Proper n-cell Polycubes in $n-k$ Dimensions[☆]

Gill Barequet[a,*], Mira Shalah[a]

[a] *Center for Graphics and Geometric Computing, Dept. of Computer Science, The Technion—Israel Inst. of Technology, Haifa 3200003, Israel*

Abstract

A d-dimensional polycube of size n is a connected set of n cubes in d dimensions, where connectivity is through $(d-1)$-dimensional faces. In this paper we develop a theoretical framework for computing the explicit formula enumerating polycubes of size n that span $n-k$ dimensions, for a fixed value of k. Besides the interest in the number of these simple combinatorial objects, known as proper polycubes, such formulae play an important role in the literature of statistical physics in the study of percolation processes and collapse of branched polymers. We use this framework to prove the known, yet unproven, conjecture about the general form of the formula for a general k, and implement a computer program which reaffirmed the known formulae for $k=2$ and $k=3$, and proved rigorously, for the first time, the formulae for $k=4$ and $k=5$.

Keywords: Polyominoes, lattice animals, inclusion-exclusion.

1. Introduction

A d-dimensional polycube of size n is a connected set of n cubes in d dimensions, where connectivity is through $(d-1)$-dimensional faces. Two *fixed* polycubes are considered the same if one can be obtained by a translation of the other. We consider here only fixed polycubes, and so we omit this adjective throughout the paper. A polycube is said to be *proper* in d dimensions if the convex hull of the centers of its cubes is d-dimensional. Following Lunnon [18], we let $\mathrm{DX}(n,d)$ denote the number of fixed polycubes of size n that are proper in d dimensions. Similarly, we denote by $\mathrm{DT}(n,d)$ the number of fixed *tree* polycubes (polycubes whose cell-adjacency graph is a tree) of size n which are proper in d dimensions. Despite the simplicity of these definitions, computing the functions $\mathrm{DX}(n,d)$ and $\mathrm{DT}(n,d)$ has shown to be an extremely difficult task.

[☆]A preliminary version of this paper appeared in EuroComb 2015 [4]. Work on this paper by the first author has been supported in part by Israel Science Foundation Grant 575/15. Work on this paper by the two authors has been supported in part by a grant from Bar-Nir Bergreen Software Technology Center of Excellence.

[*]Corresponding author

Email addresses: `barequet@cs.technion.ac.il` (Gill Barequet), `mshalah@cs.technion.ac.il` (Mira Shalah)

Enumeration of polycubes and computing their asymptotic growth rate are important problems in combinatorics and discrete geometry, originating in statistical physics [8]. While in the mathematical literature these objects are called polycubes (*polyominoes* in two dimensions), they are usually referred to as *lattice animals* in the literature of statistical physics, where they play a fundamental role in the analysis of percolation processes [12, 25] and collapse of branched polymers [9, 11, 16, 21]. To-date, no formula is known for $A_d(n)$, the number of fixed polycubes of size n in d dimensions, for any fixed value of d, let alone in the general case. Counting polyominoes is also a long-standing problem. The number of polyominoes, $A_2(n)$, is currently known up to $n = 56$ [13]. Tabulations of counts of polycubes in higher dimensions appear in the mathematics literature [1, 17, 18] as well as in the statistical-physics literature [10, 12, 22]. The main interest in the function DX stems from the fact that $A_d(n)$ can be easily computed using the formula

$$A_d(n) = \sum_{i=0}^{d} \binom{d}{i} \mathrm{DX}(n, i),$$

given originally by Lunnon [18]. The formula is proved by noting that every proper i-dimensional polycube can be embedded in the d-dimensional space in exactly $\binom{d}{i}$ different ways (according to the choice of dimensions for the polycube to occupy). In addition, if $n \le d$, the polycube simply cannot occupy all d dimensions (since a polycube of size n can occupy at most $n-1$ dimensions), and so $\mathrm{DX}(n, d) = 0$ in this case. Hence, in a matrix listing the values of DX, where the vertical coordinate is n and the horizontal coordinate is d, the top-right triangular half and the main diagonal contain only zeroes. This gives rise to the question of whether a pattern can be found in the sequences $\mathrm{DX}(n, n-k)$, where $k < n$ is the ordinal number of the diagonal. Obviously, if a simple formula is found for $\mathrm{DX}(n, n-k)$ for every k, this will yield a simple formula for $A_d(n)$ (using Lunnon's formula).

The growth constant of polycubes has also attracted much attention in the literature. Klarner [14] showed in a seminal work the existence of the limit $\lambda_2 = \lim_{n\to\infty} \sqrt[n]{A_2(n)}$. Only 32 years later, Madras [20] proved the convergence of the sequence $(A_2(n+1)/A_2(n))_{n=1}^{\infty}$ to λ_2, the growth constant of polyominoes (also known as *Klarner's constant*). The exact value of λ_2 has remained elusive till these days. The currently best known lower and upper bounds on λ_2 are roughly 4.0025 [6] and 4.649551 [15], respectively. In fact, the leading decimal digit of λ was rigorously computed only recently after remaining illusive for over 50 years. In $d > 2$ dimensions, λ_d, the growth constant of d-dimensional polycubes, also exists [20]. It was proven [7] that $\lambda_d = 2ed - o(d)$; moreover, λ_d was estimated at $(2d-3)e + O(1/d)$. It was shown [2] that λ_d^T, the growth constant of *tree* polycubes, is also $2ed - o(d)$ and was estimated at $(2d-3.5)e + O(1/d)$.

Significant progress in estimating λ_d has been obtained along the years in the literature of statistical physics, although the computations usually relied on unproven assumptions and on formulae for $\mathrm{DX}(n, n-k)$ which were interpolated empirically from known values of $A_d(n)$.

Peard and Gaunt [24] predicted that the diagonal formula $\mathrm{DX}(n, n-k)$ has the pattern $2^{n-2k+1}n^{n-2k-1}g_k(n)$ (for $k > 1$), where $g_k(n)$ is a polynomial in n. In fact, k has to be a root of $g_k(n)$ since $\mathrm{DX}(n, 0) = 0$ for $n > 1$. Therefore, the expected form is $2^{n-2k+1}n^{n-2k-1}(n-k)h_k(n)$, where $h_k(n)$ is a polynomial in n, and explicit formulae for $h_k(n)$ for $k \leq 6$ were conjectured [24]. Luther and Mertens [19] later conjectured a formula for $k = 7$. After a careful inspection of the polynomials, which revealed that the leading coefficient of $h_k(n)$ has the form $2^{k-1}/(k-1)!$, Asinowski et al. [3] refined the conjectured formula to

$$\mathrm{DX}(n, n-k) = \frac{2^{n-k}n^{n-2k-1}(n-k)}{(k-1)!}P_c(n),$$

where $P_c(n)$ is a monic polynomial in n. It has also been conjectured [7, 19] that the degree of $P_c(n)$ is $3k-4$. In this paper, we prove rigorously this refined conjecture about the general form of $\mathrm{DX}(n, n-k)$ and the degree of $P_c(n)$.

Using Cayley trees, it can be shown (see, e.g., [7]) that

$$\mathrm{DX}(n, n-1) = 2^{n-1}n^{n-3}$$

(sequence A127670 in the Online Encyclopedia of Integer Sequences [23]). Barequet et al. [7] proved rigorously, for the first time, that

$$\mathrm{DX}(n, n-2) = 2^{n-3}n^{n-5}(n-2)(2n^2-6n+9)$$

(sequence A171860). The proof uses a case analysis of the possible structures of spanning trees of the polycubes, and the various ways in which cycles can be formed in their cell-adjacency graphs. Similarly, Asinowski et al. [3] proved that

$$\mathrm{DX}(n, n-3) = 2^{n-6}n^{n-7}(n-3)(12n^5-104n^4+360n^3-679n^2+1122n-1560)/3,$$

again, by counting spanning trees of polycubes, yet the reasoning and the calculations were significantly more involved. The proof applies the inclusion-exclusion principle in order to count correctly polycubes whose cell-adjacency graphs contained certain subgraphs, so-called “distinguished structures.” In comparison with the case $k = 2$, the number of such structures for $k = 3$ is substantially higher, and the ways in which they can appear in spanning trees are much more varied. The latter proof provided a better understanding of the difficulties that one would face in applying this technique to higher values of k. The number of distinguished structures grows rapidly, the inclusion relations between them are much more complicated, and the ways in which they can be connected by forests are much more varied. This yields a much larger number of terms in the inclusion-exclusion analysis. As anticipated [3], carrying this approach beyond k=3 would create a case analysis beyond the patience of a human, making it totally impractical to manually achieve a similar proof for $k > 3$.

In this paper we create a theoretical set-up for proving the formula for $\mathrm{DX}(n, n-k)$, for a fixed value of k. Our method *fully automates* the manual method presented in [3, 7], allowing the case analysis to be made by a computer. For this nontrivial generalization we prove a few key observations about polycubes that are proper in $n-k$ dimensions. We also provide a general characterization of distinguished structures, and design algorithms that produce, analyze, and enumerate them automatically, even for complex structures, forests, and cycles that do not appear in the case k=3. Using our implementation of this method, we find the explicit formulae (which have never been proven before) for $\mathrm{DT}(n, n-4)$, $\mathrm{DX}(n, n-4)$, $\mathrm{DT}(n, n-5)$, and $\mathrm{DX}(n, n-5)$, stated in the following theorems.

Theorem 1.

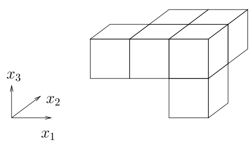
1. $\mathrm{DT}(n, n-4) = 2^{n-7}n^{n-9}(n-4)(8n^8 - 140n^7 + 1010n^6 - 3913n^5 + 9201n^4 - 15662n^3 + 34500n^2 - 120552n + 221760)/6$.
2. $\mathrm{DX}(n, n-4) = 2^{n-7}n^{n-9}(n-4)(8n^8 - 128n^7 + 828n^6 - 2930n^5 + 7404n^4 - 17523n^3 + 41527n^2 - 114302n + 204960)/6$.

Theorem 2.

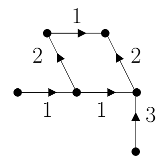
1. $\mathrm{DT}(n, n-5) = 2^{n-9}n^{n-11}(n-5)(240n^{11} - 6480n^{10} + 73640n^9 - 461232n^8 + 1778615n^7 - 4707195n^6 + 11632070n^5 - 41919528n^4 + 158857920n^3 - 483329520n^2 + 1481660640n - 2863123200)/360$.
2. $\mathrm{DX}(n, n-5) = 2^{n-12}n^{n-11}(n-5)(240n^{11} - 6000n^{10} + 62240n^9 - 356232n^8 + 1335320n^7 - 4062240n^6 + 12397445n^5 - 42322743n^4 + 150403080n^3 - 535510740n^2 + 1923269040n - 3731495040)/45$.

2. Definitions and Notations

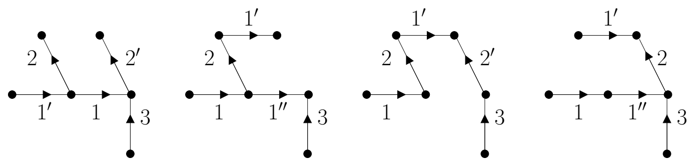
Integer Partition. A partition of a natural number $m \in \mathbb{N}$ is a way of writing m as the sum of one or more positive integers, i.e., $m = \sum_i a_i$. Two sums that differ only in the order of their summands are considered the same, and so we choose the canonical representation of a partition to be the list of its summands in nondecreasing order. Let $\Pi(m)$ denote the set of all partitions of the natural number m. For example, there are two partitions of 2 and three partitions of 3: $\Pi(2) = \{1+1, 2\}$ and $\Pi(3) = \{1+1+1, 1+2, 3\}$. For a partition $p \in \Pi(m)$, we denote by $|p|$ the number of summands in p, and by $p[i]$ the ith summand of p. In addition, we let $\oplus p = \sum_{i=1}^{|p|} p[i]$ denote the sum of the elements of p (i.e., $\oplus p = m$), and $\pi(p)$ denote the number of essentially-different permutations of the summands of p. For example, $\pi(1,1,1) = 1$ and $\pi(1,2) = 2$. Finally, we let $p_r = \oplus p + |p|$. (Note that $\oplus p + 1 \le p_r \le 2 \oplus p$ since $1 \le |p| \le \oplus p$). For two partitions p_1 and p_2, we say that p_1 contains p_2, and denote this relation by $p_2 \preceq p_1$, if there is a subpartition p_1^* of p_1 (an ordered subset of the elements of p_1), such that $|p_1^*| = |p_2|$ and $p_2[i] \le p_1^*[i]$ for all $1 \le i \le |p_2|$. For example, $2 \preceq 1+2$, but $2 \not\preceq 1+1+1$.



(a) Polycube P



(b) Adjacency graph $\gamma(P)$



(c) The spanning trees of $\gamma(P)$

Figure 1: A polycube P, its corresponding adjacency graph $\gamma(P)$, and the spanning trees of $\gamma(P)$.

Graph Isomorphism. Let $G = (V_G, E_G)$ and $H = (V_H, E_H)$ be two directed edge-labeled graphs with respective edge labels W_G and W_H, such that $|V_G| \leq |V_H|$. G is said to be *isomorphic* to H if there is a bijection $f : V_G \to V_H$ such that

- If for any $u, v \in V_G$ such that $(u, v) \in E_G$, then $(f(u), f(v)) \in E_H$; and
- If for any $e_1 = (u_1, v_1), e_2 = (u_2, v_2) \in E_G$, such that the labels of e_1 and e_2 are equal, then the labels of $(f(u_1), f(v_1))$ and $(f(u_2), f(v_2))$ are equal.

An automorphism of G is a form of symmetry in which G is mapped into itself while preserving the conditions above.

3. Overview of the Method

Denote by $\mathcal{P}_n$ the set of polycubes of size n which are proper in $n-k$ dimensions. (The value of k is fixed, therefore we omit it from the notation.) For $P \in \mathcal{P}_n$, let $\gamma(P)$ denote the directed edge-labeled graph that is constructed as follows:

- The vertices of $\gamma(P)$ correspond to the cells of P;
- Two vertices of $\gamma(P)$ are connected by an edge if the corresponding cells of P are adjacent;
- An edge has label i $(1 \leq i \leq n-k)$ if the corresponding cells have different i-coordinates (their common $(d-1)$-dimensional face is perpendicular to the x_i axis); and
- The direction of the edge is from the lower to the higher cell (with respect to the x_i direction).

See Figure 1 for an example.

Since $P \mapsto \gamma(P)$ is an injection, it suffices to count the graphs obtained from the members of $\mathcal{P}_n$ in this way. We shall count these graphs by counting their spanning trees. A spanning tree of $\gamma(P)$ has $n-1$ edges labeled by numbers from the set $\{1, 2, \ldots, n-k\}$; all these labels are present because the polycube is proper in $n-k$ dimensions. Hence, $n-k$ edges of the spanning tree are labeled with the labels $1, 2, \ldots, n-k$, and the remaining $k-1$ edges are labeled with repeated labels from the same set. Observation 3 characterizes all the different possibilities of repeated edge-labels in the spanning tree of a proper polycube.

Observation 3. *There is a bijection between the combinations of repeated edge-labels and the partitions of the integer $k-1$. Specifically, each partition $p \in \Pi(k-1)$ corresponds to the combination of $|p|$ different repeated labels in the spanning tree (and p_r repeated labels in total), in which the ith repeated label appears $p[i]+1$ times. In such a case, we say that the tree is* labeled according to p*.*

Observation 4. *Every label must occur an even number of times in any cycle of $\gamma(P)$.*

An immediate consequence of Observation 3 is that a tree can have at most $2(k-1)$ repeated edge labels, in which case the repeated labels appear in $k-1$ pairs.

In order to compute $|\mathcal{P}_n|$, we consider all possible directed edge-labeled trees of size n with combinations of edge labels as in Observation 3, and count only those that represent valid polycubes. In Section 3.2 we characterize all substructures that are present in some of these trees due to the fact that there are less dimensions than cells. By analyzing these substructures, we are able to compute how many of these trees actually represent polycubes. Then, we develop formulae for the numbers of all possible spanning trees of the polycubes, and then derive the actual number of polycubes.

3.1. Counting

Lemma 5. *[3, Lemma 7] [7, Lemma 2] The number of directed trees with n vertices and $n-1$ distinct edge labels $1, \ldots, n-1$ is $2^{n-1}n^{n-3}$, for $n \geq 2$.*

Our approach is to based on counting polycubes by enumerating spanning trees of their adjacency graphs. In order to apply Lemma 5 to counting spanning trees of polycubes, we shall distinguish between the labels repeated in a tree. As explained above, a spanning tree T of $\gamma(P)$, for a polycube $P \in \mathcal{P}_n$, must be labeled according to some partition $p \in \Pi(k-1)$. Let us, then, denote by $\ell_1, \ldots \ell_{|p|}$ the repeated labels of T, such that ℓ_i appears $p[i]$ times in T. We will distinguish between the $p[i]$ edges of T labeled with ℓ_i by relabeling them with the (*distinct*) labels $\ell_i, \ell_i', \ell_i'', \ldots$ (see, e.g., Figure 1 (c)). However, in $\gamma(P)$, the repeated labels are not distinguished. The trees that can be obtained by exchanging (permuting) $\ell_i, \ell_i', \ell_i'', \ldots$, are, in fact, also spanning trees of $\gamma(P)$.

Let, then, T_p denote the number of directed trees with n vertices that are labeled according to $p \in \Pi(k-1)$.

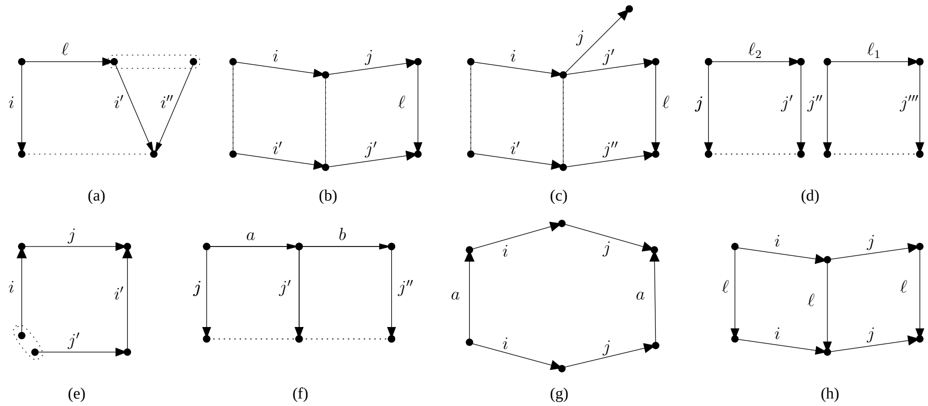


Figure 2: (a–f) A few distinguished structures for $k = 4$ (note that (d) is disconnected); (g,h) Cycle structures. A dotted line is drawn between every pair of neighboring cells and around every pair of coinciding cells.

Corollary 6. $T_p = \pi(p)\binom{n-k}{|p|}2^{n-1}n^{n-3}$.

3.2. Distinguished Structures

3.2.1. Generation

In the reasoning below, we shall consider several small structures which may be contained in the spanning trees that we count. These structures are interesting because the following two things may happen when we attempt to build the polycube corresponding to a directed edge-labeled tree:

(a) Cells may coincide (Figures 4(a,b) and 2(e)). A tree with overlapping cells is invalid and does not correspond to any polycube; and

(b) Two cells which are not connected by a tree edge may be adjacent (Figures 4(c,d) and 2(b)). Such a tree corresponds to a polycube which has cycles in its cell-adjacency graph, and therefore, its spanning tree is not unique.

Similarly to Observation 4, for every label of an edge along the path between two vertices that correspond to coinciding cells, repetitions of this label occur an even number of times on this path, and a structure that leads to a non-existing adjacency results in a path obtained by removing one edge from a cycle of an even length. Therefore, the length (number of edges) of a path that connects two coinciding cells (respectively, neighboring cells) is upper bounded by $2(k-1)$ (respectively, $2k-1$). Therefore, the length of a cycle in $\gamma(P)$ can be at most $2k$. Moreover, the number of cycles in $\gamma(P)$ is upper bounded by $k-1$.

In order to count trees correctly, we will consider several small structures contained in the trees we count, which cause the two problems above. Following [3], we will refer to such structures as distinguished structures. A *distinguished structure* is a small labeled graph that is "responsible" for the presence

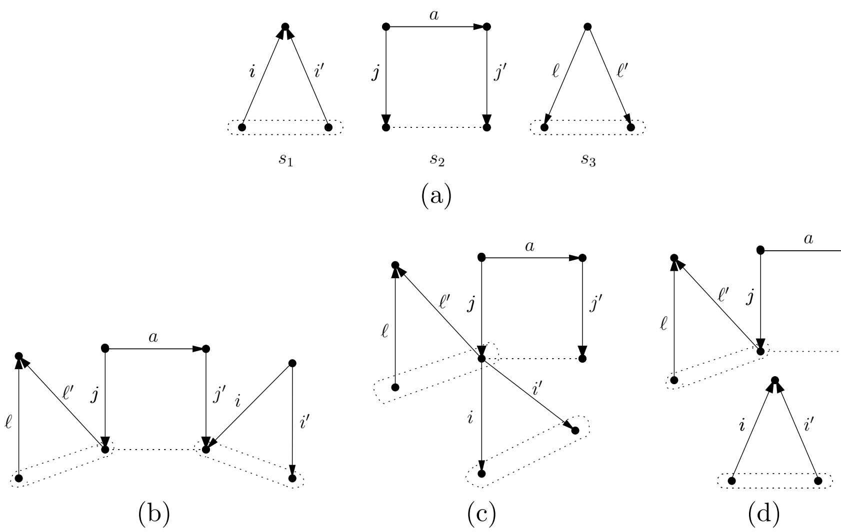


Figure 3: (a) A disconnected distinguished structure which has three connected components s_1, s_2, s_3; (b–d) A few possible configurations of s_1, s_2, s_3.

of two coinciding or adjacent cells, as explained above. More precisely, a distinguished structure is the union of all paths (edges and incident vertices) that run between two coinciding or adjacent cells. Every such path uses up some repeated labels. Therefore, the number of their occurrences in the trees that we count is limited. The enumeration of the distinguished structures is, thus, a finite task.

A distinguished structure can be classified as one of the following:

- **Basic Structures:** A basic structure is formed of a path that connects a pair of coinciding or neighboring cells. Basic structures are the building blocks of the two other types of distinguished structures.
- **Compound Structures:** A compound structure is a connected structure that contains two or more occurrences of basic structures which cover all its edges, such that every such occurrence shares an edge with another occurrence.
- **Disconnected Structures:** A disconnected structure is a collection of edge-connected structures (basic or compound). Different components of a disconnected structure can share a vertex, as illustrated in Figure 3.

Let $\mathcal{DS}_k$ denote the set of distinguished structures in $n-k$ dimensions. We hereafter refer to the *size* of a distinguished structure $\sigma \in \mathcal{DS}_k$ as the number of its vertices, and denote this quantity by $|\sigma|$.

Observation 7. *The size of a basic structure is bounded from above by $2k$.*

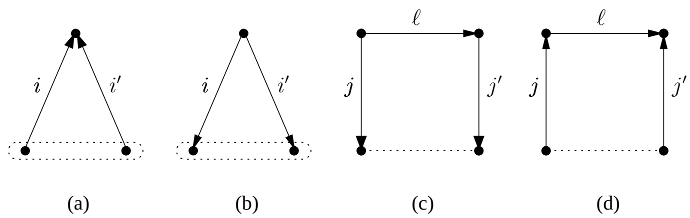


Figure 4: All members of $\mathcal{DS}_2$.

Lemma 8. *Let $\sigma \in \mathcal{DS}_k$ be a compound structure labeled according to some partition $p \in \Pi(i)$, $1 \leq i \leq k-1$. Then, $|\sigma| \leq 2(i+1)$.*

Proof: By induction on i. For the basis of the induction ($i = 1$), the only distinguished structures are shown in Figure 4 and there are no compound structures. For $i = 2$, there are six compound structures formed by identifying edges of two copies of the structure in Figure 4(a) or two copies of the structure in Figure 4(b), forming a connected structure of size 6 which is labeled according to $(2) \in \Pi(2)$, and for which the claim holds.

The induction hypothesis is the assumption that the claim is correct for $i \leq k-2$. We now proceed with the induction step. Let $\sigma \in \mathcal{DS}_k$ be a compound structure labeled according to some partition $p \in \Pi(k-1)$. The structure σ can be decomposed into two compound or basic structures σ_1 and σ_2 labeled according to $p_1 \in \Pi(j)$ and $p_2 \in \Pi(\ell)$, respectively, such that:

- $1 \leq j, \ell < k-1$;
- $|\sigma| = |\sigma_1| + |\sigma_2| - 2$; and
- $p_r = |p| + \oplus p = |p| + (k-1) = |p_1| + \oplus p_1 + |p_2| + \oplus p_2 - 1 = j + \ell + |p_1| + |p_2| - 1$. Hence, $j + \ell = k + |p| - (|p_1| + |p_2|)$.

Note that $|p| - (|p_1| + |p_2|) = -1$ since σ_1 and σ_2 share an edge. Hence, $j + \ell = k - 1$. By the induction hypothesis, we have that $|\sigma_1| \leq 2(i+1)$ and $|\sigma_2| \leq 2(j+1)$. Therefore, $|\sigma| = |\sigma_1| + |\sigma_2| - 2 \leq 2(i+1) + 2(j+1) - 2 = 2(i+j+1) = 2k$. □

Corollary 9. *The size of a compound structure is bounded from above by $2k$.*

Lemma 10. *Let $\sigma \in \mathcal{DS}_k$ be a distinguished structure that is composed of c connected components and labeled according to some partition $p \in \Pi(i)$, $1 \leq i \leq k-1$. Then, $p_r \leq |\sigma| - c \leq p_r + \oplus p$.*

Proof: The first relation, $|\sigma| - c \geq p_r$, states that the number of edges in σ is at least the number of repeated labels implied by the partition p. Indeed, this condition is necessary for p to label σ. The second relation is true since the number of *unique* labels in σ (labels each of which appears only once in σ) is bounded from above by $\oplus p$. This is because every repeated label ℓ_i (with $p[i]+1$ occurrences in σ) can add to σ at most $p[i]$ unique edge labels. $\square$

The above characterization of distinguished structures allows for the design of an algorithm for producing $\mathcal{DS}_k$. The algorithm begins with generating all "free trees" (non-isomorphic trees) of size at most $2k$, the upper bound specified in Observation 7 and Corollary 9. Then, we label the edges of every free tree T of size t according to every partition $p \in \cup_{i=1}^{k-1}\Pi(i)$ which satisfies the relation in Lemma 10 ($p_r \le t-1 \le p_r + \oplus p$) so as to obtain a directed edge-labeled tree T'. Then, we check whether T' contains coinciding or neighboring cells by using a simple depth-first traversal that starts from an arbitrary node and assigns every other node its appropriate coordinate. If such cells are found, then T' is added to $\mathcal{DS}_k$ if it is *not* isomorphic to any structure $\sigma \in \mathcal{DS}_k$ of size t, and at least one of the following conditions holds:

1. T' contains two coinciding or neighboring cells which are connected by a path of length $t-1$ (see, e.g., Figures 4(a–d) and 2(b,e));
2. T' is isomorphic to the union of $d_1, \ldots, d_m \in \mathcal{DS}_k$, such that the isomorphic copies of $d_1, \ldots, d_m$ in T' cover all its edges (see, e.g., Figures 2(c,f)).

Disconnected distinguished structures (see, e.g., Figures 2(d) and 3) are generated by checking if every collection of edge-connected structures in $\mathcal{DS}_k$ yields a single disconnected structure labeled according to some $p \in \cup_{i=1}^{k-1}\Pi(i)$.

3.2.2. Enumeration

Let us now turn to the enumeration of occurrences (i.e., isomorphic copies) of distinguished structures in directed trees with edge labels as explained earlier. In counting directed trees with $n-1$ labeled edges which have distinguished structures as subgraphs, the following logic will be used.

Lemma 11. *[7, Lemma 4] The number of ordered sequences $T = (\tau_1, \ldots, \tau_k)$ of $k \ge 1$ rooted trees with a total of $n-k$ edges and distinct edge labels $1, \ldots, n-k$ is $n^{n-k-1}k$.*

The following lemma is a generalization of the previous one.

Lemma 12. *The number of ordered sequences $\tilde{T} = (\tau_1, \ldots, \tau_k)$ of $k \ge 1$ rooted trees that contain c (additional) distinguished roots (which may coincide), such that τ_1 has at least two distinguished roots, with a total of $n-k$ edges and distinct edge labels $1, \ldots, n-k$, is $n^{n-k+c-1}$.*

Proof: Consider a sequence T as in Lemma 11, and mark c arbitrary vertices as the extra roots. In this way, $n^{n-k+c-1}k$ sequences $\tilde{T}$ are obtained. The component of $\tilde{T}$ which has an extra root is any of $\tau_1 \ldots \tau_k$ with equal probability. Therefore, in order to get the number of sequences in which τ_1 has at least two distinguished roots, we have to divide by k, obtaining $n^{n-k+c-1}$. $\square$

In fact, we will use Lemma 12 in order to prove Lemma 13, which provides the formula for the number of occurrences of any structure in $\mathcal{DS}_k$.

Lemma 13. *Let $\sigma \in \mathcal{DS}_k$ be a distinguished structure composed of $c \geq 1$ trees $s_1, \ldots, s_c$ with distinct edge labels $1, \ldots, |\sigma|-c$. The number of occurrences of σ in trees of size n with distinct edge labels $1, \ldots, n-1$ is $F_n(\sigma) = n^{n-|\sigma|+2c-3} \prod_{i=1}^{c} |s_i|$.*

Proof: For a distinguished structure σ, which has a single connected component ($c = 1$), we obtain $|\sigma| n^{n-|\sigma|-1}$ occurrences, as in Lemma 11, which equals the number of ordered sequences of $|\sigma|$ rooted trees that can be attached to the vertices of σ. For a distinguished structure that has $c > 1$ connected components, as in Lemma 12, there has to be a sequence $\tilde{t}$ of $|\sigma|-(c-1)$ trees, with $c-1$ additional distinguished roots that are needed to connect the c components of σ, as follows. The factor $\prod_{i=1}^{c} |s_i|$ stands for the number of options for choosing the *connectors*: The vertices through which the components of σ are connected. The vertices of s_1 are attached to the first $|s_1|$ roots, such that the connector of s_1 is attached to τ_1 (from Lemma 12). (This is why τ_1 must have at least two distinguished roots). Then, the vertices of rest of the components s_i (except their connectors) are attached to the remaining roots. Finally, the $c-1$ connectors are attached to the $c-1$ additional chosen roots. Applying Lemma 12 yields the claimed formula. □

A special case of Lemma 13 is provided in the Appendix.

Let now $\mathcal{F}_n(\sigma)$ denote the number of occurrences of σ in *directed* edge-labeled trees of size n.

Corollary 14. $\mathcal{F}_n(\sigma) = 2^{n-|\sigma|+c-1} F_n(\sigma)$.

Let $\sigma \in \mathcal{DS}_k$ be a distinguished structure labeled according to $p' \in \cup_{i=1}^{k-1} \Pi(i)$. Let us denote by $\mathcal{O}_p(\sigma)$ the number of occurrences of σ in directed trees of size n that are labeled according to $p \in \Pi(k-1)$.

Observation 15. *If $p' \not\preceq p$, then $O_p(\sigma) = 0$.*

The computation of $\mathcal{O}_p(\sigma)$ involves the following steps:

- Choose the $|p|$ repeated labels of the tree out of the possible $n-k$ labels.
- Choose the $|p'|$ repeated labels of σ out of the $|p|$ repeated labels of edges of the tree.
- Choose the unique labels of σ (e.g., the label ℓ in structures (b,c) in Figure 2), if there are any.
- Calculate the number of essentially-different structures that can be produced out of the $\prod_{j=1}^{|p'|} (p'[j]!)$ possible configurations of the repeated labels of σ. For example, for structure (a) in Figure 4, all the configurations yield the same structure, whereas for structure (b), there are two essentially-different structures. In one structure, the edge labeled i is attached to the head of the edge labeled ℓ, whereas in the other structure, the edge labeled i' is attached to the head of the edge labeled ℓ. Lastly, for structure (a) (Figure 2), there are six different configurations of the labels i, i', i'' which

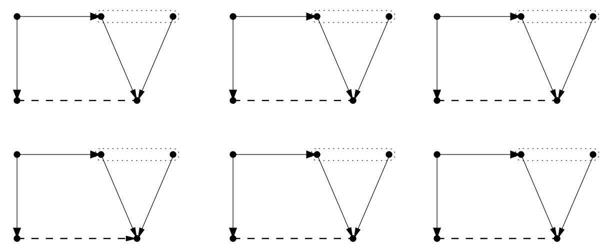


Figure 5: The six different configurations of the structure in Figure 2(a).

yield six different structures (shown in Figure 5). This number can be obtained by computing the number of symmetries (automorphisms) of σ.

- Finally, multiply the numbers obtained in the previous steps by $\mathcal{F}_n(\sigma)$.

Here are two detailed examples.

We demonstrate the computation of $\mathcal{O}_p(\sigma)$ with two of the structures shown in Figure 2. The first structure, σ_1, is the one shown in Figure 2(a). σ_1 is labeled according to $(3) \in \Pi(2)$. Therefore, for the case $k = 4$, σ_1 may appear in trees labeled according to $(2,3), (4) \in \Pi(3)$, but not according to $(2,2,2) \in \Pi(3)$ since $(3) \not\preceq (2,2,2)$. Let us detail the computation of $\mathcal{O}_{(2,3)}(\sigma_1)$. First, there are $\binom{n-4}{2}$ options to choose the two repeated labels ℓ_1, ℓ_2 in the tree. We multiply by $\pi((2,3)) = 2$ (either $\ell_1, \ell_1', \ell_2, \ell_2', \ell_2''$ or $\ell_1, \ell_1', \ell_1'', \ell_2, \ell_2'$). The label i that is repeated three times in σ_1 is thus determined uniquely. Assume without loss of generality that i is assigned the label ℓ_1. There are $n-4$ options to choose the label ℓ: it must be different from $\ell_1, \ell_1', \ell_1''$ (but may be equal to ℓ_2 or ℓ_2'). There are three options for choosing which label of $\ell_1, \ell_1', \ell_1''$ is attached to the head of ℓ, then two options for choosing the label that is attached to the tail of ℓ. This number is calculated from the number of automorphisms of σ_1. To complete the computation of $\mathcal{O}_{(2,3)}(\sigma_1)$, we multiply by $\mathcal{F}_n(\sigma_1)$.

The second structure, σ_2, is the one shown in Figure 2(e). σ_2 is labeled according to the partition $(2,2) \in \Pi(2)$. Hence, it may appear in trees labeled according to $(2,2,2)$, $(2,3) \in \Pi(3)$, but not according to $(4) \in \Pi(3)$ since $(2,2) \not\preceq (4)$. For computing $O_{(2,2,2)}(\sigma_2)$, there are $\binom{n-4}{3}$ options for choosing the three labels repeated in the tree. This uniquely yields the labels $\ell_1, \ell_1', \ell_2, \ell_2', \ell_3, \ell_3'$ (note that $\pi((2,2,2)) = 1$). Then, there are $\binom{3}{2}$ options to choose the repeated labels i, j. Assume without loss of generality that the chosen labels are ℓ_1 and ℓ_2. Note the symmetry in this structure: It does not matter if i is assigned the label ℓ_1 and j is assigned the label ℓ_2, or vice versa, since the two options yield the same structure. Again, the number of symmetries of σ_2 is calculated using the number of its automorphisms and the computation is completed by multiplying by $\mathcal{F}_n(\sigma_2)$.

4. Inclusion-Exclusion Graph

When counting the occurrences of a distinguished structure $\sigma \in \mathcal{DS}_k$, other distinguished structures which contain multiple occurrences of σ are counted multiple times. Obviously, if a distinguished structure σ_b contains c occurrences of a smaller structure σ_s, then σ_b is accounted for c times when counting the occurrences of σ_s. The inclusion-exclusion principle is applied to resolve this dependency between the different structures. In order to obtain the number of *trees* that contain σ as a subtree (using the quantity $\mathcal{O}_p(\sigma)$), we build an inclusion-exclusion graph IE $= (\mathcal{V}, \mathcal{E})$. This graph contains a vertex corresponding to each structure $\sigma \in \mathcal{DS}_k$. There is an edge $e = \sigma_1 \to \sigma_2$ labeled with c if σ_1 contains c occurrences of σ_2. Let $\ell(e)$ denote the label of the edge e, $h(\sigma)$ denote the length of the longest path from σ to a root of IE, and $I(\sigma_2) = \{\sigma_1 \in \mathcal{V} : (\sigma_1, \sigma_2) \in \mathcal{E}\}$. The set of roots $\mathcal{R} = \{v \in \mathcal{V} : I(v) = \emptyset\}$ of the IE graph contains all the structures that are not contained in any other structure; in a sense, those are the "big" structures. Figure 6 shows a subgraph of the IE graph for $k = 4$. Let us denote by $T_p(\sigma)$ the number of trees of size n labeled according to $p \in \Pi(k-1)$ that contain σ but no $\sigma' \in I(\sigma)$ as a subtree.

Lemma 16. $T_p(\sigma) = \mathcal{O}_p(\sigma) - \sum_{\sigma' \in I(\sigma)} \ell((\sigma', \sigma)) T_p(\sigma')$.

Proof: By Induction on $h(\sigma)$.

- The roots of the IE graph $\sigma \in \mathcal{R}$, for which $h(\sigma) = 0$, represent distinguished structures that are not contained in any other structure. Therefore, for any partition $p \in \Pi(k-1)$, the number of trees that contain σ as a subtree equals the number of occurrences of σ in directed trees labeled according to p. Thus, $T_p(\sigma) = \mathcal{O}_p(\sigma)$.
- Induction hypothesis: The claim is correct for vertices of height $h < h_0$.
- Induction step: Let $\sigma \in \mathcal{V}$ be at height h_0 ($h(\sigma) = h_0$), and let $\sigma' \in I(\sigma)$. The trees that contain σ' as a subtree are counted $\ell((\sigma', \sigma)) T_p(\sigma')$ times in $\mathcal{O}_p(\sigma)$. Therefore, subtracting $\ell((\sigma', \sigma)) T_p(\sigma')$ from $\mathcal{O}_p(\sigma)$ excludes all the trees that contain σ' as a subtree. Thus, $T_p(\sigma) = \mathcal{O}_p(\sigma) - \sum_{\sigma' \in I(\sigma)} \ell((\sigma', \sigma)) T_p(\sigma')$.

$\square$

A simple bottom-up procedure traverses the IE graph, computing $T_p(\sigma)$ for every structure $\sigma \in \mathcal{V}$.

5. Counting Polycubes

A proper tree polycube is a polycube $P \in \mathcal{P}_n$ for which $\gamma(P)$ is a tree. The other polycubes $P' \in \mathcal{P}_n$ are those for which $\gamma(P')$ contains cycles.

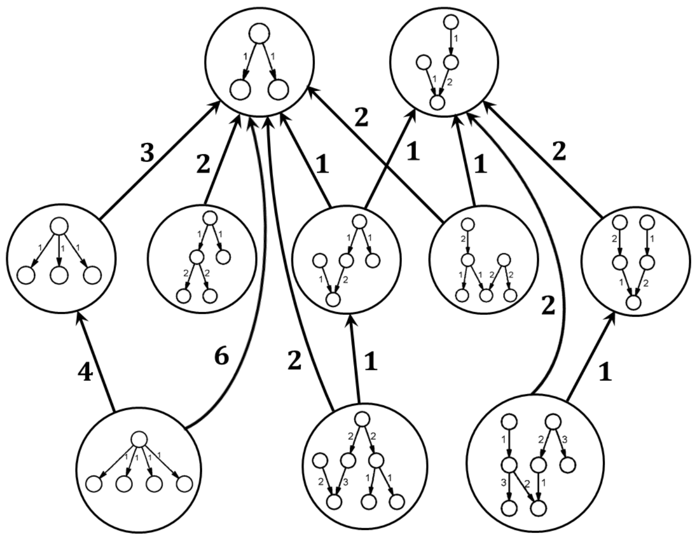


Figure 6: A snapshot of the IE graph for $k = 4$.

5.1. Trees

Every tree polycube gives rise to a unique spanning tree. For every choice of repeated labels $p \in \Pi(k-1)$, let $\mathrm{DT}_p(n)$ denote the number of proper tree polycubes which corresponding (unique) spanning trees are labeled according to p. Corollary 6 specifies T_p, the total number of directed trees with n vertices that are labeled according to p. Every such tree corresponds to a tree polycube in $\mathcal{P}_n$ unless it contains a distinguished structure as a subtree. (Indeed, it can neither contain a distinguished structure that has coinciding cells because the latter is illegal, nor can it contain a distinguished structure that has neighboring cells since it is a tree.) Therefore, all the trees that contain a distinguished structure as a subtree must be excluded. Hence,

$$\mathrm{DT}(n, n-k) = \sum_{p \in \Pi(k-1)} \mathrm{DT}_p(n) = \sum_{p \in \Pi(k-1)} \frac{T_p - \sum_{\sigma \in \mathcal{DS}_k} T_p(\sigma)}{\prod_{j=1}^{|p|} p[j]!} \qquad (1)$$

The division by $\prod_{j=1}^{|p_i|}(p_i[j]!)$ is because each tree polycube is counted that many times.

5.2. Nontrees

Let $\sigma \in \mathcal{DS}_k$ be a distinguished structure which contains only adjacent cells and no coinciding cells. Let σ_c denote the graph that is constructed by adding to σ all the missing cycle-edges between every pair of adjacent cells. The structure σ_c is a *cycle*. For example, the distinguished structure shown in Figure 2(b) is a spanning tree of the cycle shown in Figure 2(h). Two cycle structures c_1

and c_2 are considered distinct if either c_1 is not isomorphic to c_2, or c_2 is not isomorphic to c_1. Let $\mathcal{C}$ denote the set of all cycle structures of polycubes in $\mathcal{P}_n$. The set $\mathcal{C}$ can be found using $\mathcal{DS}_k$ as described. Note that two different distinguished structures may be spanning trees of the same cycle. For every cycle $\mathcal{C}_i \in \mathcal{C}$, let $P_{\mathcal{C}_i}$ denote the number of polycubes $P \in \mathcal{P}_n$ that contain $\mathcal{C}_i$ in their cell-adjacency graph $\gamma(P)$. Suppose that a distinguished structure $\sigma \in \mathcal{DS}_k$ has c occurrences in $\mathcal{C}_i$. Then, we have that

$$P_{\mathcal{C}_i} = \sum_{p \in \Pi(k-1)} \frac{T_p(\sigma)}{c \prod_{j=1}^{|p|} p[j]!}. \qquad (2)$$

This follows from the definition of $T_p(\sigma)$. Finally, we reach the desired formula.

$$\mathrm{DX}(n, n-k) = \mathrm{DT}(n, n-k) + \sum_{i=1}^{|\mathcal{C}|} P_{\mathcal{C}_i}. \qquad (3)$$

Theorem 17. *The general pattern of* $\mathrm{DX}(n, n-k)$*, for a fixed* $k > 0$*, is* $\frac{2^{n-k}}{(k-1)!} n^{n-2k-1}(n-k)P_{3k-4}(n)$*, where* $P_c(n)$ *is a monic polynomial in* n *of order* c.

Proof: From the discussion in Sections 3 and 4 about the terms of the inclusion-exclusion formula and equations (1), (2), and (3), we conclude that for any $p \in \Pi(k-1)$, by Corollary 6, the degree of T_p is at least $n-3$, and that the highest power of n, namely, $n+k-4$, is contributed by $T_{(2,2,\ldots,2)} = \binom{n-k}{k-1} n^{n-3} 2^{n-1}$: The edge-repetition configuration in the labeled tree, corresponding to the partition $(1, 1, \ldots, 1) \in \Pi(k-1)$, contains $k-1$ pairs of repeated edge labels. The power of n contributed by $T_{p'}$, for all the other partitions $p' \in \Pi(k-1)$, is smaller than $n+k-4$.

Now let $\sigma \in \mathcal{DS}_k$ be a distinguished structure composed of c connected components, and labeled according to some partition $p' \in \cup_{i=1}^{k-1} \Pi(i)$. Let u denote the number of unique labels in σ. We claim that the power of n in $\mathcal{O}_p(\sigma)$, for any partition $p \in \Pi(k-1)$ for which $p' \preceq p$, is bounded by $n-2k-1$ from below and by $n+k-5$ from above. As a result, since $T_p(\sigma)$ consists of linear combinations of $\mathcal{O}_p(\sigma)$ and $\mathcal{O}_p(\sigma')$ for other structures $\sigma' \in \mathcal{DS}_k$ (where the coefficients of the combinations are functions of k only), the power of n in $T_p(\sigma)$ is also bounded by $n-2k-1$ from below and by $n+k-5$ from above.

The power of n in $\mathcal{O}_p(\sigma)$ is contributed by the following three factors:

- $\binom{n-k}{|p|}$: This factor corresponds to choosing $|p|$ repeated labels, and it clearly contributes at most $|p|$ powers of n to $\mathcal{O}_p(\sigma)$.
- $\mathcal{F}_n(\sigma)$: The power of n contributed by $\mathcal{F}_n(\sigma)$ is $n-|\sigma|+2c-3$ (Lemma 13).
- u: Naturally, the power of n contributed by the choice of these u unique labels can be at most u.

Upper bound. The power of n is bounded from above by the sum of these three factors. We now prove that $n{-}|\sigma|{+}2c{-}3{+}u{+}|p| \leq n{+}k{-}5$:

1. By Lemma 10, $|\sigma| - c \geq p'_r$. Moreover, $|\sigma| - c = p'_r + u$ since, clearly, the number of edges in σ ($|\sigma| - c$) equals the total number of edge-labels in it ($p'_r + u$).
2. Trivially, $|p| \leq k - 1$.

Therefore, $n - |\sigma| + 2c - 3 + u + |p| \leq_2 n - |\sigma| + 2c + u + k - 4$. To show that $n - |\sigma| + 2c + u + k - 4 \leq n + k - 5$, it suffices to show that $-|\sigma| + 2c + u \leq -1$. By multiplying this relation by -1, we obtain $|\sigma| - 2c - u \geq 1$. We have that $|\sigma| - 2c - u \geq_1 p'_r + u - c - u = p'_r - c$. We now claim that indeed $p'_r - c \geq 1$. This is because every connected component in σ must have at least one pair of coinciding or neighboring cells, and in order to have such cells, every connected component must have at least two repeated labels, implying that the difference between p'_r (the total number of repeated labels in σ) and c (the number of connected components of σ) is at least 1.

Lower bound. In order to show that the power of n in $\mathcal{O}_p(\sigma)$ is bounded from below by $n{-}2k{-}1$, we prove that the power of n contributed by the second factor, namely, $n{-}|\sigma|{+}2c{-}3$, is at least $n{-}2k{-}1$, that is, $-|\sigma|{+}2c \geq -2(k{-}1)$.

Lemma 18. *Let $\sigma \in \mathcal{DS}_k$ be a distinguished structure composed of c connected components and labeled according to some partition $p' \in \cup_{i=1}^{k-1}\Pi(i)$. Then, $-|\sigma|{+}2c \geq -2i$.*

Proof: By induction on i.

- **Induction basis ($\boldsymbol{i = 1}$):** The only structures labeled according to $\Pi(1) = 1$ are two paths of length 2 and two paths of length 4 (all shown in Figure 4), for which we have $-|\sigma| + 2 \geq -2$.

- **Induction hypothesis:** The claim is correct for structures labeled according to $i \leq k - 2$.

- **Induction step:** Let $\sigma \in \mathcal{DS}_k$ be labeled according to $p' \in \Pi(k{-}1)$. If σ is a basic or compound structure, then, by Corollary 7 and Lemma 8, we have that $-|\sigma|{+}2 \geq -2k{+}2 = -2(k{-}1)$. Otherwise, σ can be decomposed into two structures σ_1 and σ_2 with c_1 and c_2 connected components, that are labeled according $p_1 \in \Pi(i)$ and $p_2 \in \Pi(j)$, respectively, such that

  - $1 \leq i, j < k - 1$;
  - $|\sigma| = |\sigma_1| + |\sigma_2|$;
  - $c = c_1 + c_2$;
  - $p_r = \oplus p{+}|p| = (k{-}1){+}|p| = \oplus p_1{+}|p_1|{+}\oplus p_2{+}|p_2| = i{+}j{+}|p_1|{+}|p_2|$; thus, $i + j = (k - 1) + |p| - (|p_1| + |p_2|)$.

By the above and the induction hypothesis, $-|\sigma| + 2c = -(|\sigma_1| + |\sigma_2|) + 2(c_1 + c_2) \geq -2i - 2j = -2(i+j) \geq -2(k-1)$. The last relation is correct since, clearly, $|p| \leq |p_1| + |p_2|$.

□

We return to the proof of the main theorem.

In Equation (1), $T_{(2,2,\ldots,2)}$ is divided by 2^{k-1} ($\prod_{j=1}^{|k-1|} 2$). Thus, the coefficient of the highest power of n is $\frac{2^{n-k}}{(k-1)!}$. Hence, we obtain a global formula of the form $\frac{2^{n-k}}{(k-1)!}(n^{n+k-4} + \cdots + cn^{n-2k-1})$, where c is some constant independent of n. We can now factor out the quantity n^{n-2k-1} to obtain a formula of the form $\frac{2^{n-k}}{(k-1)!} n^{n-2k-1} P_{3k-3}(n)$. Finally, k must be a root of $\mathrm{DX}(n, n-k)$ since a polycube of size $n = k$ cannot span $n - k = 0$ dimensions (unless $n = k = 1$). Factoring out $n-k$ yields the claimed pattern. □

Note that $3k - 3$ known values of $\mathrm{DX}(n, n-k)$ (for a specific value of k), including the two trivial values $\mathrm{DX}(k, 0) = 0$ and $\mathrm{DX}(k+1, 1) = 1$, suffice for interpolating uniquely $P_{3k-4}(n)$. However, a "physical" argument [19] implies that as few as k values suffice for interpolating the polynomial.[1] In a nutshell, this argument is based on the unproven assumption that the "free energy" $((\log \mathrm{CX}(n, d))/n)$ has a well-defined $1/d$-expansion whose coefficients depend on n and are bounded when n tends to infinity. Then, the powers of n in the terms of the expansion are tuned so as to avoid the explosion of the terms, thereby imposing constraints which allow the computation of $\mathrm{DX}(n, n-k)$ by knowing only k values of it.

6. Results

The method outlined in the preceding sections was fully implemented in a parallel C++ program, using *Wolfram Mathematica* for simplifying the final formulae. All calculations were performed on a supercomputer with 132 GB of RAM and 20 processors.[2] Our results, summarized in Table 7, agree completely with the formulae conjectured in the literature of statistical physics. The program produced data files which document the entire computation, serving as proofs of the formulae. This completes the proof of Theorems 1 and 2.

7. Conclusion

In this paper, we present a theoretical setup and an automatic tool for computing the diagonal formula $\mathrm{DX}(n, n-k)$ for any $k > 0$. Using this setup, we prove the known conjecture about the form of $\mathrm{DX}(n, n-k)$ for a general

[1]The cited reference actually claims that $k + 1$ values are needed, not taking into account that k is a root of the polynomial (except in the first diagonal formula).

[2]The results reported in the conference version of this paper [4] were obtained by running the program on a different computer, hence the difference in the reported running times.

$k=3$	
$\lvert\mathcal{DS}_3\rvert$	147
$\lvert\mathcal{C}_3\rvert$	13
$\mathrm{DT}_{(2,2)}(n)$	$2^{n-6}n^{n-7}(n-3)(n-4)(4n^4-28n^3+97n^2-200n+300)$
$\mathrm{DT}_3(n)$	$2^{n-3}n^{n-7}(n-3)(2n^2-21n^3+106n^2-282n+360)/3$
$\mathrm{DT}(n,n-3)$	$2^{n-3}n^{n-7}(n-3)(2n^4-21n^3+106n^2-282n+360)/3$
$\sum_{i=1}^{13} P_{\mathcal{C}_i}$	$2^{n-6}n^{n-7}(n-3)(n-4)(4n^3-17n^2+11n+70)$
$\mathrm{DX}(n,n-3)$	$2^{n-6}n^{n-7}(n-3)(12n^5-104n^4+360n^3-679n^2+1122n-1560)/3$

$k=4$	
$\lvert\mathcal{DS}_4\rvert$	8,397
$\lvert\mathcal{C}_4\rvert$	179
$\mathrm{DT}_{(2,2,2)}(n)$	$2^{n-7}n^{n-9}(n-4)(n-5)(n-6)(8n^6-84n^5+438n^4-1543n^3+4236n^2$ $-9020n+19040)/6$
$\mathrm{DT}_{2,3}(n)$	$2^{n-4}n^{n-9}(n-4)(n-5)(4n^6-56n^5+383n^4-1654n^3+5106n^2-10920n$ $+14112)/6$
$\mathrm{DT}_4(n)$	$2^{n-5}n^{n-9}(n-4)(4n^6-84n^5+851n^4-5191n^3+20190n^2-47552n$ $+53760)/6$
$\mathrm{DT}(n,n-4)$	$2^{n-7}n^{n-9}(n-4)(8n^8-140n^7+1010n^6-3913n^5+9201n^4-15662n^3$ $+34500n^2-120552n+221760)/6$
$\sum_{i=1}^{179} P_{\mathcal{C}_i}$	$2^{n-7}n^{n-9}(n-4)(n-5)(12n^6-122n^5+373n^4+68n^3-1521n^2-578n$ $+3360)/6$
$\mathrm{DX}(n,n-4)$	$2^{n-7}n^{n-9}(n-4)(8n^8-128n^7+828n^6-2930n^5+7404n^4-17523n^3$ $+41527n^2-114302n+204960)/6$

$k=5$	
$\lvert\mathcal{DS}_5\rvert$	652,060
$\lvert\mathcal{C}_5\rvert$	3,680
$\mathrm{DT}_{(2,2,2,2)}(n)$	$2^{n-12}n^{n-11}(n-5)(n-6)(n-7)(n-8)(16n^8-224n^7+1560n^6$ $-7544n^5+29089n^4-98032n^3+319752n^2-819200n+2324880)/3$
$\mathrm{DT}_{(2,2,3)}(n)$	$2^{n-8}n^{n-11}(n-5)(n-6)(n-7)(8n^8-140n^7+1206n^6-6917n^5$ $+30322n^4-107966n^3+333720n^2-816696n+1321920)/3$
$\mathrm{DT}_{(3,3)}(n)$	$2^{n-7}n^{n-11}(n-5)(n-6)(8n^8-168n^7+1730n^6-11736n^5+59912n^4$ $-238071n^3+722025n^2-1517688n+1814400)/9$
$\mathrm{DT}_{(2,4)}(n)$	$2^{n-8}n^{n-11}(n-5)(n-6)(8n^8-196n^7+2338n^6-17731n^5+95521n^4$ $-384154n^3+1161728n^2-2462976n+2903040)/3$
$\mathrm{DT}_{(5)}(n)$	$2^{n-6}n^{n-11}(n-5)(4n^8-140n^7+2375n^6-25215n^5+183076n^4$ $-932080n^3+3256940n^2-7149000n+7560000)/15$
$\mathrm{DT}(n,n-5)$	$2^{n-9}n^{n-11}(n-5)(240n^{11}-6480n^{10}+73640n^9-461232n^8+1778615n^7$ $-4707195n^6+11632070n^5-41919528n^4+158857920n^3-483329520n^2$ $+1481660640n-2863123200)/360$
$\sum_{i=1}^{3680} P_{\mathcal{C}_i}$	$2^{n-12}n^{n-11}(n-5)(n-6)(32n^9-568n^8+3592n^7-8001n^6-5009n^5$ $+20971n^4+98945n^3+30014n^2-3298664n+9648576)/3$
$\mathrm{DX}(n,n-5)$	$2^{n-12}n^{n-11}(n-5)(240n^{11}-6000n^{10}+62240n^9-356232n^8+1335320n^7$ $-4062240n^6+12397445n^5-42322743n^4+150403080n^3-535510740n^2$ $+1923269040n-3731495040)$

Table 7: Results for $k=3,4,5$.

fixed value of k. As k grows, the number of distinguished structures grows too, and the complexity of the calculations grows as well. We implemented the entire method so that the formulae are obtained completely automatically. As a byproduct, our software also provides a full *proof* of the formula: A complete listing of all structures, all the intermediate computations, a full description of the inclusion-exclusion relations between the structures, and a detailed account of all the calculations. We applied our method to the cases $k \leq 5$, reaffirming the known formulae for $k = 2, 3$ and proving rigorously for the first time the conjectured formulae for $\mathrm{DX}(n, n-4)$ and $\mathrm{DX}(n, n-5)$. Running the program for higher values of k will be done in future work. However, given that already for $k = 5$, the number of distinguished structures surpasses half a million, we do not believe that going beyond $k = 7$ will be feasible.

References

[1] G. Aleksandrowicz and G. Barequet, Counting polycubes without the dimensionality curse, *Discrete Mathematics*, 309 (2009), 576–583.

[2] G. Aleksandrowicz and G. Barequet, The growth rate of high-dimensional tree polycubes, *European J. of Combinatorics*, 35 (2014), 32–38.

[3] A. Asinowski, G. Barequet, R. Barequet, and G. Rote, Proper n-cell polycubes in $n-3$ dimensions, *J. of Integer Sequences*, 15 (2012), #12.8.4.

[4] G. Barequet and M. Shalah, Automatic proofs for formulae enumerating proper polycubes, *Proc. 8th European Conf. on Combinatorics, Graph Theory, and Applications*, Bergen, Norway, August-September 2015, *Electronic Notes in Discrete Mathematics*, 49, 145–151, November 2015.

[5] G. Barequet, M. Moffie, A. Ribó, and G. Rote, Counting polyominoes on twisted cylinders, *INTEGERS: Electronic J. of Combinatorial Number Theory*, 6 (2006), #A22, 37 pp.

[6] G. Barequet, G. Rote, and M. Shalah, $\lambda > 4$: An improved lower bound on the growth constant of polyominoes, *Comm. of the ACM*, 59 (7), July 2016, to appear.

[7] R. Barequet, G. Barequet, and G. Rote, Formulae and growth rates of high-dimensional polycubes, *Combinatorica*, 30 (2010), 257–275.

[8] S.R. Broadbent and J.M. Hammersley, Percolation processes: I. Crystals and mazes, *Proc. Cambridge Philosophical Society*, 53 (1957), 629–641.

[9] B. Derrida and H.J. Herrmann, Collapse of branched polymers, *J. de Physique*, 44 (1983), 1365–1376.

[10] D.S. Gaunt, The critical dimension for lattice animals, *J. of Physics A: Math. Gen.*, 13 (1980), L97–L101.

[11] S. Flesia, D.S. Gaunt, C.E. Soteros, and S.G. Whittington, Statistics of collapsing lattice animals, *J. of Physics A: Mathematical and General*, 27 (1994), 5831–5846.

[12] D.S. Gaunt, M.F. Sykes, and H. Ruskin, Percolation processes in d-dimensions, *J. of Physics A: Mathematical and General*, 9 (1976), 1899–1911.

[13] I. Jensen, Counting polyominoes: A parallel implementation for cluster computing, *Proc. Int. Conf. on Computational Science*, part III, Melbourne, Australia and St. Petersburg, Russia, *Lecture Notes in Computer Science*, 2659, Springer, 203–212, June 2003.

[14] D.A. Klarner, Cell growth problems, *Canadian J. of Mathematics*, 19 (1967), 851–863.

[15] D.A. Klarner and R.L. Rivest, A procedure for improving the upper bound for the number of n-ominoes, *Canadian J. of Mathematics*, 25 (1973), 585–602.

[16] T.C. Lubensky, J. Isaacson, Statistics of lattice animals and dilute branched polymers, *Physical Review A*, **20**, 2130–2146, 1979.

[17] W.F. Lunnon, Symmetry of cubical and general polyominoes, in: *Graph Theory and Computing*(R.C. Read, ed.), Academic Press, New York, NY, 1972, 101–108.

[18] W. F. Lunnon, Counting multidimensional polyominoes, *The Computer Journal*, 18 (1975), 366–367.

[19] S. Luther and S. Mertens, Counting lattice animals in high dimensions, *J. of Statistical Mechanics: Theory and Experiment*, 9 (2011), 546–565.

[20] N. Madras, A pattern theorem for lattice clusters, *Annals of Combinatorics*, 3 (1999), 357–384.

[21] N. Madras, C.E. Soteros, S.G. Whittington, J.L. Martin, M.F. Sykes, S. Flesia, D.S. Gaunt, The free energy of a collapsing branched polymer, *J. of Physics, A: Math. and General*, **23**, 5327–5350, 1990.

[22] J.L. Martin, The impact of large-scale computing on lattice statistics, *J. of Statistical Physics*, 58 (1990), 749–774.

[23] *The On-Line Encyclopedia of Integer Sequences*, available at `http://oeis.org`

[24] P.J. Peard and D.S. Gaunt, $1/d$-expansions for the free energy of lattice animal models of a self-interacting branched polymer, *J. Physics, A: Mathematical and General*, 28 (1995), 6109–6124.

[25] M.F. Sykes and M. Glen, Percolation processes in two dimensions: I. Low-density series expansions, *J. Physics, A: Mathematical and General*, 9 (1976), 87–95.

Appendix

The theorem given in this appendix considers the given structure σ to be fixed, in the sense that if σ is disconnected, then the different ways in which the connected components can share a vertex (as illustrated in Figure 3) are not taken into account, and the number of occurrences of σ, when the components of σ cannot be connected directly by sharing a vertex, are computed. In a sense, this theorem is a special case of Lemma 13.

The original version of our program generated all possible configurations in which the connected components of a disconnected structure σ may be connected, considering every such configuration to be a different distinguished structure. In fact, the number of distinguished structures and cycles reported in Figure 7 include such structures. Clearly, for every disconnected structure σ, summing up the expressions for the number of occurrences for every configuration of σ (provided by the following theorem) yields the exact expression for σ stated in Lemma 13.

Theorem 19. *Let σ be a distinguished structure composed of $c \geq 1$ trees $s_1, \ldots, s_c$ with distinct edge labels $1, \ldots, |\sigma| - c$. The number of occurrences of σ in trees of size n with distinct edge labels $1, \ldots, n-1$ is*

$$F_n(\sigma) = \tfrac{(n-|\sigma|+c-1)!}{(n-|\sigma|)!} n^{n-|\sigma|+c-2} \prod_{i=1}^{c} |s_i|.$$

Proof: We proceed by double counting, enumerating in two ways the different sequences of directed edges that can be added to a graph composed of the union of $n-|\sigma|$ vertices and the distinguished structure σ, so as to form a rooted tree with n vertices.

One way to count these sequences is to add the edges one by one, and to count the number of options available at each step. There are $\mathcal{N} = \prod_{i=1}^{c} |s_i|$ ways to choose a root for each component s_i of σ. At the beginning, we have a forest with $n-|\sigma|+c$ rooted trees. After adding a collection of edges, forming a rooted forest with i trees, there are $n(i-1)$ options for the next edge to add: Its origin can be any one of the n vertices of the graph, and its terminus can be any one of the $i-1$ roots other than the root of the tree containing the origin. Therefore, the total number of options is

$$\mathcal{N} \prod_{i=2}^{n-|\sigma|+c} n(i-1) = \mathcal{N} n^{n-|\sigma|+c-1}(n-|\sigma|+c-1)!. \tag{4}$$

An alternative way to count these edge sequences is to start with one of the $F_n(S)$ possible unrooted edge-labeled trees which contains σ, choose one of its n vertices as a root, and choose one of the $(n-|\sigma|)!$ possible sequences, say, η, then label the $(n-|\sigma|)$ vertices of the tree according to η (the vertices that do not belong to σ), and "shift" each vertex-label to the incident edge towards the root, producing an edge-labeled tree. The total number of sequences that can be formed this way is

$$nF_n(\sigma)(n-|\sigma|)!. \tag{5}$$

Finally, we conclude from Equations (4) and (5) that the number of occurrences of σ in unrooted trees with edge labels $1, \ldots, n-1$ is

$$F_n(\sigma) = \frac{(n-|\sigma|+c-1)!}{(n-|\sigma|)!} n^{n-|\sigma|+c-2} \mathcal{N}.$$

$\square$