

Counting 4-Patterns in Permutations Is Equivalent to Counting 4-Cycles in Graphs

Bartłomiej Dudek 

Institute of Computer Science, University of Wrocław, Poland
bartlomiej.dudek@cs.uni.wroc.pl

Paweł Gawrychowski 

Institute of Computer Science, University of Wrocław, Poland
gawry@cs.uni.wroc.pl

Abstract

Permutation σ appears in permutation π if there exists a subsequence of π that is order-isomorphic to σ . The natural algorithmic question is to check if σ appears in π , and if so count the number of occurrences. Only since very recently we know that for any fixed length k , we can check if a given pattern of length k appears in a permutation of length n in time linear in n , but being able to count all such occurrences in $f(k) \cdot n^{o(k/\log k)}$ time would refute the exponential time hypothesis (ETH). Together with practical applications in statistics, this motivates a systematic study of the complexity of counting occurrences for different patterns of fixed small length k . We investigate this question for $k = 4$. Very recently, Even-Zohar and Leng [arXiv 2019] identified two types of 4-patterns. For the first type they designed an $\tilde{O}(n)$ time algorithm¹, while for the second they were able to provide an $\tilde{O}(n^{1.5})$ time algorithm. This brings up the question whether the permutations of the second type are inherently harder than the first type.

We establish a connection between counting 4-patterns of the second type and counting 4-cycles (not necessarily induced) in a sparse undirected graph. By designing two-way reductions we show that the complexities of both problems are the same, up to polylogarithmic factors. This allows us to leverage the work done on the latter to provide a reasonable argument for why there is a difference in the complexities for counting 4-patterns of the first and the second type. In particular, even for the seemingly simpler problem of detecting a 4-cycle in a graph on m edges, the best known algorithm works in $\mathcal{O}(m^{4/3})$ time. Our reductions imply that an $\mathcal{O}(n^{4/3-\epsilon})$ time algorithm for counting occurrences of any 4-pattern of the second type in a permutation of length n would imply an exciting breakthrough for counting (and hence also detecting) 4-cycles. In the other direction, by plugging in the fastest known algorithm for counting 4-cycles, we obtain an algorithm for counting occurrences of any 4-pattern of the second type in $\mathcal{O}(n^{1.48})$ time.

2012 ACM Subject Classification Theory of computation → Algorithm design techniques

Keywords and phrases Permutations, pattern avoidance, counting cycles

Digital Object Identifier 10.4230/LIPIcs.ISAAC.2020.23

Funding *Bartłomiej Dudek*: partially supported by the National Science Centre, Poland, grant number 2017/27/N/ST6/02719.

1 Introduction

Permutations are arguably the most basic combinatorial objects. A natural question in discrete mathematics is to count permutations with certain properties, like consisting of a given number of cycles or having no fixed points. A whole class of such questions is obtained by fixing a permutation σ , called the pattern, and defining a permutation π to avoid σ if σ is not a sub-permutation of π , or in other words if π does not contain a subsequence that is

¹ $\tilde{O}(\cdot)$ hides factors polylogarithmic in n .



© Bartłomiej Dudek and Paweł Gawrychowski;
licensed under Creative Commons License CC-BY

31st International Symposium on Algorithms and Computation (ISAAC 2020).

Editors: Yixin Cao, Siu-Wing Cheng, and Minming Li; Article No. 23; pp. 23:1–23:18

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

order-isomorphic to σ . For example, 21 is avoided only by $12 \dots n$. Otherwise, we say that π contains σ . One of the first results concerning pattern avoidance is by Erdős and Szekeres [23], who proved that every permutation of at least $(k-1)(\ell-1)+1$ elements contains either $12 \dots k$ or $\ell \dots 21$. Another classical result in pattern avoidance is due to Knuth [37], who showed that π can be sorted by a stack if and only if π avoids 231. Together with the systematic study of patterns in permutations by Simion and Schmidt [40], this sparked an interest in counting and characterising permutations that avoid a given pattern (or multiple patterns). A remarkable result in this area is by Marcus and Tardos [39], who showed that the number of permutations of length n avoiding σ is bounded by $c(\sigma)^n$, where $c(\sigma)$ is a function independent of n . This was conjectured in early 1990s independently by Stanley and Wilf. For further discussion we refer the reader to surveys and textbooks [41, 10, 36].

We approach pattern avoidance from an algorithmic perspective. We cannot hope for an efficient algorithm for arbitrary patterns, as in general it is NP-hard to check if π contains σ [12] when σ is part of the input. However, if we restrict our attention to patterns of length k , we might hope to check if a given permutation on n elements avoids such pattern faster than using the trivial algorithm in $\mathcal{O}(n^k)$ time. Indeed, Albert et al. [3] and Ahal and Rabinovich [2] improved this complexity to $\mathcal{O}(n^{2k/3+1})$ and $n^{0.47k+o(k)}$, respectively. In a recent breakthrough result, Guillemot and Marx [27] developed a fixed-parameter tractable (FPT) algorithm that runs in $2^{\mathcal{O}(k^2 \log k)} \cdot n$ time. Later, by refining the proof of Marcus and Tardos [39], Fox [25] removed the $\log k$ factor in the exponent to arrive at $2^{\mathcal{O}(k^2)} \cdot n$ complexity. For $k \geq n/\log n$, $\mathcal{O}(1.79^n)$ and $\mathcal{O}(1.618^n)$ time algorithms are known [16, 8]. Hence even though the problem is NP-hard, by now we have a range of efficient algorithms for different special cases of checking pattern avoidance.

However, some applications bring the need to not only detect but also count occurrences of the pattern. A basic example is calculating the so-called Kendall's τ correlation coefficient [35], which requires counting inversions. Generalizations of Kendall's test used in statistics require counting occurrences of larger patterns. Bergsma-Dassios [9] and Yanagimoto [48] used patterns of length 4 in their tests. Finally, patterns of length 5 appear in the Hoeffding's dependence coefficient [31]. Also see Heller et al. [29] for a general family of such tests. We refer the reader to [24] for a more detailed description of the viewpoint of permutations in nonparametric statistics of bivariate data. Unfortunately, hardly any of the aforementioned algorithms for detecting patterns generalize to counting. A recent result by Berendsohn et al. [8] shows that this is, in fact, inevitable, as if patterns of length k can be counted in $f(k)n^{o(k/\log k)}$ time then the exponential-time hypothesis fails. This shows that we cannot hope for a general FPT algorithm, and considering the applications in statistics we should focus on understanding the best possible exponent for small values of k .

Patterns of length k can be trivially counted in $\mathcal{O}(n^k)$ time, which was improved by Albert et al. [3] to $\mathcal{O}(n^{2k/3+1})$ and then by Berendsohn et al. [8] to $\mathcal{O}(n^{k/4+o(k)})$ time. However, it is clear that among all patterns of the same length k some are easier to count than the others. For example, occurrences of $12 \dots k$ can be easily counted in $\tilde{\mathcal{O}}(nk)$ time using dynamic programming and range queries. This motivates a systematic study of the complexity of counting occurrences of different patterns of fixed small length. For $k=2$, this is exactly the well-known exercise of counting inversions (or in other words, the pattern 21) in a permutation (or its reverse), which can be solved in $\mathcal{O}(n \log n)$ time with merge sort or in $\mathcal{O}(n\sqrt{\log n})$ in the Word RAM model [17]. For $k=3$, all patterns can be counted in $\tilde{\mathcal{O}}(n)$ time by using appropriate range counting structures. For $k=4$, various algorithms were designed to compute efficiently the Bergsma-Dassios test, which asks about the value $\tau^* = (\#1234 + \#1243 + \#2134 + \#2143 + \#3412 + \#3421 + \#4312 + \#4321) / \binom{n}{4} - \frac{1}{3}$ [9]. First approaches

brought the complexity down to $\mathcal{O}(n^2)$ [43, 42, 30] and finally, very recently, Even-Zohar and Leng [24] observed that the patterns counted in this test possess some structural property that allows to design an $\tilde{\mathcal{O}}(n)$ time algorithm. For the remaining patterns of size 4, they obtained an algorithm working in $\tilde{\mathcal{O}}(n^{1.5})$ time. Defining the k -profile of a permutation π to be the sequence of $k!$ numbers with the number of occurrences for every possible pattern σ of length k , this brings us to the following natural open question:

► **Question 1** (Even-Zohar and Leng [24]). *What is the computational complexity of finding the full 4-profile of a given permutation of length n ?*

In fact, Even-Zohar and Leng [24] showed that among all the twenty-four 4-patterns, there are eight that can be counted in $\tilde{\mathcal{O}}(n)$ time, while the remaining ones can be counted in $\tilde{\mathcal{O}}(n^{1.5})$ time. Additionally, they showed that all patterns of the second type are equivalent in terms of computational complexity, that is after counting one of them, we can retrieve all the other in $\tilde{\mathcal{O}}(n)$ time. These two types in fact coincide with the notion of concordant and discordant patterns as defined by Bergsma and Dassios [9]. Using the notation of Fox [25], the permutation matrix of patterns of the second type contains J_2 as an interval minor. This raises the challenge of finding a reason why some 4-patterns seem harder to count than the others.

► **Question 2.** *Why some 4-patterns seem more difficult to count than the others?*

Related work. Many efforts have been devoted to understand which patterns are more difficult to detect [4, 12, 3, 32, 49, 28]. Recently Jelínek and Kynčl [34] established that it is possible to detect σ in polynomial time if σ avoids α , for $\alpha \in \{1, 12, 21, 132, 213, 231, 312\}$ and NP-complete otherwise. This was later strengthened by Berendsohn et al. [8] by considering treewidth of the incidence graph of σ . Even though the problem is NP-hard in general, more efficient algorithms are known for many families of patterns, such as vincular [7], bivincular [13], mesh [14], boxed mesh [6] and consecutive [22]. See the survey by Bruner and Lackner [15] for a more detailed description of these variants.

Fine-grained complexity. Although the traditional notion of “easy” and “hard” problems is defined with respect to the polynomial time solvability, in the last two decades commenced the study on “fine-grained” theory which tries to understand relationships between polynomial-time solvable problems. They can be employed to state conditional lower bounds based on one of a few believable conjectures on complexities of some basic problems, such as SETH, APSP, or 3SUM. See a recent survey by Vassilevska Williams [45] for a summary.

Counting short cycles in graphs. Similarly as for permutation patterns, a natural question is to detect or count small substructures of a graph, with perhaps the most fundamental example being counting cycles of particular length. Already the smallest case, triangle, is highly non-trivial to count, as the fastest known approach for a n -node graph runs in $\mathcal{O}(n^\omega) = \mathcal{O}(n^{2.38})$ using fast matrix multiplication algorithm [26, 44].

Surprisingly, Vassilevska Williams and Williams [47] proved that this is essentially inevitable, as the two problems are, in a certain sense, equivalent: a practical advance for detecting triangles would imply a practical algorithm for Boolean matrix multiplication. As in many applications the graphs are sparse, it is desirable to design algorithm with running time depending on the number of edges m . Alon et al. [5] developed an $\mathcal{O}(m^{2\omega/(\omega+1)}) = \mathcal{O}(m^{1.41})$ time algorithm for counting triangles (in fact their algorithm is stated for finding a single triangle, but can be easily extended). Going one step further, 4-cycles can also be counted in

$\mathcal{O}(n^\omega)$ time [5], but the fastest known counting algorithm for sparse graphs runs in $\mathcal{O}(m^{1.48})$ time [46]. Interestingly, one can *find* a $2k$ -cycle, for any constant $k \geq 2$, in $\mathcal{O}(n^2)$ time [50]. If the graph is given as an adjacency matrix, this is clearly optimal, but it seems plausible to conjecture that this is also the case if the graph is given as adjacency lists.

► **Conjecture 1** (Yuster and Zwick [50]). *For every $\varepsilon > 0$, there is no algorithm that detects 4-cycles in a graph on n nodes in $\mathcal{O}(n^{2-\varepsilon})$ time.*

The best known algorithm for finding a 4-cycle in a sparse graph runs in $\mathcal{O}(m^{4/3})$ time [5]. This was recently extended by Dahlgaard et al. [19] who showed how to find a $2k$ -cycle in $\mathcal{O}(m^{2k/(k+1)})$ time. Furthermore, they showed that this is in fact optimal, assuming Conjecture 1 and using a general combinatorial result of Bondy and Simonovits that a graph with $m = 100kn^{1+1/k}$ edges must contain a $2k$ -cycle [11]. See also Abboud and Vassilevska Williams [1] for a similar conjecture on the complexity of detecting a 3-cycle.

► **Conjecture 2** (Dahlgaard, Knudsen and Stöckel [19]). *For every $\varepsilon > 0$, there is no algorithm that detects a 4-cycle in a graph with m edges in $\mathcal{O}(m^{4/3-\varepsilon})$ time.*

Dudek and Gawrychowski [20] recently used this conjecture to provide an explanation for why there is no $\tilde{\mathcal{O}}(n)$ time algorithm for computing the so-called quartet distance between two trees on n nodes. Very recently Duraj et al. developed an equivalence class between range query problems and detecting triangles in sparse graphs [21].

Our contribution. As in the previous works we divide the patterns into two types and we call them trivial and non-trivial respectively. Our main contribution is a two-way reduction between counting occurrences of a non-trivial pattern and counting 4-cycles in an undirected sparse graph. This provides a reasonable answer for Question 2, as any $\tilde{\mathcal{O}}(n)$ time algorithm for such patterns would imply an exciting breakthrough for counting 4-cycles, and confirms that the two types of 4-patterns identified in the previous work are inherently different.

We partially answer Question 1 about the exact complexity of computing 4-profile of permutation of length n . Our two-way reductions imply that, by plugging in the asymptotically faster known algorithm for counting 4-cycles in a sparse graph [46], we are able to compute the full 4-profile of a permutation of length n in $\mathcal{O}(n^{1.48})$ time. In the other direction, we argue that an $\mathcal{O}(n^{4/3-\varepsilon})$ time algorithm is unlikely, as long as one is willing to believe Conjecture 2.

Our reductions are summarised in Figure 1. A corollary from these reductions is an alternative proof for the equivalence between the non-trivial patterns, which avoids using the notion of corner tree formulas and a computer-aided argument used in [24].

► **Theorem 1.** *An algorithm for counting 4-cycles in a graph on m edges in $\tilde{\mathcal{O}}(m^\gamma)$ time implies an algorithm for counting non-trivial patterns in a permutation of length n in $\tilde{\mathcal{O}}(n^\gamma)$ time and vice versa.*

We can plug in the fastest known algorithm for counting 4-cycles that runs in time $\mathcal{O}(m^{\frac{4\omega-1}{2\omega+1}}) = \mathcal{O}(m^{2-\frac{3}{2\omega+1}})$ [46]. As $\omega < 2.373$ [26, 44], we obtain a more efficient algorithm for computing the full 4-profile in $\mathcal{O}(n^{1.48})$ time.

► **Corollary 2.** *For every $\varepsilon > 0$, there exists no algorithm that can count non-trivial 4-patterns in permutation of length n in $\mathcal{O}(n^{4/3-\varepsilon})$ time unless Conjecture 2 is false.*

We stress that even though we use Conjecture 2 about detecting 4-cycles, the reduction proceeds by creating multiple instances and subtracting some of the obtained result. Hence, it does not imply anything about the complexity of detecting 4-patterns, and in fact for this problem Guillemot and Marx [27] showed an $\mathcal{O}(n)$ time algorithm.

Overview of the methods. Most of our reductions exploit the additional structure of pattern occurrences in the plane which is divided by a horizontal and a vertical line. We group the occurrences by shapes corresponding to the number of points in each quadrant and count them separately. It turns out that the hard case is when the four points are all in distinct quadrants. This is the heart of our main reductions between counting patterns and 4-cycles. All other shapes can be counted in almost linear time with a careful application of range queries. To simplify the presentation, we split the reductions into many steps, between different classes of graphs and patterns so as to work with 4-partite patterns and graphs which have more structure for our application. Our reductions are based on the divide and conquer paradigm, applied to each of the four half-planes separately. We present them using Minimum Base Ranges corresponding to nodes of the full binary tree on n leaves.

Our reduction from counting 4-cycles to counting 4-patterns uses somewhat similar techniques to Berendsohn et al. [8]. However, their approach works for arbitrary subgraphs on k nodes, which comes at a cost of increasing the size of permutation pattern and in our case would result in a pattern of 29 elements. This would not give us the desired connection between counting 4-cycles and 4-patterns, so we need a new argument tailored for 4-cycles.

2 Preliminaries

Permutation π of length n is a bijective mapping $\pi : [n] \rightarrow [n]$, where $[n] = \{1, \dots, n\}$ and a k -pattern σ is a permutation of length k . A permutation π contains a k -pattern σ if there exist indices $1 \leq i_1 < i_2 < \dots < i_k \leq n$ such that $\sigma(j) < \sigma(j')$ iff $\pi(i_j) < \pi(i_{j'})$ for distinct $j, j' \in [k]$. A sequence of k increasing indices with the above properties is called an *occurrence* of σ in π . For example, in permutation 5246173 the underlined positions 4, 5 and 7 form an occurrence of pattern 312. By counting a k -pattern in a permutation we mean counting occurrences of the pattern. Unless stated otherwise, a pattern refers to a 4-pattern.

Shapes. We represent permutation π as a set of points in the plane: $S_\pi = \{(i, \pi(i)) : i \in [n]\}$ and we interchangeably use points and their corresponding elements from the permutation. For instance, four points $\{(i_j, \pi(i_j)) : j \in [4]\}$ are an occurrence of pattern σ iff positions $i_1 < \dots < i_4$ are an occurrence of σ in π . We say that a horizontal line divides a plane into top and bottom part and vertical line divides into left and right part. *Division* of a plane with both horizontal and vertical line splits the points from S_π into four *regions* and we abbreviately denote each of them by capital letters denoting horizontal and vertical location of the region: TL, TR, BL or BR. Slightly abusing the notation, by a region we mean either the region or the set of points from S_π that belong to the region, with the appropriate order

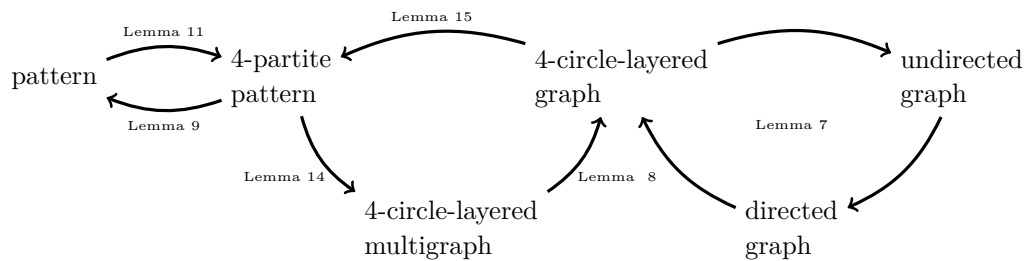


Figure 1 Sequence of reductions used to prove the equivalence between counting non-trivial 4-patterns and 4-cycles. The right part of the figure describes different kinds of graphs in which we count 4-cycles.

between them. Returning to the correspondence between the elements of π and S_π , notice that the division of the plane with horizontal line $y = h$ and vertical line $x = v$ also partitions elements from π into four groups, for instance $(i, \pi(i)) \in TL$ iff $i < v \wedge \pi(i) > h$. We will only consider such divisions of the plane that the dividing lines never pass through a point from S_π .

Given a division of the plane, we say that an occurrence of pattern σ forms *shape* $\frac{a|b}{c|d}$ if among the 4 points, there are respectively a, b, c and d points in top-left, top-right, bottom-left and bottom-right region of the plane. By counting a particular shape for a division we mean counting the number of quadruples of points forming the shape with appropriate number of points in each of the regions. Note that one pattern may form multiple shapes, i.e. $\frac{1|1}{0|2}, \frac{1|2}{0|1}, \frac{2|0}{2|0}$ or $\frac{1|1}{1|1}$, depending on the pattern and the position of the dividing lines. However, some shapes cannot be formed by all patterns, no matter how we divide the plane, i.e. $\frac{1|1}{1|1}$ can be formed by 2314, but not by 2134, and similarly (but the opposite) for $\frac{0|2}{2|0}$. As we can always reflect points in the plane over a dividing line, while discussing a shape we will not mention other shapes obtained by a sequence of such operations, because all such shapes can be counted in exactly the same way. For instance $\frac{3|0}{0|1}, \frac{0|3}{1|0}, \frac{1|0}{0|3}$ and $\frac{0|1}{3|0}$ are all rotations of the same shape, but $\frac{3|0}{1|0}$ is not. To sum up, there are the following possible shapes: $\frac{4|0}{0|0}, \frac{3|0}{0|1}, \frac{3|0}{1|0}, \frac{2|0}{0|2}, \frac{2|0}{2|0}, \frac{1|1}{0|2}, \frac{1|2}{0|1}, \frac{1|1}{1|1}$ and all their rotations. We call shapes $\frac{4|0}{0|0}, \frac{3|1}{0|0}, \frac{2|0}{2|0}$ and their rotations *non-proper*, because the division does not split the pattern both horizontally and vertically. All other shapes are called *proper*. Now we are ready to state the crucial property that distinguishes two main groups of patterns:

► **Definition 3.** A pattern that can form the shape $\frac{1|1}{1|1}$ is called *non-trivial*, and all other patterns are called *trivial*.

Notice that there are 8 trivial patterns: 1234, 1243, 2134, 2143, 4321, 4312, 3421, 3412, all other patterns are *non-trivial*. All trivial patterns can form $\frac{0|2}{2|0}$ (or its reflection $\frac{2|0}{0|2}$), which cannot be formed by non-trivial patterns. For a particular division of the plane, we say that an occurrence of a 4-pattern σ is *4-partite* if all its points belong to pairwise distinct regions, that is they form the shape $\frac{1|1}{1|1}$. To simplify notation, by counting 4-partite pattern σ_4 we mean counting 4-partite occurrences of the pattern σ . Clearly, only non-trivial 4-patterns can be 4-partite. We denote $\#_\sigma(P)$ as the number of occurrences of pattern σ among the points from P . For a 4-partite pattern σ_4 , we slightly abuse the notation and by $\#_{\sigma_4} \left(\frac{TL|TR}{BL|BR} \right)$ we denote the number of 4-partite occurrences of the pattern σ_4 in the plane divided into 4 regions: TL, TR, BL, BR .

MBRs. Let \mathcal{T}_n be a full binary tree with $n' = 2^{\lceil \log n \rceil}$ leaves numbered from 1 to n' and with internal nodes corresponding to the range of indices of leaves from their subtrees. We call the ranges corresponding to the nodes in the tree *base ranges*. Clearly, any number from $[n']$ is contained in $\log n' = \mathcal{O}(\log n)$ base ranges. For a subset $S \subseteq [n]$, we define its *minimum base range* $MBR(S)$ as the smallest base range from \mathcal{T}_n containing all elements from S . Notice that it is the lowest common ancestor (LCA) of all leaves corresponding to the elements from S .

We construct the full binary tree \mathcal{T}_n separately for x - and y -coordinates of points from S_π and consider the Cartesian product $\mathcal{T}_n \times \mathcal{T}_n$ of the trees. For every pair $(R_x, R_y) \in \mathcal{T}_n \times \mathcal{T}_n$ of ranges, let $P_{R_x, R_y} = \{(i, \pi(i)) \in S_\pi : i \in R_x \wedge \pi(i) \in R_y\}$ be the set of points from S_π with their coordinates in appropriate ranges. We call a pair (R_x, R_y) *relevant* if its set P_{R_x, R_y} is non-empty. As every number belongs to $\mathcal{O}(\log n)$ base ranges, every point belongs to $\mathcal{O}(\log^2 n)$ sets P_{R_x, R_y} and hence we have:

► **Observation 4.** *There are $\mathcal{O}(n \log^2 n)$ relevant pairs of ranges.*

General remarks. All the reductions we show in this paper are split into several intermediate steps. Unless stated otherwise, each presented reduction runs in time linear in the total size of the input and the sum of sizes of the created instances of the other problem we reduce to.

2.1 Range Queries and Short Patterns

Some of our algorithms use range queries for counting points in rectilinear (aligned with the x - and the y -axis) rectangles efficiently. Below we provide the precise interface for such queries.

► **Lemma 5** ([18, 33]). *There exists a deterministic data structure that preprocesses a set of n weighted points in $\mathcal{O}(n \log n)$ time and answers queries about the number or the sum of weights of points inside rectilinear rectangles in $\mathcal{O}(\log n)$ time.*

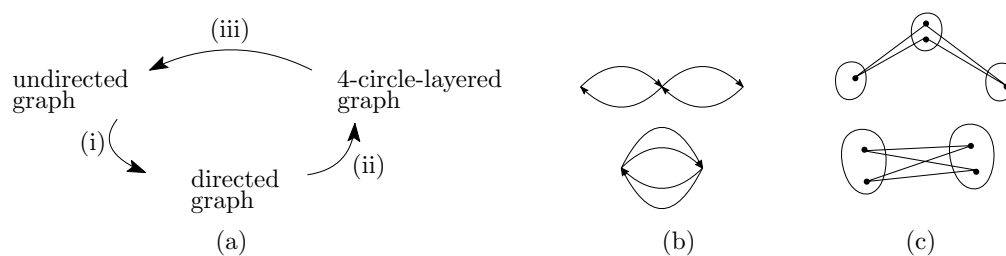
The following theorem about counting patterns with less than 4 elements is folklore.

► **Theorem 6** (cf. [24, Corollary 2]). *For any pattern σ , $|\sigma| < 4$ there exists an algorithm counting σ in permutations of length n in $\tilde{\mathcal{O}}(n)$ time.*

2.2 Counting 4-Cycles

Whenever we talk about counting 4-cycles in a graph we mean simple cycles (with all nodes distinct) of length 4, but not necessarily induced. For counting 4-cycles self-loops and isolated nodes are irrelevant, but there might be multiple edges, and then we count the cycle multiple times: the product of the multiplicities of the relevant edges. Following the naming convention from [38], we define a *4-circle-layered graph* to be a 4-partite directed graph with four disjoint groups of nodes V_0, \dots, V_3 such that every edge in the graph is from the group V_i to $V_{(i+1) \bmod 4}$ for some $0 \leq i \leq 3$.

First, we show that, informally, counting 4-cycles in undirected graphs is equivalent to counting 4-cycles in 4-circle-layered graphs. More precisely, we provide a sequence of reductions for counting 4-cycles in different graphs, starting from undirected graphs, through directed graphs to 4-circle-layered graphs and then back to undirected graphs. We show that counting 4-cycles in a graph of each type can be reduced in $\mathcal{O}(m)$ time to a constant number of instances of counting 4-cycles in graphs of the next type.



■ **Figure 2** (a) Sequence of reductions showing equivalence between counting 4-cycles in undirected, directed and 4-circle-layered graphs. (b) Non-simple cycles from G' to subtract in reduction (ii). (c) Cycles to subtract (top) and add (bottom) in reduction (iii).

► **Lemma 7.** *Counting 4-cycles in undirected graphs on m edges can be reduced to a constant number of instances of counting 4-cycles in 4-circle-layered graphs on $\mathcal{O}(m)$ edges and vice versa.*

Proof. We consider three types of graphs, first undirected graphs, then directed graphs and finally 4-circle-layered graphs. For each of them we show that counting 4-cycles in graphs of this type can be reduced in $\mathcal{O}(m)$ time to a constant number of instances of counting 4-cycles in the graphs of the next type, as presented in Figure 2(a). We describe each of the reductions separately:

- (i) **undirected \rightarrow directed.** Given an undirected graph G we construct a directed graph G' replacing every undirected edge with two directed edges. Then the number of 4-cycles in G' is twice the number of 4-cycles in G , as every cycle can be traversed in both directions. Then we have: $\#_{C_4}(G) = \frac{1}{2}\#_{C_4}(G')$.
- (ii) **directed \rightarrow 4-circle-layered.** Given a directed graph G' we construct a 4-circle-layered graph G'' by copying nodes of G' four times and adding edges between corresponding nodes from two consecutive groups. More precisely, let $v''_i \in V''_i$ in G'' be the copy of node v' from G' in the i -th group. For every directed edge (u', v') in G' we add the edge $(u''_i, v''_{(i+1) \bmod 4})$ to G'' for all $0 \leq i \leq 3$. Then the number of 4-cycles in G'' is 4 times the number of 4-cycles in G' plus some additional cycles which do not correspond to simple cycles in G' . More precisely, all the additional 4-cycles in G'' correspond to non-simple (on 2 or 3 distinct nodes) 4-cycles in G' , which are shown in Figure 2(b) and can be counted in linear time. Formally, let $b(u') = |\{v' \in V' : (u', v') \in E' \wedge (v', u') \in E'\}|$ be the number of neighbors of a node u' connected to v' in both directions, which can be obtained by sorting the adjacency lists in linear time. Then we have: $\#_{C_4}(G') = \frac{1}{4} \left(\#_{C_4}(G'') - \sum_{u' \in V'} \left(4 \binom{b(u')}{2} + b(u') \right) \right)$.
- (iii) **4-circle-layered \rightarrow undirected.** Given a 4-circle-layered graph G'' we create an undirected graph G by undirecting all edges from G'' . Then we can no longer ensure that the 4-cycles pass through 4 different groups of nodes, so we need to subtract 4-cycles fully contained in three groups of nodes and add 4-cycles fully contained in two groups, as shown in Figure 2(c). The number of such cycles can be obtained by counting 4-cycles in the graph G restricted only to the particular groups of nodes. Formally, let V_i be the group of nodes corresponding to V''_i in G'' and $G[W]$ be the subgraph of G restricted to the nodes from W and edges between them. Then we have: $\#_{C_4}(G'') = \#_{C_4}(G) + \sum_{0 \leq i \leq 3} \#_{C_4}(G[V_i \cup V_{i+1}]) - \#_{C_4}(G[V_i \cup V_{i+1} \cup V_{i+2}])$ where the indices $i+1$ and $i+2$ are taken modulo 4. \blacktriangleleft

A multigraph is a triple (V, E, MULT) , where E is a set of m edges and the function $\text{MULT} : E \rightarrow \{1, \dots, U\}$ denotes multiplicity of an edge. For simple graphs it holds that $\text{MULT}(e) = 1$ for all edges $e \in E$ and the function is omitted. Throughout this paper we focus mainly on simple graphs, but in one of the provided reductions we obtain a 4-circle-layered graph with multiplicities on every edge (or in other words, a 4-circle-layered multigraph), so in the following lemma we show how to reduce counting 4-cycles in such graphs to counting 4-cycles in 4-circle-layered simple graphs.

► **Lemma 8.** *Counting 4-cycles in a 4-circle-layered multigraph with edge multiplicities bounded by U can be reduced to $\mathcal{O}(\log^4 U)$ instances of counting 4-cycles in 4-circle-layered simple graphs of the same size as the original graph.*

Proof. Intuitively, we split every edge of the graph into edges with multiplicities being powers of two and iterate over all possible combinations of powers of two forming the cycle.

More precisely, we iterate over all quadruples $(p_0, p_1, p_2, p_3) \in \{0, \dots, \lfloor \log U \rfloor\}^4$ and for each of them create a simple, unweighted 4-circle-layered graph on the same set of nodes as the original graph and a subset of its edges. For all $0 \leq i \leq 3$ we keep only the edges between groups V_i and $V_{(i+1) \bmod 4}$ such that their multiplicity contains 2^{p_i} in its binary

representation. Then we count the number of 4-cycles in the obtained graph and multiply it by $2^{\sum_i p_i}$. Finally, the total number of 4-cycles in the original multigraph is the sum of results obtained for each quadruple. ◀

3 Counting Patterns

In this section we show that counting 4-partite patterns is equivalent, up to logarithmic factors, to counting 4-patterns. The flavor of our arguments is similar to the ones used in [24], but we avoid the notion of corner tree formulas and explicitly state two technical lemmas that are required for our main result. First we show that counting 4-partite patterns can be reduced to counting 4-patterns by omitting the division of the plane and using inclusion-exclusion principle.

► **Lemma 9.** *Counting 4-partite pattern σ_4 on n elements can be reduced to a constant number of instances of counting 4-pattern σ in permutations of total size $\mathcal{O}(n)$.*

Proof. When we omit the division of the plane and count the pattern σ in the plane, we additionally count also the quadruples of points forming the pattern but coming from not all of the 4 regions of the plane. To address this, we use inclusion-exclusion principle and add or subtract patterns on points from all possible subsets of regions. Then the number of 4-partite patterns is:

$$\#_{\sigma_4} \left(\frac{TL|TR}{BL|BR} \right) = \sum_{S \subseteq \{TL, TR, BL, BR\}} (-1)^{|S|} \cdot \#_{\sigma} \left(\bigcup_{Q \in S} Q \right)$$

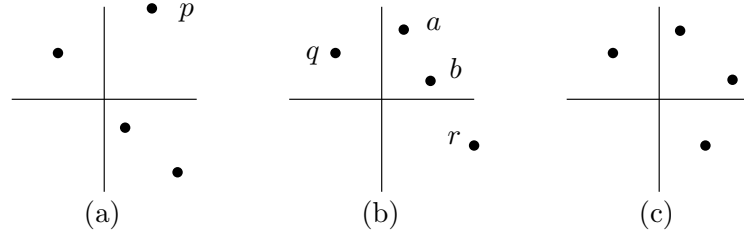
where the union over regions chooses the specific subset of points preserving the relative order between them, as in the original setting. ◀

For the reduction in the other direction, first we need a technical lemma showing that all proper shapes but $\frac{11}{11}$ can be counted in $\tilde{\mathcal{O}}(n)$ time. Recall that we do not have to consider rotations of shapes separately, as they are equivalent under linear-time transformations of the input.

► **Lemma 10.** *For any 4-pattern σ and division of the plane with n points, the shapes $\frac{3|0}{0|1}, \frac{2|0}{0|2}, \frac{1|1}{0|2}, \frac{1|2}{0|1}$ can be counted in $\tilde{\mathcal{O}}(n)$ time.*

Proof. To simplify the presentation, we use the graphical symbols to denote particular regions of the plane: $\frac{\blacksquare|\square}{\square|\square}, \frac{\square|\blacksquare}{\square|\square}, \frac{\square|\square}{\blacksquare|\square}$ and $\frac{\square|\square}{\square|\blacksquare}$ that denote TL, TR, BL and BR respectively. Notice the difference between the notion for 4-partite patterns σ_4 where $\#_{\sigma_4} \left(\frac{TL|TR}{BL|BR} \right) = \#_{\sigma_4} \left(\frac{\blacksquare|\square|\square|\blacksquare}{\square|\square|\square|\square} \right)$ and non-4-partite patterns σ , for which we use division of the plane only to specify the subset of points in which we count patterns, e.g. $\#_{\sigma} \left(\frac{\blacksquare|\square}{\square|\square} \right) = \#_{\sigma}(TL)$. In order to count shapes $\frac{3|0}{0|1}$ and $\frac{2|0}{0|2}$ it suffices to count appropriate 3-, 2- or 1-patterns on points in $\frac{\blacksquare|\square}{\square|\square}$ or $\frac{\square|\square}{\square|\blacksquare}$ and multiply the two numbers. By Theorem 6, this approach runs in $\tilde{\mathcal{O}}(n)$ time.

Now we show how to count the shape $\frac{1|1}{0|2}$. Suppose that in the pattern σ , the two points in $\frac{\square|\square}{\square|\blacksquare}$ form the pattern 21, see Figure 3(a) for an example. For the other case of the pattern 12 we can apply horizontal reflection for points in both the bottom regions. First we preprocess $\frac{\square|\square}{\square|\blacksquare}$ and for every point there we count points from $\frac{\square|\square}{\square|\blacksquare}$ “to the right and down” of it and “to the left and up” using range queries. The precise interface for range queries used in this proof is provided in Lemma 5. Next, we iterate over all points p in $\frac{\square|\blacksquare}{\square|\square}$ and for each of them



■ **Figure 3** (a) A quadruple of points forming $\frac{111}{012}$, where the two bottom points alone form the pattern 21. (b) Naming of points in $\frac{112}{011}$. (c) 3412 is the most difficult pattern to count.

need to count points in $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ and pairs of pairs of points in $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ that together form the pattern σ . The former number is computed with a range query about the number of points from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ that are below or above p , depending on σ . To compute the latter number, notice that the point p can be in three positions with respect to the two points from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$: either to the left of both of them, to the right or in-between (as in Figure 3(a)). Each of the cases can be retrieved by either:

- (a) counting points “to the right and down” for all points from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ to the right of p , or
- (b) counting points “to the left and up” for all points from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ to the left of p , or
- (c) subtracting the values obtained in (a) and (b) from $\#_{21}(\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare})$, the total number of pairs of points from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$, such that one of them is “to the right and down” from the other.

All the above values can be obtained in $\mathcal{O}(\log n) = \tilde{\mathcal{O}}(1)$ time with range queries about the sum of weights of points in a rectangle.

Counting the shape $\frac{112}{011}$ is slightly more involved as now we do not have a single “central” region in which we can iterate over points and obtain the answer, as it was the case with points $p \in \frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ for the shape $\frac{111}{012}$. In order to refer to the points more easily, we use the naming of points as in Figure 3(b), that is q is the point from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$, r from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ and a and b from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$, where a is to the left of b . Again we focus only on the case when points from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ form the pattern 21, that is a is “to the left and up” of b . For the other case of the pattern 12 we can horizontally reflect points in both the top regions. Consider the case when the last element in the pattern σ is the smallest (equals 1), so is the point r , in $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$. Then the allowed location of r depends only on the point b , as r must be to the right of b , so for every point b in $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ we can count points from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ that are to the right of b . Next we proceed similarly as while counting the shape $\frac{111}{012}$, that is we iterate through points q from $\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}$ and count pairs of points a and b in the appropriate order with respect to q , where additionally points b have weights.

The above approach can be also applied to all shapes in which r is to the left of both points a and b , or q is below both a and b , or q is above both a and b . In other words, this covers all patterns in which q is not between a and b or r is not between a and b . Hence it remains to consider the patterns in which both q and r are between a and b . Notice that for the fixed relation between points a and b (21 in our case) there is exactly one such pattern σ : 3412, see Figure 3(c). To sum up, there are 9 possible patterns (3 locations for points r and q are possible independently) forming the considered shape $\frac{112}{011}$ and 8 of them we can count in $\tilde{\mathcal{O}}(n)$ time. Moreover, the sum of counts of all the 9 patterns is exactly $|\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}| \cdot |\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare}| \cdot \#_{21}(\frac{\blacksquare\blacksquare}{\blacksquare\blacksquare})$. Subtracting from the total count the 8 values that we can compute efficiently gives us the number of occurrences of the last pattern. Thus, all patterns forming the shape $\frac{112}{011}$ can be counted in $\tilde{\mathcal{O}}(n)$ time. ◀

Recall that, given a division of the plane into 4 regions, an occurrence of a 4-pattern σ is 4-partite if all its elements are in pairwise distinct regions. In the following lemma we show that we can count 4-patterns by counting 4-partite patterns. At a high level, every occurrence of the pattern is counted while considering the division of the plane aligned with the division of minimum base ranges containing all coordinates of the four points.

► **Lemma 11.** *Counting a 4-pattern σ on n elements can be reduced in $\tilde{O}(n)$ time to multiple instances of counting 4-partite patterns σ_4 of total size $\tilde{O}(n)$.*

Proof. Recall that $\text{MBR}(S)$, the minimum base range of a set $S \subseteq [n]$ is the minimum base range containing all elements of S in the full binary tree \mathcal{T}_n on $n' = 2^{\lceil \log n \rceil}$ leaves and $R \in \mathcal{T}_n$ is a set of consecutive elements from $[n]$. By Observation 4 we have that there are $\tilde{O}(n)$ pairs $(R_x, R_y) \in \mathcal{T}_n \times \mathcal{T}_n$ for which there exists an $i \in [n]$ such that $i \in R_x$ and $\pi(i) \in R_y$. We can retrieve all such pairs in $\tilde{O}(n)$ time by iterating through all points from S_π and generating the set of all relevant pairs of ranges. Recall that $P_{R_x, R_y} = \{(i, \pi(i)) \in S_\pi : i \in R_x \wedge \pi(i) \in R_y\}$. In terms of the permutation π , R_x corresponds to its substring and R_y restricts its values.

For every relevant pair of ranges (R_x, R_y) with P_{R_x, R_y} of at least 4 points inserted, we consider the plane restricted only to points from P_{R_x, R_y} and divided in the following way. As all points from S_π have distinct coordinates and $|P_{R_x, R_y}| \geq 4$, the range R_x contains at least 4 elements, so is not a leaf in \mathcal{T}_x and has two children R_x^L, R_x^R in \mathcal{T}_x . The two ranges R_x^L and R_x^R are disjoint so we can find a vertical line that separates them, i.e. that passes through the middle of segment between the rightmost element from R_x^L and the leftmost element from R_x^R . Notice that this line does not pass through a point from P_{R_x, R_y} as R_x^L and R_x^R are two consecutive ranges in \mathcal{T}_n . We find a horizontal line separating the range R_y in the same way. For the set of points P_{R_x, R_y} and the above division of the plane, we count all shapes $\frac{3|0}{0|1}, \frac{2|0}{0|2}, \frac{1|1}{0|2}, \frac{1|2}{0|1}$ and all their possible rotations in $\tilde{O}(|P_{R_x, R_y}|)$ time, by Lemma 10. Finally, we need to count the shape $\frac{1|1}{1|1}$, the 4-partite pattern σ_4 on the set P_{R_x, R_y} and sum up all the obtained results.

Now we show that the above procedure counts every occurrence of the pattern σ exactly once, while considering the pair of minimum base ranges for both coordinates of the points from the occurrence. Formally, an occurrence g of σ on positions $i_1 < i_2 < i_3 < i_4$ is counted only for the pair of ranges (R_x, R_y) where $R_x = \text{MBR}(\{i_1, i_2, i_3, i_4\})$ and $R_y = \text{MBR}(\{\pi(i_1), \pi(i_2), \pi(i_3), \pi(i_4)\})$ and the appropriate shape, depending on the position of points from $\{(i_j, \pi(i_j)) : j \in [4]\}$ with respect to the division. Suppose the contrary, that g is counted for another pair of ranges (R'_x, R'_y) where $R'_x \neq R_x$, for $R'_y \neq R_y$ the reasoning is similar. If $\{i_1, i_2, i_3, i_4\} \not\subseteq R'_x$, then for some j the point $(i_j, \pi(i_j))$ will not be present in the considered instance. Otherwise, from the structure of base ranges we have that $\text{MBR}(\{i_1, i_2, i_3, i_4\})$ is fully contained in one half of R'_x . In this case g also will not be counted, because it forms a non-proper shape for the considered division ($\frac{2|0}{2|0}, \frac{3|1}{0|0}$ or $\frac{4|0}{0|0}$ or their rotations) and we do not count such shapes.

As every point from S_π is included in $\mathcal{O}(\log^2 n)$ sets P_{R_x, R_y} , the total size of all the considered sets is $\tilde{O}(n)$ and hence counting shapes different than $\frac{1|1}{1|1}$ takes $\tilde{O}(n)$ time. Similarly, the total size of the instances of counting 4-partite pattern σ_4 is $\tilde{O}(n)$. ◀

By definition, trivial patterns do not form the $\frac{1|1}{1|1}$ shape, so the reduced instances have always 0 occurrences of the 4-partite pattern, which can be returned in constant time. Hence:

► **Corollary 12** (cf. [24, Corollary 3]). *All trivial 4-patterns (1234, 1243, 2134, 2143, 4321, 4312, 3421, 3412) in permutations of length n can be counted in $\tilde{O}(n)$ time.*

4 Equivalence of Counting 4-Partite Patterns and Cycles

First we show that in fact all (non-trivial) 4-partite patterns are equivalent by a linear-time transformation of the considered set of points. At a high level, we will show that reversing the order of points in any of the four parts of the plane (left, top, ...) allows us to slightly modify the pattern.

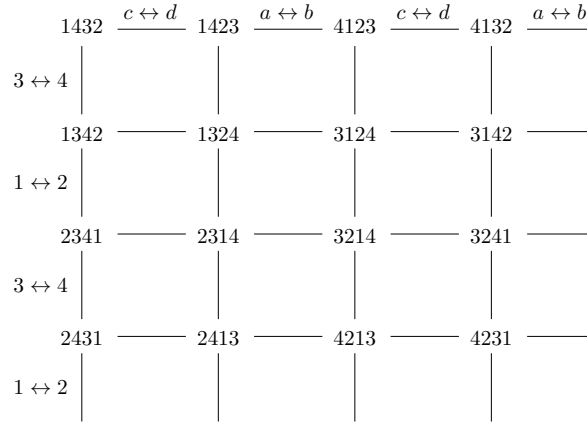
► **Lemma 13.** *Counting any non-trivial 4-partite pattern σ_4 can be reduced to counting any other non-trivial 4-partite pattern σ'_4 .*

Proof. We start with showing that by reversing the points in the left part of the plane we can swap the first two elements of the pattern:

$$\#_{abcd_4} \left(\frac{TL|TR}{BL|BR} \right) = \#_{bacd_4} \left(\text{REV} \left(\frac{TL}{BL} \right) \middle| \frac{TR}{BR} \right).$$

Formally, suppose that we need to count the 4-partite pattern $abcd$ in the plane divided as follows: $\frac{TL|TR}{BL|BR}$ and the leftmost and rightmost points from the left part ($TL \cup BL$) have the x -coordinate respectively x_1 and x_2 . We replace every point (x, y) from the left part with $(x_1 + x_2 - x, y)$. Then, only the horizontal order of points from the left part is reversed and any 4-partite occurrence of the pattern $abcd$ in the original instance corresponds to a 4-partite occurrence of the pattern $bacd$ in the transformed instance. Similarly, after reversing the right part we obtain the pattern $abdc$ from $abcd$. When we reverse the (vertical) order of the top or bottom part, we swap respectively elements 3 and 4 or 1 and 2 in the pattern. For example, by reversing the top part, from the pattern 1324 we obtain the pattern 1423.

Observe that operations in any two parts of the plane are independent, we can apply any subset of them and obtain either of the 16 possible non-trivial 4-partite patterns. See Figure 4 with the precise description of operations between the patterns. Thus, we can transform in linear time any instance of counting non-trivial 4-partite pattern σ_4 to an instance of counting either of the 16 possible non-trivial 4-partite patterns. ◀

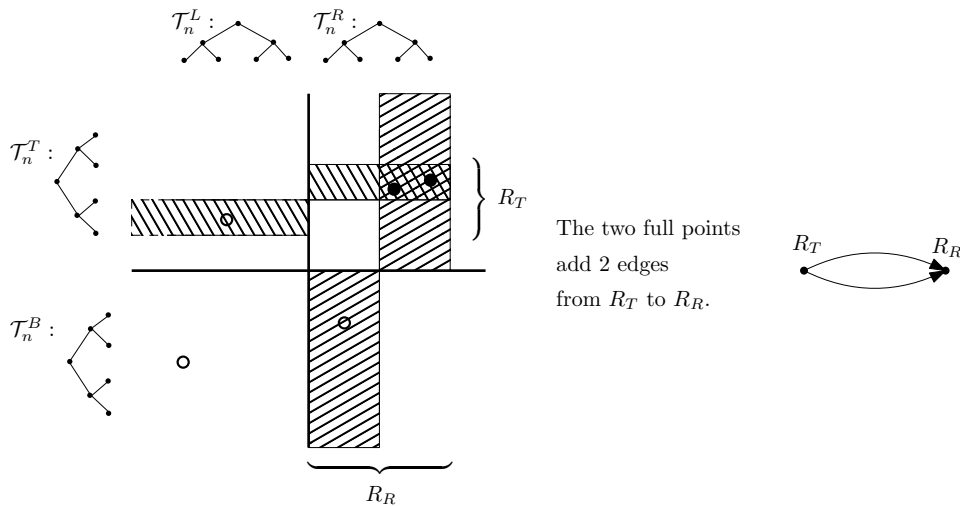


■ **Figure 4** Reductions between non-trivial patterns described in Lemma 13. Operation $a \leftrightarrow b$ ($c \leftrightarrow d$) swaps the first (second) pair of elements in the pattern and corresponds to reversing left (right) part of the plane. Operation $1 \leftrightarrow 2$ ($3 \leftrightarrow 4$) swaps elements 1 and 2 (3 and 4) in the pattern and corresponds to reversing bottom (top) part of the plane.

Hence in the following claims it suffices to consider only one non-trivial 4-partite pattern and we will focus on counting the pattern 1324₄. Notice that in $\tilde{O}(n)$ time we can shift any set of n points in such a way that the division lines are aligned with x - and y - axes and all points have integer coordinates from $\mathcal{N} = \{-n, \dots, -1, 1, \dots, n\}$, preserving the relative order between the parts. In the following lemma we show that counting non-trivial 4-partite patterns can be reduced to counting 4-cycles in 4-circle-layered multigraphs. At a high level, we will group all occurrences of the pattern by the minimum base ranges of coordinates of points in each of the parts of the plane.

► **Lemma 14.** *Counting a non-trivial 4-partite pattern on n points can be reduced to an instance of counting 4-cycles in a 4-circle-layered multigraph on $\tilde{O}(n)$ edges with multiplicities bounded by n .*

Proof. For a permutation π and division of the plane with points S_π we need to construct a 4-circle-layered multigraph in such a way that the number of 4-cycles in the graph gives us the number of occurrences of the pattern. Recall that we can operate on points from \mathcal{N}^2 and the division of the plane along the x - and y -axes. We consider four full binary trees $\mathcal{T}_n^L, \mathcal{T}_n^R, \mathcal{T}_n^B, \mathcal{T}_n^T$ for each part of the plane separately. For each base range in the trees we create a separate node in the new 4-partite graph.



■ **Figure 5** We consider four full binary trees $\mathcal{T}_n^L, \mathcal{T}_n^R, \mathcal{T}_n^B, \mathcal{T}_n^T$ for each part of the plane separately and group occurrences of patterns by the MBRs of coordinates in each part of the plane. Points from appropriate halves of MBRs from each two consecutive parts add a new edge to the multigraph.

Now we process all points from S_π grouped by their region. Suppose we process a point $(x, y) \in S_\pi$ from the top-right region. We iterate over all pairs $(R_R, R_T) \in \mathcal{T}_n^R \times \mathcal{T}_n^T$ of base ranges such that $x \in R_R$ and $y \in R_T$ and the ranges are not singletons (leaves in \mathcal{T}_n), so contain at least two elements from $[n]$. Recall that we focus on the pattern 1324, because now the choice of the particular pattern is crucial in the following condition. We add edge (R_T, R_R) to the 4-circle-layered multigraph if x is in the right half of R_R and y is in the top half of R_T . This means that the point (x, y) can be a part of an occurrence of the 1324 pattern in which R_T is the MBR of y -coordinates of the top points and R_R is the MBR of x -coordinates of the right points. See Figure 5. We proceed similarly for the remaining three regions, modifying only the condition for including an edge, based on the position of elements of the pattern 1324 inside the considered region.

If an edge is inserted more than once, we simply increment its multiplicity, which can be stored e.g. in a balanced binary search tree. As every point from S_π adds at most $\mathcal{O}(\log^2 n)$ edges, in total there are $\mathcal{O}(n \log^2 n) = \tilde{\mathcal{O}}(n)$ edges in the graph. Clearly, the constructed directed multigraph is 4-partite as we connect nodes from \mathcal{T}_n^T to the nodes from \mathcal{T}_n^R , from \mathcal{T}_n^R to \mathcal{T}_n^B etc. Finally, observe that the multiplicity of an edge connecting nodes corresponding to ranges R and R' is the number of points in the intersection of their appropriate halves. Hence multiplicities of edges in the graph are bounded by n . \blacktriangleleft

The reduction from 4-circle-layered multigraphs to 4-circle-layered simple graphs was shown in Lemma 8. Finally, to conclude the equivalence between counting 4-partite patterns and cycles in 4-circle-layered graphs, we describe the reduction from counting 4-cycles in 4-circle-layered graphs to counting non-trivial patterns. The idea is to first embed the graph in the plane so that every group V_i of nodes corresponds to a half-plane and edges to points in the plane. Then every 4-cycle corresponds to a rectangle with all corners in distinct quadrants. Now we appropriately tilt each quadrant, so that every rectangle corresponds to an occurrence of the pattern 1324_4 . However, this change introduces many more occurrences of the pattern as now we have slightly weaker constraints on the relative position of points. This is corrected by subtracting the surplus by applying the inclusion-exclusion principle for different ways of tilting the quadrants.

We remark that our approach is similar to that of Berendsohn et al. [8, Section 5]. They showed a reduction from Partitioned Subgraph Isomorphism to counting short patterns in permutations by embedding the input graph in the plane with appropriate tilting and using the inclusion-exclusion principle. However, while their reduction works for arbitrary subgraphs of size k , this comes at the cost of increasing the size of the permutation pattern to $7k + 1$, which in our case would result in a permutation pattern on 29 elements, hence not giving us the desired tight connection between counting 4-cycles and 4-patterns.

► **Lemma 15.** *Counting 4-cycles in a 4-circle-layered simple graph on m edges can be reduced in $\tilde{\mathcal{O}}(m)$ time to a constant number of instances of counting a non-trivial pattern in a permutation of length m .*

Proof. Given a 4-circle-layered graph $G = (V_0 \dot{\cup} V_1 \dot{\cup} V_2 \dot{\cup} V_3, E)$, where $E \subseteq \bigcup_i V_i \times V_{i+1 \bmod 4}$, we will embed it in the plane and construct a constant number of instances of counting a non-trivial 4-partite pattern. As Lemma 13 guarantees that all such patterns are equivalent, we can focus only on the pattern 1324 .

Every half-plane corresponds to a part of the graph in the clockwise order: negative x -coordinates correspond to nodes from V_0 , positive y -coordinates correspond to nodes from V_1 , positive x -coordinates correspond to nodes from V_2 and negative y -coordinates correspond to nodes from V_3 . The order of points in every half-plane projected on the appropriate axis is arbitrary, so we can use any injective mapping from V_0 and V_3 to $\{-n, \dots, -1\}$ and from V_1 and V_2 to $\{1, \dots, n\}$. Next, every edge in the graph corresponds to a point in the plane, so we get a subset of m points from \mathcal{N}^2 . Then every 4-cycle in G corresponds to a rectangle with corners in points in distinct quadrants.

Now we would like to transform the constructed set of points into a number of point sets S_π for some permutations π . Intuitively, every 4-cycle from G will correspond to an occurrence of the pattern 1324_4 . Notice that there might be many edges incident to a node, so in the beginning some points have equal x - or y -coordinate, which we need to avoid. At first we will guarantee that no two points from distinct quadrants have equal x - or y -coordinates, which is already sufficient to be able to define an occurrence of the 4-partite pattern 1324_4 .

In the end we will show that we can slightly shift all points preserving relationships between points from distinct quadrants and additionally ensuring uniqueness of coordinates inside each quadrant. Consider the following transformation of the plane:

$$\frac{TL|TR}{BL|BR} \rightarrow \frac{TL + (\frac{1}{5}, 0)|TR + (0, \frac{1}{5})}{BL + (0, -\frac{1}{5})|BR + (-\frac{1}{5}, 0)}$$

where by adding a vector to the region we denote shifting all points from the region by the vector. Informally, we shift TL slightly right, TR slightly up etc, see Figure 6(a). Observe that now every 4-cycle from G corresponds to an occurrence of 1324_4 (see Figure 6(b) and its explanation in the caption), but there are also many more other occurrences of the pattern, which do not correspond to a cycle from G . More precisely, every occurrence of the pattern 1324_4 corresponds to 4 edges from G , but we cannot ensure that they form a cycle, or equivalently, that every two consecutive edges share an endpoint, see Figure 6(c).

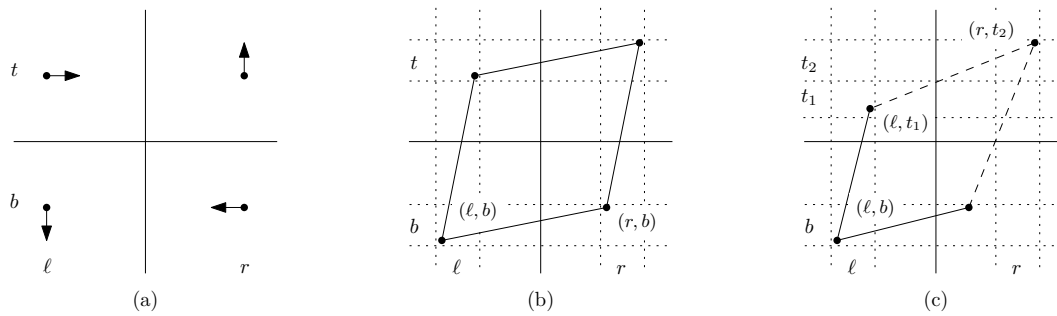


Figure 6 (a) Slightly shifting all points guarantees that points from distinct quadrants do not share a coordinate. (b) Every cycle from the graph corresponds to an occurrence of 1324_4 . We mark the area of the “small shifts” between the dashed lines, so i.e. all points that initially had y -coordinate equal to t now are between the two horizontal dashed lines surrounding t . (c) Some occurrences of 1324_4 do not correspond to a cycle in G , as the consecutive edges do not share endpoints. Points corresponding to consecutive edges that share an endpoint are connected with a solid line (i.e. (ℓ, b) and (ℓ, t_1)) and with a dashed line if they do not share (i.e. (ℓ, t_1) and (r, t_2)).

In particular, after the above transformation, in every occurrence of 1324_4 the two points from the left half-plane: $(x, y - \frac{1}{5}) \in BL$ and $(x' + \frac{1}{5}, y') \in TL$ satisfy that $x' + \frac{1}{5} \geq x$, but the two edges corresponding to these points share an endpoint only when $x' = x$. On the other hand, if we slightly modify the above transformation and set $TL \rightarrow TL + (-\frac{1}{5}, 0)$, we obtain that $x' - \frac{1}{5} \geq x$, so $x' > x$ and certainly the two edges cannot share an endpoint. Now we use this property for all half-planes and plug the modified transformations into the inclusion-exclusion principle:

$$\#_{C_4}(G) = \sum_{S \subseteq \{L, R, T, B\}} (-1)^{|S|} \#_{1324_4} \left(\frac{TL + (\delta_L(S), 0)|TR + (0, \delta_T(S))}{BL + (0, -\delta_B(S))|BR + (-\delta_R(S), 0)} \right)$$

where $\delta_X(S) = \frac{1}{5}$ if $X \in S$ or $-\frac{1}{5}$ otherwise. Finally, to ensure that no two points in a single quadrant have equal x - or y -coordinate we first transform every point (x, y) into $(x + \frac{y}{10n}, y + \frac{x}{10n})$ and then shift accordingly. For instance, a point $(x, y) \in TL$ is transformed to $(x + \frac{y}{10n} + \delta_L(S), y + \frac{x}{10n})$. Notice that the choice of lengths of the shifts guarantees that no two points have the same x - or y -coordinate and the new coordinates are within $[-\frac{3}{10}, \frac{3}{10}] \times [-\frac{3}{10}, \frac{3}{10}]$ square comparing to the original location of points. In the obtained instances all points have non-integer coordinates, but we can normalize them into \mathcal{N}^2 preserving the relative order between the points. ◀

References

- 1 Amir Abboud and Virginia Vassilevska Williams. Popular conjectures imply strong lower bounds for dynamic problems. In *55th FOCS*, pages 434–443. IEEE Computer Society, 2014.
- 2 Shlomo Ahal and Yuri Rabinovich. On complexity of the subpattern problem. *SIAM J. Discrete Math.*, 2:629–649, 2008.
- 3 Michael H. Albert, Robert E. L. Aldred, Mike D. Atkinson, and Derek A. Holton. Algorithms for pattern involvement in permutations. In *12th ISAAC*, Lecture Notes in Computer Science, pages 355–366. Springer, 2001.
- 4 Michael H. Albert, Marie-Louise Lackner, Martin Lackner, and Vincent Vatter. The complexity of pattern matching for 321-avoiding and skew-merged permutations. *Discrete Mathematics & Theoretical Computer Science*, 18(2), 2016.
- 5 Noga Alon, Raphael Yuster, and Uri Zwick. Finding and counting given length cycles. *Algorithmica*, 3:209–223, 1997.
- 6 Sergey V. Avgustinovich, Sergey Kitaev, and Alexandr Valyuzhenich. Avoidance of boxed mesh patterns on permutations. *Discrete Applied Mathematics*, 1-2:43–51, 2013.
- 7 Eric Babson and Einar Steingrímsson. Generalized permutation patterns and a classification of the mahonian statistics. *Séminaire Lotharingien de Combinatoire*, pages B44b, 18 p., 2000.
- 8 Benjamin Aram Berendsohn, László Kozma, and Dániel Marx. Finding and counting permutations via CSPs. In *14th IPEC, LIPIcs*, pages 1:1–1:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
- 9 Wicher Bergsma and Angelos Dassios. A consistent test of independence based on a sign covariance related to Kendall’s tau. *Bernoulli*, 2:1006–1028, 2014.
- 10 Miklós Bóna. *Combinatorics of Permutations, Second Edition*. Discrete mathematics and its applications. CRC Press, 2012.
- 11 John Adrian Bondy and Miklós Simonovits. Cycles of even length in graphs. *Journal of Combinatorial Theory, Series B*, 2:97–105, 1974.
- 12 Prosenjit Bose, Jonathan F. Buss, and Anna Lubiw. Pattern matching for permutations. *Inf. Process. Lett.*, 5:277–283, 1998.
- 13 Mireille Bousquet-Mélou, Anders Claesson, Mark Dukes, and Sergey Kitaev. $(2+2)$ -free posets, ascent sequences and pattern avoiding permutations. *J. Comb. Theory, Ser. A*, 7:884–909, 2010.
- 14 Petter Brändén and Anders Claesson. Mesh patterns and the expansion of permutation statistics as sums of permutation patterns. *Electr. J. Comb.*, 2, 2011.
- 15 Marie-Louise Bruner and Martin Lackner. The computational landscape of permutation patterns. *CoRR*, abs/1301.0340, 2013. [arXiv:1301.0340](#).
- 16 Marie-Louise Bruner and Martin Lackner. A fast algorithm for permutation pattern matching based on alternating runs. *Algorithmica*, 75(1):84–117, 2016.
- 17 Timothy M. Chan and Mihai Pătraşcu. Counting inversions, offline orthogonal range counting, and related problems. In *21st SODA*, pages 161–173. SIAM, 2010.
- 18 Bernard Chazelle. A functional approach to data structures and its use in multidimensional searching. *SIAM J. Comput.*, 17(3):427–462, 1988.
- 19 Søren Dahlgaard, Mathias Bæk Tejs Knudsen, and Morten Stöckel. Finding even cycles faster via capped k -walks. In *49th STOC*, pages 112–120. ACM, 2017.
- 20 Bartłomiej Dudek and Paweł Gawrychowski. Computing quartet distance is equivalent to counting 4-cycles. In *51st STOC*, pages 733–743. ACM, 2019.
- 21 Lech Duraj, Krzysztof Kleiner, Adam Polak, and Virginia Vassilevska Williams. Equivalences between triangle and range query problems. In *30th SODA*, pages 30–47. SIAM, 2020.
- 22 Sergi Elizalde and Marc Noy. Consecutive patterns in permutations. *Adv. Appl. Math.*, 1-2:110–125, 2003.
- 23 Paul Erdős and George Szekeres. A combinatorial problem in geometry. *Compositio Mathematica*, pages 463–470, 1935.

- 24 Chaim Even-Zohar and Calvin Leng. Counting small permutation patterns. *CoRR*, abs/1911.01414, 2019. [arXiv:1911.01414](#).
- 25 Jacob Fox. Stanley-Wilf limits are typically exponential. *CoRR*, abs/1310.8378, 2013. [arXiv:1310.8378](#).
- 26 Francois Le Gall. Powers of tensors and fast matrix multiplication. In *25th ISSAC*, pages 296–303. ACM, 2014.
- 27 Sylvain Guillemot and Dániel Marx. Finding small patterns in permutations in linear time. In *25th SODA*, pages 82–101. SIAM, 2014.
- 28 Sylvain Guillemot and Stéphane Vialette. Pattern matching for 321-avoiding permutations. In *20th ISAAC*, volume 5878 of *Lecture Notes in Computer Science*, pages 1064–1073. Springer, 2009.
- 29 Ruth Heller, Yair Heller, Shachar Kaufman, Barak Brill, and Malka Gorfine. Consistent distribution-free k -sample and independence tests for univariate random variables. *J. Mach. Learn. Res.*, pages 29:1–29:54, 2016.
- 30 Yair Heller and Ruth Heller. Computing the Bergsma Dassios sign-covariance. *CoRR*, abs/1605.08732, 2016. [arXiv:1605.08732](#).
- 31 Wassily Hoeffding. A non-parametric test of independence. *The Annals of Mathematical Statistics*, pages 546–557, 1948.
- 32 Louis Ibarra. Finding pattern matchings for permutations. *Inf. Process. Lett.*, 61(6):293–295, 1997.
- 33 Joseph JáJá, Christian Worm Mortensen, and Qingmin Shi. Space-efficient and fast algorithms for multidimensional dominance reporting and counting. In *15th ISAAC*, volume 3341 of *Lecture Notes in Computer Science*, pages 558–568. Springer, 2004.
- 34 Vít Jelínek and Jan Kynčl. Hardness of permutation pattern matching. In *28th SODA*, pages 378–396. SIAM, 2017.
- 35 Maurice G. Kendall. A new measure of rank correlation. *Biometrika*, pages 81–93, 1938.
- 36 Sergey Kitaev. *Patterns in Permutations and Words*. Monographs in Theoretical Computer Science. An EATCS Series. Springer, 2011.
- 37 Donald E. Knuth. *The Art of Computer Programming, Volume I: Fundamental Algorithms*. Addison-Wesley, 1968.
- 38 Andrea Lincoln, Virginia Vassilevska Williams, and R. Ryan Williams. Tight hardness for shortest cycles and paths in sparse graphs. In *29th SODA*, pages 1236–1252. SIAM, 2018.
- 39 Adam Marcus and Gábor Tardos. Excluded permutation matrices and the Stanley-Wilf conjecture. *J. Comb. Theory, Ser. A*, 1:153–160, 2004.
- 40 Rodica Simion and Frank W. Schmidt. Restricted permutations. *Eur. J. Comb.*, 4:383–406, 1985.
- 41 Vincent Vatter. Permutation classes. In Miklós Bóna, editor, *Handbook of Enumerative Combinatorics*. CRC Press, 2015. Preprint at [arXiv:1409.5159](#).
- 42 Luca Weihs, Mathias Drton, and Dennis Leung. Efficient computation of the Bergsma–Dassios sign covariance. *Comput. Stat.*, 1:315–328, 2016.
- 43 Luca Weihs, Mathias Drton, and Nicolai Meinshausen. Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika*, 3:547–562, 2018.
- 44 Virginia Vassilevska Williams. Multiplying matrices faster than Coppersmith-Winograd. In *44th STOC*, pages 887–898. ACM, 2012.
- 45 Virginia Vassilevska Williams. On some fine-grained questions in algorithms and complexity. In *International Congress of Mathematicians (ICM)*, 2018.
- 46 Virginia Vassilevska Williams, Joshua R. Wang, Richard Ryan Williams, and Huacheng Yu. Finding four-node subgraphs in triangle time. In *26th SODA*, pages 1671–1680. SIAM, 2015.
- 47 Virginia Vassilevska Williams and R. Ryan Williams. Subcubic equivalences between path, matrix, and triangle problems. *J. ACM*, 5:27:1–27:38, 2018.

23:18 Counting 4-Patterns in Permutations Is Equivalent to Counting 4-Cycles in Graphs

- 48 Takemi Yanagimoto. On measures of association and a related problem. *Annals of the Institute of Statistical Mathematics*, 1:57–63, 1970.
- 49 V. Yugandhar and Sanjeev Saxena. Parallel algorithms for separable permutations. *Discrete Applied Mathematics*, 3:343–364, 2005.
- 50 Raphael Yuster and Uri Zwick. Finding even cycles even faster. *SIAM J. Discrete Math.*, 2:209–222, 1997.