota_k$.*
3. *There is a $Q \in \mathbb{N}$ such that every permutation $\pi \in \mathcal{C}$ of size n has at most Qn inversions.*

Proof. The implication $3 \implies 1$ is clear, since both \boxtimes and \mathcal{R} contain permutations of size n with $\Omega(n^2)$ inversions, namely δ_n and $\iota_{\lfloor n/2 \rfloor} \ominus \iota_{\lfloor n/2 \rfloor}$. Likewise, the implication $1 \implies 2$ follows directly from the fact that any permutation in \mathcal{R} is a subpermutation of $\iota_k \ominus \iota_k$ for some k .

It remains to prove $2 \implies 3$. Fix k such that every permutation of \mathcal{C} avoids both δ_k and $\iota_k \ominus \iota_k$, and let π be a permutation of size n in \mathcal{C} . Since π avoids δ_k , it can be partitioned into $k - 1$ increasing subsequences I_1, \dots, I_{k-1} . We will show that π has at most $(k - 1)^3 n$ inversions.

Suppose that $(\pi(i), \pi(j))$ is an inversion, that is, $i < j$ and $\pi(i) > \pi(j)$. We say that the inversion has *type* (a, b) if $\pi(i)$ belongs to I_a and $\pi(j)$ belongs to I_b . Note

that this implies $a \neq b$. We will show that for any $(a, b) \in [k-1] \times [k-1]$ there are at most $(k-1)n$ inversions of type (a, b) . To this end, we define a graph $G_{a,b}$ on the vertex set $I_a \cup I_b$ whose edges are the pairs of vertices that form an inversion of type (a, b) . We claim that the graph $G_{a,b}$ is $(k-1)$ -degenerate, that is, each of its nonempty subgraphs has a vertex of degree at most $k-1$.

Suppose for contradiction that $G_{a,b}$ has a nonempty subgraph G' whose every vertex has degree at least k . The vertex set of G' has the form $I'_a \cup I'_b$ for some $I'_a \subseteq I_a$ and $I'_b \subseteq I_b$. Let x_1 be the rightmost vertex in I'_a , and let $N(x_1)$ be the vertices of I'_b adjacent to x_1 . By assumption, $|N(x_1)| \geq k$. Let y_1 be the rightmost (and therefore also topmost) vertex in $N(x_1)$, and let $N(y_1) \subseteq I'_a$ be its neighbors. Since any vertex in $N(x_1)$ is to the right of x_1 , it follows that any vertex in $N(x_1)$ is to the right of all the vertices in $N(y_1)$ (recall that x_1 is the rightmost vertex of I'_a). Moreover, since y_1 is topmost in $N(x_1)$, all the vertices of $N(x_1)$ are below all the vertices in $N(y_1)$. It follows that the set $N(x_1) \cup N(y_1)$ induces in π a permutation containing $\iota_k \ominus \iota_k$, which is a contradiction.

We conclude that the graph $G_{a,b}$ is $(k-1)$ -degenerate for any $(a, b) \in [k-1] \times [k-1]$, and therefore has at most $(k-1)(|I_a| + |I_b|) \leq (k-1)n$ edges. Summing over all values of a and b , we conclude that π has at most $(k-1)^3 n$ inversions. \square

The key part in the proof of Theorem 6.1 is a structural characterization of \mathfrak{X} -avoiding classes. To state it, we need to introduce some terminology.

An *integer interval* (or just *interval*, if there is no risk of confusion) is a finite set of consecutive positive integers, i.e., a set of the form $\{i \in \mathbb{N}; a \leq i \leq b\}$ for some $a, b \in \mathbb{N}$. For a finite set $X \subseteq \mathbb{N}$, its *intervalicity*, denoted $\text{Int}(X)$, is the smallest $k \in \mathbb{N}_0$ such that X can be expressed as a union of k integer intervals. For a finite point set $P \subseteq \mathbb{N} \times \mathbb{N}$, we let $P|_x$ and $P|_y$ denote its projections to the x -axis and to the y -axis, respectively. The *intervalicity* of the point set P , denoted again $\text{Int}(P)$, is the maximum of $\text{Int}(P|_x)$ and $\text{Int}(P|_y)$.

Proposition 6.3. *For every \mathfrak{X} -avoiding class \mathcal{C} there is a constant $K \equiv K(\mathcal{C})$, such that every $\pi \in \mathcal{C}$ admits a gridding with at most K rows and K columns such that for every cell X of the gridding, the following holds:*

1. *the point set $\pi[X]$ has intervalicity at most K ,*
2. *the permutation induced by $\pi[X]$ avoids at least one of ι_K and δ_K ,*
3. *the permutation induced by $\pi[X]$ avoids both $\iota_K \ominus \iota_K$ and $\delta_K \oplus \delta_K$, and*
4. *if n_X is the number of elements in $\pi[X]$, then either $\pi[X]$ or its reverse has at most $K \cdot n_X$ inversions.*

Before we prove the proposition, let us collect several known facts about permutation classes.

Fact 6.4 (Vatter [29, Corollary 5]). *For every permutation class \mathcal{C} that does not contain \mathcal{L} or \mathcal{L}^r as a subclass, there is a constant $K = K_{6.4}$ such that every permutation $\pi \in \mathcal{C}$ is a union of at most K monotone subsequences.*

Fact 6.5 (Vatter [28, Theorem 3.1]). *Let \mathcal{C} be a permutation class, and let α be a permutation. The following two statements are equivalent:*

- *There is a constant $K = K_{6.5}$ such that every permutation in \mathcal{C} has a gridding of size $K \times K$ in which every cell induces a permutation that avoids α .*
- *There is a constant Q such that neither $\bigoplus^Q \alpha$ nor $\bigotimes^Q \alpha$ belongs to \mathcal{C} .*

A *horizontal alternation* is a permutation in which all the odd values appear to the left of all the even values or vice versa. A *vertical alternation* is an inverse of a horizontal alternation. An *alternation* is a horizontal or vertical alternation.

Fact 6.6 (Huczynska–Vatter [23, Proposition 3.2]). *For a permutation class \mathcal{C} that does not contain any monotone juxtaposition as a subclass there is a constant $K = K_{6.6}$ such that \mathcal{C} does not contain any alternation of size K .*

Proof of Proposition 6.3. Let \mathcal{C} be an \mathfrak{X} -avoiding class, and fix $\pi \in \mathcal{C}$.

By Fact 6.6, there is a constant $K_{6.6}$ such that \mathcal{C} does not contain any alternation of size $K_{6.6}$ or more. We claim that this implies that in any gridding of π , every cell has intervalicity less than $K_{6.6}$. To see this, suppose for contradiction that there is a cell X such that the projection of $\pi[X]$ on the horizontal or vertical axis has intervalicity at least $K_{6.6}$. Without loss of generality, suppose that the projection is onto the horizontal axis, let $\ell \geq K_{6.6}$ be its intervalicity, and let P_1, P_2, \dots, P_ℓ be the intervals, numbered left to right, whose union is the projection of $\pi[X]$. For any $i \in [\ell - 1]$, fix an integer j_i such that $\max P_i < j_i < \min P_{i+1}$. Since $\pi(j_i)$ is not in $\pi[X]$, it is either smaller than all the values in $\pi[X]$ or larger than all the values in $\pi[X]$. Without loss of generality, there are at least $(\ell - 1)/2$ values of $i \in [\ell - 1]$ for which $\pi(j_i)$ is larger than all the values in $\pi[X]$. We can then choose at least $(\ell + 1)/2$ values in $\pi[X]$ which form a vertical alternation with these $(\ell - 1)/2$ ‘large’ values $\pi(j_i)$, showing that π has an alternation of size $\ell \geq K_{6.6}$, which is impossible. Thus, in any gridding of π , every cell has intervalicity less than $K_{6.6}$.

By Fact 6.4, there is a constant $K_{6.4}$ such that every permutation of \mathcal{C} is a union of at most $K_{6.4}$ monotone sequences. Let us therefore fix such a partition of π into monotone sequences, and suppose that the partition contains p increasing sequences I_1, \dots, I_p and q decreasing sequences D_1, \dots, D_q , for some $p + q \leq K_{6.4}$. Define $I_\cup = \bigcup_{i=1}^p I_i$ and $D_\cup = \bigcup_{i=1}^q D_i$. Notice that I_\cup avoids δ_{p+1} and D_\cup avoids ι_{q+1} .

We will now show that π has a gridding with bounded number of cells in which each cell is either disjoint from I_\cup or disjoint from D_\cup . For any choice of $i \in [p]$ and $j \in [q]$, we can easily construct a 2×2 gridding of π such that the increasing subsequence I_i intersects neither the top-left nor the bottom-right cell, while D_j intersects neither the bottom-left nor the top-right cells. Combining such gridgings for all the choices of $(i, j) \in [p] \times [q]$, we obtain a single gridding of size $(pq + 1) \times (pq + 1)$ in which every cell avoids either I_\cup or D_\cup , and in particular, every cell avoids ι_{q+1} or δ_{p+1} .

For the next step, we note that there are constants A, B such that neither $\bigoplus^A(\iota_B \ominus \iota_B)$ nor $\bigotimes^A(\iota_B \ominus \iota_B)$ belongs to \mathcal{C} . Indeed, if \mathcal{C} contained $\bigoplus^A(\iota_B \ominus \iota_B)$ for each A and B , then \mathcal{C} would contain \mathcal{PBT} as a subclass, and if \mathcal{C} contained $\bigotimes^A(\iota_B \ominus \iota_B)$ for all A, B , then \mathcal{C} would contain \mathcal{L}^r as a subclass. By Fact 6.5 applied

to $\alpha = \iota_B \ominus \iota_B$, we can refine the gridding of π obtained in the previous step by adding at most $K_{6,5}$ rows and columns to obtain a gridding in which no cell contains $\iota_B \ominus \iota_B$. By a symmetric argument, another round of refinement ensures that no cell of the gridding contains $\delta_{\overline{B}} \oplus \delta_{\overline{B}}$ for a constant \overline{B} .

Invoking Proposition 6.2 (or its symmetric version) to each cell of the resulting gridding, and choosing K large enough in terms of the constants appearing in the previous arguments, we conclude that the gridding obtained by this procedure has all the properties from the statement of the proposition. \square

For a permutation π , the *adjacency graph* $G(\pi)$ is a graph whose vertices are the elements of π , and two vertices are connected by an edge if and only if the corresponding two elements have adjacent positions or adjacent values. In particular, the adjacency graph is a union of two paths, one visiting the vertices in left-to-right order, and the other in bottom-to-top order.

The *tree-width* of the permutation π , denoted $\text{tw}(\pi)$, is then defined as the tree-width of $G(\pi)$. Note that there are permutations of size n with tree-width $\Omega(n)$ [1].

Now, we introduce the concept of *sorting diagram*, which will offer a convenient way to represent a sequence of steps performed during the sorting of a sequence. Let t and n be positive integers, and let $\vec{\sigma} = (\sigma^1, \sigma^2, \dots, \sigma^t)$ be a t -tuple of permutations of $[n]$. Define the permutation $\pi = \sigma^t \circ \sigma^{t-1} \circ \dots \circ \sigma^2 \circ \sigma^1$. In particular, by applying the steps $\sigma^1, \dots, \sigma^t$ in this order, we would sort the sequence π^{-1} into an increasing sequence.

Let P be the set $[(t+1)n] \times [(t+1)n]$, viewed as an integer grid in the plane. We partition P into blocks of size $n \times n$ as follows: for $a, b \in [t+1]$, let $P_{a,b}$ be the set $\{(i, j) \in P; (a-1)n < i \leq an \wedge (b-1)n < j \leq bn\}$. For a permutation $\sigma \in \mathcal{S}_n$, let $P_{a,b}[\sigma]$ denote the set of n points forming a translated copy of the diagram of σ inside the block $P_{a,b}$; formally,

$$P_{a,b}[\sigma] = \{((a-1)n + i, (b-1)n + \sigma(i)); i \in [n]\}.$$

We are now ready to describe the sorting diagram of the t -tuple of permutations $\vec{\sigma} = (\sigma^1, \sigma^2, \dots, \sigma^t)$; refer to Figure 5. Recall that ι_n is the increasing permutation $(1, 2, \dots, n)$. The *sorting diagram* of the t -tuple $\vec{\sigma} = (\sigma^1, \dots, \sigma^t)$, denoted $SD \equiv SD(\vec{\sigma})$, is the graph whose vertex set is the point set

$$V(SD) = \bigcup_{i=1}^{t+1} P_{t+2-i,i}[\iota_n] \cup \bigcup_{i=1}^t P_{t+2-i,i+1}[\sigma^i].$$

In particular, the vertex set consists of $t+1$ copies of the diagram of ι_n inside the blocks forming the decreasing diagonal of P , together with the copies of the diagrams of $\sigma^1, \dots, \sigma^t$, appearing right-to-left inside the blocks just above the decreasing diagonal. We refer to the n vertices inside the bottom-right block $P_{t+1,1}$ as *starting vertices*, and we label them s_1, \dots, s_n in left-to-right order. Similarly, the vertices inside the top-left block $P_{1,t+1}$ are the *terminal vertices*, and are labelled t_1, \dots, t_n in left-to-right order. The edge set of SD is determined as follows:

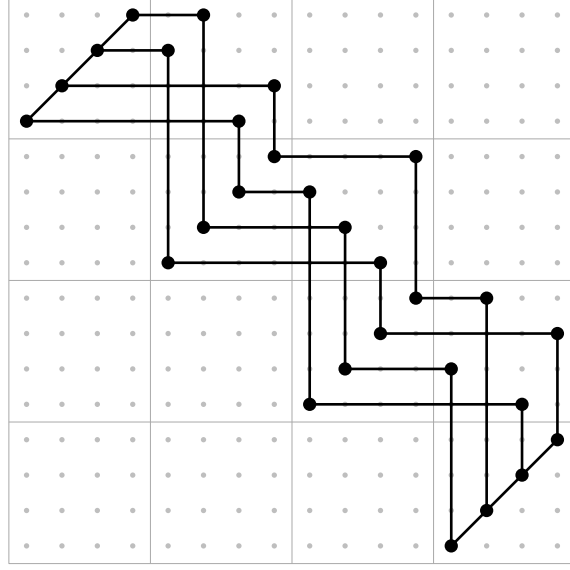


Figure 5: The sorting diagram of the triple of permutations $\sigma^1 = 2413, \sigma^2 = 3214, \sigma^3 = 3412$.

- Any two vertices of $V(SD)$ that lie on the same horizontal line are connected by an edge. We call these edges *horizontal edges*. Note that every vertex is incident to a unique horizontal edge, except for the n starting vertices.
- Any two vertices of $V(SD)$ that lie on the same vertical line are connected by an edge. We call these edges *vertical edges*. Note that every vertex is incident to a unique vertical edge, except for the n terminal vertices. Note also that the horizontal and vertical edges together form n vertex-disjoint paths, each connecting a starting vertex to a distinct terminal vertex, with the vertex s_i being connected to $t_{\pi(i)}$, where $\pi = \sigma^t \circ \dots \circ \sigma^1$.
- Finally, for any $i \in [n-1]$, the vertex s_i is connected to s_{i+1} and t_i to t_{i+1} . We call these edges as the *diagonal edges*.

This completes the description of the graph SD .

Lemma 6.7. *For a t -tuple of permutations $\vec{\sigma} = (\sigma^1, \dots, \sigma^t)$ and their composition $\pi = \sigma^t \circ \dots \circ \sigma^1$, the tree-width of the graph $SD(\vec{\sigma})$ is at least as large as the tree-width of the incidence graph $G(\pi)$ of π .*

Proof. The graph $G(\pi)$ is a minor of $SD(\vec{\sigma})$, since by contracting all the horizontal and vertical edges in $SD(\vec{\sigma})$, we obtain precisely the graph $G(\pi)$. The lemma follows, since tree-width is minor-monotone. \square

Proposition 6.8. *Let \mathcal{C} be an \mathfrak{X} -avoiding class. There are constants Q and R , such that for any $t \in \mathbb{N}$, any $n \in \mathbb{N}$, and any t -tuple $\vec{\sigma} = (\sigma^1, \dots, \sigma^t)$ of permutations from \mathcal{C}_n , the sorting diagram $SD(\vec{\sigma})$ can be drawn on a surface of genus at most $Q \cdot t$ with at most $R \cdot t \cdot n$ edge-crossings.*

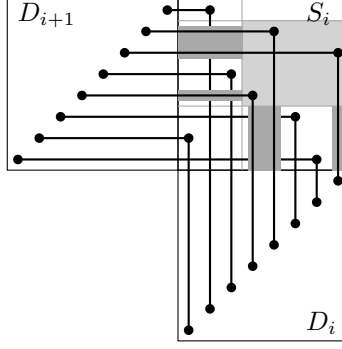


Figure 6: A fragment of a sorting diagram, induced by three consecutive blocks D_i , S_i and D_{i+1} . The light gray rectangle is a cell of the gridding of σ^i . The four dark gray rectangles represent the two horizontal and two vertical handles attached to this cell.

Proof. The definition of the graph $SD = SD(\vec{\sigma})$ already implies a drawing in the plane, where vertices are points and edges are straight-line segments. Unfortunately, such a drawing may have too many edge-crossings. Our strategy is to insert handles and cross-caps into this drawing, so that we eliminate enough of the crossings, without increasing the genus of the surface too much.

Recall from the definition of sorting diagram that the vertices of SD are organized into $2t + 1$ square blocks, which include $t + 1$ diagonal blocks $P_{t+1,1}, P_{t,2}, \dots, P_{1,t+1}$, as well as t blocks right above the diagonal, namely $P_{t+1,2}, P_{t,3}, \dots, P_{2,t+1}$. For ease of notation, for $i \in [t + 1]$ let D_i refer to the diagonal block $P_{t+2-i,i}$, and for $i \in [t]$ let S_i refer to the block right above D_i , i.e., $P_{t+2-i,i+1}$. The vertices of SD induce a copy of ι_n inside every D_i , and a copy of σ^i inside S_i .

In the straight-line drawing of SD , the only edge-crossings that occur are inside the blocks S_i , between a vertical edge connecting a vertex x from D_i to a vertex y in S_i , and a horizontal edge connecting a vertex x' in S_i to a vertex y' in D_{i+1} . Notice that such a pair of edges has a crossing if and only if (y, x') corresponds to a decreasing subsequence of σ^i . We will now modify the drawing of edges (and the surface they are drawn upon) inside each block S_i for $i \in [t]$, while the parts of the drawing inside the D_i 's are unaffected.

Fix $i \in [t]$. Let $K = K(\mathcal{C})$ be the constant from Proposition 6.3. As the first step, apply to the copy of σ^i inside S_i the gridding described by Proposition 6.3. We will now add handles to our surface, and route the edges along these handles, to ensure that no two edges cross, unless they have endpoint inside the same cell of the gridding. Refer to Figure 6.

Let X be a cell of the gridding of σ^i . By Proposition 6.3, $\sigma^i[X]$ has intervalicity at most K . In particular, there are K disjoint intervals I_1, \dots, I_K (some possibly empty) such that a vertical edge from D_i to S_i has an endpoint in X if and only if its projection to the x -axis is a point from $I_1 \cup I_2 \cup \dots \cup I_K$. For each $j \in [K]$ such that $I_j \neq \emptyset$, we attach a new handle H_j^\uparrow to our surface, oriented vertically, whose bottom end is near the bottom edge of S_i , the top end is near the bottom edge of X , and its horizontal position corresponds to the interval I_j (in case X is in the bottom row of the gridding, we omit these handles and leave the vertical edges

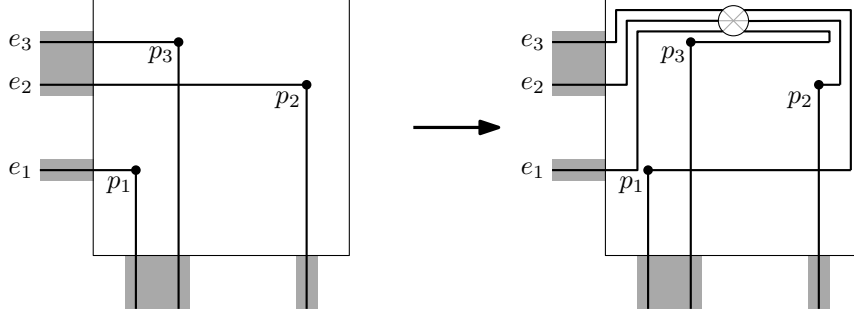


Figure 7: A modification of horizontal edges inside a cell X of the gridding of σ^i . The gray rectangles represent the handles adjacent to X , and the circle near the top boundary of the modified drawing represents a cross-cap.

into X unchanged). All the vertical edges whose projection on the x -axis lies in I_j are then routed through H_j^+ , from the point when they cross the boundary from D_i to S_i , till the point when they reach the bottom of edge of the cell X . This way, we remove any crossing between a vertical edge incident with a point in X and a horizontal edge incident with a cell X' in the gridding which lies below X .

We perform an analogous operation with the horizontal edges. Again, for a cell X , the projection of $\sigma^i[X]$ on the y -axis forms K intervals J_1, \dots, J_K , and for each such nonempty interval J_ℓ we attach a horizontal handle H_ℓ^- , whose left endpoint is near the boundary between S_i and D_{i+1} and its right endpoint is near the left edge of X ; if X is in the leftmost column of the gridding, these handles are omitted. The horizontal edges projecting into J_ℓ are then routed through the handle H_ℓ^- .

For each $i \in [t]$, this operation adds at most K vertical and K horizontal handles incident to any given cell X of the gridding of σ^i , so at most $2K^3t$ handles overall. After these handles are added, a vertical edge from D_i to S_i may cross a horizontal edge from S_i to D_{i+1} only if their endpoints inside S_i belong to the same cell of the gridding of σ^i . Unfortunately, there may still be too many such edge-crossings, and so we may need to perform additional modifications inside a cell X to reduce the number of crossings. The goal is for the number of crossings inside X to be linear in the size of $\sigma^i[X]$.

Consider again a cell X of the gridding of σ^i , and let n_X be the size of $\sigma^i[X]$. Observe that at this point, the number of edge-crossings inside X is equal to the number of decreasing subsequences of length 2 in $\sigma^i[X]$. By Proposition 6.3, $\sigma^i[X]$ contains either at most $K \cdot n_X$ decreasing subsequences of length 2, or at most $K \cdot n_X$ increasing subsequences of length 2. In the former case, we do not need to modify the drawing of the edges incident to X any further. Consider now that the second situation occurs; refer to Figure 7.

Let p_1, \dots, p_{n_X} be the points of $\sigma^i[X]$, numbered bottom to top, and let e_j be the horizontal edge incident to p_j . We will modify the drawings of all the horizontal edges inside X as follows: the edge e_j , instead of heading straight to the left from p_j towards the horizontal handle leading to D_{i+1} , will instead head straight to the right from p_j , until it reaches close to the right edge of X , then it will turn upwards, follow the right edge towards the top-right corner of X , then turn left and follow the top edge of X . All the horizontal edges will perform this ‘U-turn’, keeping their

mutual positions so that they do not introduce any mutual crossings. In particular, as the edges reach the top edge of X and head to the left, the edge e_1 will be the topmost, followed by e_2, e_3 etc. However, we need the edges to reach the handles in the opposite bottom-to-top order. To fix this, we attach a cross-cap near the top edge of X and let all the edges e_i pass through it. This reverses their bottom-to-top order, and it is easy to then extend the edges along the top and left edge of X towards their designated handles, without introducing any new crossings. After this modification, an edge e_j crosses a vertical edge incident to a vertex p_k if and only if (p_j, p_k) form an increasing subsequence of length 2 in $\sigma^i[X]$. In particular, there are now at most $K \cdot n_X$ crossings inside X .

In the end, we obtain a drawing of $SD(\vec{\sigma})$ on a surface with at most $2K^3t$ handles and at most K^2t cross-caps, and in particular, the genus of the surface is at most $(4K^3 + K^2)t$. Each block S_i contains at most Kn edge-crossings, so there are at most Knt edge-crossings overall. \square

Our next tool is an inequality that relates the genus, the crossing number and the tree-width of a graph. Its proof is based on an idea of Dujmović, Eppstein and Wood [15].

Proposition 6.9. *Every graph on n vertices that can be drawn on a surface of genus g with at most x edge-crossings has treewidth $O\left(\sqrt{(g+1)(n+x)}\right)$.*

Proof. Suppose a graph G on n vertices has been drawn on a surface of genus g with x edge-crossings. Let us orient the edges of G arbitrarily, so that each edge e has a designated tail t_e and head h_e . Next, let us replace each edge-crossing by a new vertex, thereby obtaining a crossing-free drawing of a new graph H with $n_H = n + x$ vertices. Since H can be drawn on a surface of genus g , it has treewidth at most $O(\sqrt{g \cdot n_H}) = O(\sqrt{g(n+x)})$ [16, 20]. Let us fix an optimal tree decomposition (T_H, β_H) of the graph H , where T_H is the corresponding decomposition tree, and $\beta_H: V(T_H) \rightarrow 2^{V(H)}$ is the function that assigns to each node q of T_H a bag $\beta(q)$ of vertices of H of size at most $O(\sqrt{g(n+x)})$.

We now modify this decomposition to a tree decomposition of G , as follows. Let w be any vertex of H that is not a vertex of G . This means that in the original drawing of G , w corresponded to a crossing of a pair of edges e and f . Let t_e and t_f be the tails of the two edges. We modify the decomposition (T_H, β) by replacing any occurrence of the vertex w in any of the bags by the two vertices t_e and t_f . After we perform these replacements for every vertex $w \in V(H) \setminus V(G)$, we may easily check that we obtain a tree decomposition of G of width at most twice as large as the width of (T_H, β) (see Dujmović et al. [15] for details). \square

We now have all the ingredients to prove the harder part of Theorem 6.1.

Proposition 6.10. *Any \mathfrak{X} -avoiding class \mathcal{C} has $\text{wst}(\mathcal{C}; n) = \Omega(\sqrt{n})$.*

Proof. Let $\pi \in \mathcal{S}_n$ be a permutation of tree-width $\Omega(n)$ (such permutations exist by a result of Ahal and Rabinovich [1, Theorem 3.4 and Proposition 3.6]). Let us put $t := \text{st}(\mathcal{C}; \pi)$, and let $\vec{\sigma} = (\sigma^1, \dots, \sigma^t)$ be a t -tuple of permutations from \mathcal{C} such that $\pi = \sigma^t \circ \sigma^{t-1} \circ \dots \circ \sigma^1$.

Consider the sorting diagram $SD = SD(\vec{\sigma})$. By Lemma 6.7, $\text{tw}(SD) \geq \text{tw}(\pi) = \Omega(n)$. On the other hand, SD has $(2t + 1)n$ vertices, and by Proposition 6.8, it can be drawn on a surface of genus $O(t)$ with $O(tn)$ edge-crossings. Therefore, by Proposition 6.9, SD has tree-width $O(t\sqrt{n})$. Combining these bounds, we conclude that

$$\text{wst}(\mathcal{C}; n) \geq \text{st}(\mathcal{C}; \pi) = t = \Omega(\sqrt{n}),$$

proving the proposition. \square

To complete the proof of Theorem 6.1, we now turn to upper bounds.

Proposition 6.11. *Any \mathfrak{X} -containing class \mathcal{C} has $\text{wst}(\mathcal{C}; n) = O(\log^2 n)$.*

Proof. It is clearly enough to prove the proposition for a class \mathcal{C} that belongs to \mathfrak{X} . Let us first consider the monotone juxtapositions, starting with $\mathcal{C}_1 = \text{Grid} \left(\begin{smallmatrix} \square \\ \square \end{smallmatrix} \right)$. We may easily observe that $\text{wst}(\mathcal{C}_1; n) = O(\log n)$, e.g., by considering the elements of $[n]$ as binary strings of length $O(\log n)$ and sort them using radix-sort, which can be performed by $O(\log n)$ operations of \mathcal{C}_1 .

With the help of Observation 2.3, we conclude that also the classes \mathcal{C}_1^r , \mathcal{C}_1^{-1} , and $(\mathcal{C}_1^r)^{-1}$ have worst-case sorting time $O(\log n)$.

Consider now the class $\mathcal{C}_2 = \text{Grid} \left(\begin{smallmatrix} \square \\ \square \end{smallmatrix} \right)$. Notice that $\mathcal{C}_1 \subseteq \mathcal{C}_2^{\circ 2}$, which implies $\text{wst}(\mathcal{C}_2; n) = O(\log n)$ via Observation 2.2. Again, the same bound applies to all the four symmetries of \mathcal{C}_2 , showing that all the eight monotone juxtapositions have worst-case sorting time $O(\log n)$.

We now deal with the four classes \mathcal{L} , \mathcal{L}^r , \mathcal{PBT} and \mathcal{PBT}^r . Noting that $\mathcal{PBT} \subseteq \mathcal{L}^{\circ 2}$, it is enough to prove that $\text{wst}(\mathcal{PBT}; n) = O(\log^2 n)$, and invoke Observations 2.2 and 2.3 to see that the same upper bound applies to the other three classes as well.

Composing a sequence π with a permutation $\sigma \in \mathcal{PBT}$ corresponds to partitioning π into an even number of blocks $B_1 B_2 B_3 B_4 \cdots B_{2k-1} B_{2k}$, each block being a (possibly empty) subsequence of consecutive entries of π , and then exchanging the position of B_{2j-1} and B_{2j} for each $j \in [k]$. For instance, the permutation $\pi = (3, 8, 1, 6, 2, 5, 9, 4, 7) \in \mathcal{S}_9$ can be partitioned as $B_1 = (3, 8)$, $B_2 = (1)$, $B_3 = \emptyset$, $B_4 = (6, 2)$, $B_5 = (5, 9, 4)$ and $B_6 = (7)$, which after exchanging adjacent odd and even blocks yields the permutation $B_2 B_1 B_4 B_3 B_6 B_5 = (1, 3, 8, 6, 2, 7, 5, 9, 4)$, which is equal to $\sigma \circ \pi$ for $\sigma = (3, 1, 2, 4, 5, 9, 6, 7, 8) \in \mathcal{PBT}_9$.

We now describe a strategy to sort an arbitrary $\pi \in \mathcal{S}_n$ by $O(\log^2 n)$ steps of the form described above. We say that an entry $\pi(i)$ is *small* if $\pi(i) \leq n/2$, otherwise it is *large*. We will use a divide-and-conquer approach, where in the initial phase, which will take $O(\log n)$ steps, we move all the small entries to the left of all the large ones, and then proceed recursively to separately and concurrently sort the small entries and the large ones.

In the initial phase, in every step we first partition the current input sequence π' into small runs and large runs, where a *small run* is a maximal sequence of consecutive small values, and a *large run* is defined similarly. Suppose π' starts with a small element, the other case being analogous. It is then partitioned as $S_1 L_1 S_2 L_2 \cdots S_k L_k$, where each S_j is a small run and L_j a large run, with L_k possibly empty. Then in a single sorting step, for every even $j \in [k]$ we exchange the positions

of the L_{j-1} and S_j , leaving all the other runs unaffected. In this manner, S_j becomes adjacent to S_{j-1} and L_j (if it exists) to L_{j-1} . Thus, every small and large run will merge with another run of the same type, except perhaps the rightmost small run and the rightmost large run. In $O(\log n)$ steps, we reach a permutation with only a single small run and a single large run, which can be swapped to ensure that the small run is to the left of the large one.

We then recursively apply the same strategy to concurrently sort the $\lfloor n/2 \rfloor$ small elements and the $\lfloor n/2 \rfloor$ large ones. In $O(\log^2 n)$ steps, the entire sequence is sorted. \square

Propositions 6.10 and 6.11 together prove Theorem 6.1.

We remark that sorting by layered permutations corresponds to sorting by a sequence of pop-stacks in “genuine series”, which has been considered (for two pop-stacks) by Atkinson and Stitt [5].

We also note that sorting by layered permutations is related to the setting of sorting by length-weighted reversals with linear weights, considered by Bender et al. [8]. More precisely, a single step of sorting by layered permutations can be simulated by a sequence of length-weighted reversals of total cost $O(n)$ in the linear-weight setting of [8], and in particular, an improvement to the $O(\log^2 n)$ upper bound for $\text{wst}(\mathcal{L}; n)$ would also improve the bounds in [8].

7 From $\Omega(\log n)$ to 1

A simple counting argument, combined with the famous Marcus–Tardos theorem [26] shows that any proper permutation class \mathcal{C} , i.e., a class \mathcal{C} not containing all permutations, has at least logarithmic worst-case sorting time.

Proposition 7.1. *If \mathcal{C} is a permutation class that does not contain all permutations, then $\text{wst}(\mathcal{C}; n) = \Omega(\log n)$.*

Proof. By the Marcus–Tardos theorem, there is a constant $c > 0$ such that $|\mathcal{C}_n| \leq c^n$ for all n . Consequently, $|\mathcal{C}_n^{\circ k}| \leq c^{kn}$ for any $k \geq 1$. Taking $k := \text{wst}(\mathcal{C}; n)$ yields

$$n! = |S_n| = \left| \bigcup_{m=1}^k \mathcal{C}_n^{\circ m} \right| \leq \sum_{m=1}^k |\mathcal{C}_n^{\circ m}| \leq k c^{kn},$$

which implies $k = \Omega(\log n)$. \square

Obviously, if \mathcal{C} is the class of all permutations, then $\text{wst}(\mathcal{C}; n) = 1$ for all $n \geq 2$. This completes the proof of Theorem 1.1.

8 Conclusion and open problems

The two main open problems concern the two levels of our hierarchy where we do not have matching upper and lower bounds. Recall that the set \mathfrak{X} contains the eight monotone juxtapositions, the class \mathcal{L} of layered permutations, its reverse \mathcal{L}^r , the class \mathcal{PBT} of parallel block transpositions, and its reversal \mathcal{PBT}^r . We know that

for any monotone juxtaposition, the worst-case sorting time is $\Theta(\log n)$, but for \mathcal{L} , \mathcal{PBT} and their symmetries, we only have a lower bound of order $\Omega(\log n)$ and an upper bound of order $O(\log^2 n)$.

Note that Observations 2.2 and 2.3 imply that $\text{wst}(\mathcal{L}; n) = \Theta(\text{wst}(\mathcal{L}^r; n))$, $\text{wst}(\mathcal{PBT}; n) = \Theta(\text{wst}(\mathcal{PBT}^r; n))$, and $\text{wst}(\mathcal{L}; n) = O(\text{wst}(\mathcal{PBT}; n))$. We can therefore restrict our attention to \mathcal{L} and \mathcal{PBT} .

Problem 8.1. *What is the worst-case sorting time of the class \mathcal{L} of layered permutations, and of the class \mathcal{PBT} of parallel block transpositions?*

Another gap in our bounds concerns the broad family of \mathfrak{X} -avoiding classes. We proved that any \mathfrak{X} -avoiding class has worst-case sorting time of order $\Omega(\sqrt{n})$. However, every example of such a class that we know of has worst-case sorting time of order $\Omega(n)$, suggesting that the lower bound can be further improved.

Problem 8.2. *Is there an \mathfrak{X} -avoiding class \mathcal{C} with $\text{wst}(\mathcal{C}; n) = o(n)$?*

References

- [1] S. Ahal and Y. Rabinovich. On complexity of the subpattern problem. *SIAM Journal on Discrete Mathematics*, 22(2):629–649, 2008. doi:10.1137/S0895480104444776.
- [2] M. H. Albert, M. D. Atkinson, M. Bouvel, N. Ruškuc, and V. Vatter. Geometric grid classes of permutations. *Trans. Amer. Math. Soc.*, 365(11):5859–5881, 2013. doi:10.1090/S0002-9947-2013-05804-7.
- [3] M. H. Albert, C. Homberger, J. Pantone, N. Shar, and V. Vatter. Generating permutations with restricted containers. *Journal of Combinatorial Theory, Series A*, 157:205–232, 2018. doi:10.1016/j.jcta.2018.02.006.
- [4] M. D. Atkinson and R. Beals. Permutation involvement and groups. *Q. J. Math.*, 52(4):415–421, 2001. doi:10.1093/qjmath/52.4.415.
- [5] M. D. Atkinson and T. Stitt. Restricted permutations and the wreath product. *Discrete Math.*, 259(1-3):19–36, 2002. doi:10.1016/S0012-365X(02)00443-0.
- [6] L. Babai and Ákos Seress. On the diameter of Cayley graphs of the symmetric group. *Journal of Combinatorial Theory, Series A*, 49(1):175–179, 1988. doi:10.1016/0097-3165(88)90033-7.
- [7] V. Bafna and P. A. Pevzner. Genome rearrangements and sorting by reversals. In *34th Annual Symposium on Foundations of Computer Science (Palo Alto, CA, 1993)*, pages 148–157. IEEE Comput. Soc. Press, Los Alamitos, CA, 1993. doi:10.1109/SFCS.1993.366872.
- [8] M. A. Bender, D. Ge, S. He, H. Hu, R. Y. Pinter, S. Skiena, and F. Swidan. Improved bounds on sorting by length-weighted reversals. *J. Comput. System Sci.*, 74(5):744–774, 2008. doi:10.1016/j.jcss.2007.08.008.

- [9] S. Bhatia, P. Feijão, and A. R. Francis. Position and content paradigms in genome rearrangements: the wild and crazy world of permutations in genomics. *Bull. Math. Biol.*, 80(12):3227–3246, 2018. doi:10.1007/s11538-018-0514-3.
- [10] M. Bóna. A survey of stack-sorting disciplines. *Electron. J. Combin.*, 9(2):Article 1, 16, 2002/03. Permutation patterns (Otago, 2003). doi:10.37236/1693.
- [11] L. Bulteau, G. Fertin, and I. Rusu. Pancake flipping is hard. *J. Comput. System Sci.*, 81(8):1556–1574, 2015. doi:10.1016/j.jcss.2015.02.003.
- [12] G. Cerbai, A. Claesson, and L. Ferrari. Stack sorting with restricted stacks. *J. Combin. Theory Ser. A*, 173:105230, 19, 2020. doi:10.1016/j.jcta.2020.105230.
- [13] G. Cerbai, A. Claesson, L. Ferrari, and E. Steingrímsson. Sorting with pattern-avoiding stacks: the 132-machine. *Electron. J. Combin.*, 27(3):Paper No. 3.32, 27, 2020. doi:10.37236/9642.
- [14] G. Cerbai and L. Ferrari. Permutation patterns in genome rearrangement problems: The reversal model. *Discrete Applied Mathematics*, 279:34–48, 2020. doi:10.1016/j.dam.2019.10.012.
- [15] V. Dujmović, D. Eppstein, and D. R. Wood. Structure of graphs with locally restricted crossings. *SIAM J. Discrete Math.*, 31(2):805–824, 2017. doi:10.1137/16M1062879.
- [16] Z. Dvořák and S. Norin. Treewidth of graphs with balanced separations. *Journal of Combinatorial Theory, Series B*, 137:137–144, 2019. doi:10.1016/j.jctb.2018.12.007.
- [17] A. Elvey Price and A. J. Guttmann. Permutations sortable by two stacks in series. *Advances in Applied Mathematics*, 83:81–96, 2017. doi:10.1016/j.aam.2016.09.003.
- [18] G. Fertin, A. Labarre, I. Rusu, E. Tannier, and S. Vialette. *Combinatorics of genome rearrangements*. Computational Molecular Biology. MIT Press, Cambridge, MA, 2009. doi:10.7551/mitpress/9780262062824.001.0001.
- [19] W. H. Gates and C. H. Papadimitriou. Bounds for sorting by prefix reversal. *Discrete Math.*, 27(1):47–57, 1979. doi:10.1016/0012-365X(79)90068-2.
- [20] J. R. Gilbert, J. P. Hutchinson, and R. E. Tarjan. A separator theorem for graphs of bounded genus. *Journal of Algorithms*, 5(3):391–407, 1984. doi:10.1016/0196-6774(84)90019-1.
- [21] N. Habermann. Parallel neighbor-sort (or the glory of the induction principle). *Technical Report*, 1972.
- [22] C. Homberger and V. Vatter. On the effective and automatic enumeration of polynomial permutation classes. *J. Symbolic Comput.*, 76:84–96, 2016. doi:10.1016/j.jsc.2015.11.019.

- [23] S. Huczynska and V. Vatter. Grid classes and the Fibonacci dichotomy for restricted permutations. *Electron. J. Combin.*, 13(1):Research Paper 54, 14, 2006. doi:10.37236/1080.
- [24] J. Kececioglu and D. Sankoff. Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. *Algorithmica*, 13(1-2):180–210, 1995. doi:10.1007/BF01188586.
- [25] D. E. Knuth. *The Art of Computer Programming, Volume I: Fundamental Algorithms*. Addison-Wesley, 1968.
- [26] A. Marcus and G. Tardos. Excluded permutation matrices and the Stanley-Wilf conjecture. *J. Combin. Theory Ser. A*, 107(1):153–160, 2004. doi:10.1016/j.jcta.2004.04.002.
- [27] R. Tarjan. Sorting using networks of queues and stacks. *J. Assoc. Comput. Mach.*, 19:341–346, 1972. doi:10.1145/321694.321704.
- [28] V. Vatter. Small permutation classes. *Proc. Lond. Math. Soc. (3)*, 103(5):879–921, 2011. doi:10.1112/plms/pdr017.
- [29] V. Vatter. An Erdős-Hajnal analogue for permutation classes. *Discrete Math. Theor. Comput. Sci.*, 18(2):Paper No. 4, 5, 2016. doi:10.46298/dmtcs.1328.