Journal of Integer Sequences, Vol. 15 (2012),
Article 12.8.4

Proper n-Cell Polycubes in $n - 3$ Dimensions

Andrei Asinowski
Department of Mathematics
The Technion—Israel Institute of Technology
Haifa 32000
Israel
andrei@tx.technion.ac.il

Gill Barequet
Department of Computer Science
The Technion—Israel Institute of Technology
Haifa 32000
Israel
barequet@cs.technion.ac.il

Ronnie Barequet
Department of Computer Science
Tel Aviv University
Tel Aviv 69978
Israel
ronnieba@post.tau.ac.il

Günter Rote
Institut für Informatik
Freie Universität Berlin
Takustraße 9
D-14195 Berlin
Germany
rote@inf.fu-berlin.de

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. Enumeration of polycubes, and, in particular, specific types of polycubes, as well as computing the asymptotic growth rate of polycubes, is a popular problem in combinatorics and discrete geometry. This is also an important tool in statistical physics for computations and analysis of percolation processes and collapse of branched polymers. A polycube is said to be *proper* in d dimensions if the convex hull of the centers of its cubes is d-dimensional. In this paper we prove that the number of polycubes of size n that are proper in $n-3$ dimensions is $2^{n-6}n^{n-7}(n-3)(12n^5-104n^4+360n^3-679n^2+1122n-1560)/3$.

1 Introduction

A d-dimensional *polycube* of size n is a connected set of n cubical cells on the lattice $\mathbb{Z}^d$, where connectivity is through $(d-1)$-faces. Two *fixed* polycubes are considered equivalent if one can be transformed into the other by a translation. A polycube is called *proper* in d dimensions if the convex hull of the centers of all its cubes is d-dimensional. While in the mathematical literature these objects are called polycubes (*polyominoes* in two dimensions), they are usually referred to as (strongly-embedded) *lattice animals* in the literature of statistical physics. Following Lunnon [7], we let $\mathrm{DX}(n,d)$ denote the number of n-cell fixed polycubes that are proper in d dimensions.

Counting polycubes (or animals) is a long-standing problem in discrete geometry, originating in statistical physics [3]. To-date, no formula is known for $A_d(n)$, the number of polycubes of size n in d dimensions, for any fixed value of d, not to mention the general case. Klarner [5] showed that the limit $\lambda_2 = \lim_{n\to\infty} \sqrt[n]{A_2(n)}$ exists. Thirty two years have passed until Madras [9] proved the convergence of the sequence $A_2(n+1)/A_2(n)$ to λ_2 (and, in fact, a similar claim in any fixed dimension d), as n tends to infinity. Thus, λ_2 is the *growth rate* limit (also known as the *connective constant*) of polyominoes. Its exact value has remained elusive till these days. The currently best lower and upper bounds known on λ_2 are roughly 3.9801 [1] and 4.6496 [6], respectively.

Much less is known in higher dimensions, let alone in a general dimension d. Significant progress has been obtained along the years in the literature of statistical physics, although the computations usually relied on unproven assumptions or on formulae which were interpolated empirically from a few known values of $A_d(n)$. The expansion

$$\ln \lambda_d = \ln \sigma + 1 - \frac{2}{\sigma} - \frac{79}{24\sigma^2} - \frac{317}{24\sigma^3} - \frac{18321}{320\sigma^4} - \frac{123307}{240\sigma^5} + O(\frac{1}{\sigma^6}),$$

where $\sigma = 2d-1$ (one less than the *coordination number* of the cubical lattice $\mathbb{Z}^d$), is provided by Gaunt and Peard [4]. This $1/d$-expansion (of the *free energy* of animals, in their terminology) is partly based on so-called "diagonal formulae," that is, formulae for $\mathrm{DX}(n, n-k)$, where $k > 0$ is a small constant. It turned out that this expansion is consistent with the main result obtained by Barequet et al. [2], namely, that the growth-rate limit of the number of polycubes in d dimensions is asymptotically $2ed - o(d)$, conjectured to asymptotically be $(2d-3)e - \frac{31e}{48d} + O(\frac{1}{d^2})$.

In the literature of statistical physics [8, 11], formulae for $\mathrm{DX}(n, n-k)$ are interpolated for $1 \leq k \leq 7$. (See also a discussion by Barequet et al. [2, p. 265] about how to derive these formulae from more general formulae provided by Peard and Gaunt [11].) It is rather easy to show, using Cayley trees, that

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

(sequence A127670 in the On-Line Encyclopedia of Integer Sequences [10]). Barequet et al. [2] 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 in the on-line encyclopedia of integer sequences). 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.

In this paper we find the explicit formula for $\mathrm{DX}(n, n-3)$, the number of n-cell polycubes that are proper in $d = n-3$ dimensions (sequence A191092 in the on-line encyclopedia of integer sequences), stated in the following theorem.

Theorem 1.

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

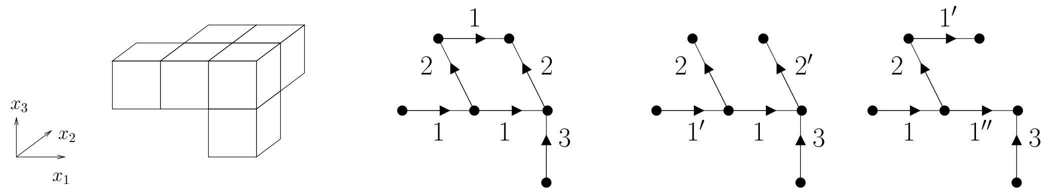
Similarly to our proof [2] of the formula for $\mathrm{DX}(n, n-2)$, we prove the formula for $\mathrm{DX}(n, n-3)$ by counting spanning trees of polycubes, yet the reasoning and the calculations are significantly more involved. Indeed, we use the inclusion-exclusion principle in order to count correctly polycubes whose cell-adjacency graphs contain cycles. Spanning trees of such polycubes contain certain subgraphs, which we call "distinguished structures." In comparison with the case $k = 2$, the number of such structures is substantively higher, and the ways in which they can appear in spanning trees are much more varied. Therefore, the proof of Theorem 1 provides us with a better understanding of the difficulties that can be expected in applying this technique for higher values of k. In particular, we believe that it is not practical to achieve a similar proof manually for $k > 3$.

2 Overview of the Method

Denote by $\mathcal{P}_n$ the family of polycubes under consideration.

Let $P \in \mathcal{P}_n$, and let $\gamma(P)$ denote the directed graph with labeled edges that is constructed as follows: The vertices of $\gamma(P)$ correspond to 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 ($1 \leq i \leq n-3$) if the corresponding cells have different i-coordinate (that is, 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.

It is clear that $P \mapsto \gamma(P)$ is an injection. Thus, it suffices to count the graphs obtained from the members of $\mathcal{P}_n$ in this way. We shall accomplish this task by counting their spanning trees.



(a) Polycube (b) Adjacency graph (c) Two spanning trees

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

Consider a spanning tree of $\gamma(P)$. It has $n-1$ edges labeled by numbers from the set $\{1, 2, \ldots, n-3\}$; all these labels actually are present, otherwise the polycube is not proper. Thus, either there are two edge labels (say, i and j) that appear twice, or there is one edge label (say, i) that appears three times. In the former case we distinguish members of such pairs by labeling them i, i' and j, j', while in the latter—by labeling them i, i', i'' (see Figure 1(c)). Whenever we consider a spanning tree of $\gamma(P)$, we assume that its repeated labels are distinguished this way. In contrast, when considering $\gamma(P)$, repeated labels are assumed not to be distinguished (as in Figure 1(b)).

Observation 2. Every label must occur an even number of times in any cycle of $\gamma(P)$.

In addition, the number of cycles in $\gamma(P)$, as well as the length of each such cycle, are limited due to the limited multiplicity of labels. Therefore, we must have the following:

Observation 3. There are three possible cases concerning the structure of $\gamma(P)$:

1. $\gamma(P)$ is a tree itself. The number of such *tree-like* polycubes in $\mathcal{P}_n$ will be denoted by $\mathbf{X}$.

2. $\gamma(P)$ has exactly one cycle. The number of such polycubes in $\mathcal{P}_n$ will be denoted by $\mathbf{Y}$. The length of this cycle in $\gamma(P)$ is either 4 or 6.

3. $\gamma(P)$ has two 4-cycles. The number of such polycubes in $\mathcal{P}_n$ will be denoted by $\mathbf{Z}$. In this case, the two cycles are either edge-disjoint or have exactly one common edge, thus forming a 6-cycle with a chord.

Indeed, all other possibilities (say, more than two 4-cycles, several 6-cycles, or an 8-cycle) are excluded: They would cause $\gamma(P)$ to have a spanning tree with too many repeated labels.

Thus, $\mathrm{DX}(n, n-3) = \mathbf{X} + \mathbf{Y} + \mathbf{Z}$. In order to find the formulae for $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, we shall count those directed trees with $n-1$ labeled edges with two pairs of repeated labels or with one triple of repeated labels, which are actually spanning trees of $\gamma(P)$ for some polycube $P \in \mathcal{P}_n$. Note that:

- If a tree with two pairs of repeated labels, i, i' and j, j', is a spanning tree of $\gamma(P)$ for some polycube $P \in \mathcal{P}_n$, then the trees obtained by exchanging i and i' and/or j and j' are also spanning trees of $\gamma(P)$. Similarly, if a tree with a triple of repeated labels, i, i', i'', is a spanning tree of a $\gamma(P)$, then the trees obtained by permuting i, i' and i'' are also spanning trees of $\gamma(P)$.

- In particular, if $\gamma(P)$ itself is a tree with two pairs of repeated labels, then it has four spanning trees.[1] If $\gamma(P)$ is a tree with a triple of repeated labels, then it has six spanning trees.

- The situation may be more complicated when $\gamma(P)$ is not a tree at all. There are polycubes P such that $\gamma(P)$ has both types of spanning trees: those with two pairs of repeated labels and those with a triple of repeated labels. (Such a polycube is shown in Figure 1.)

In the next section we characterize all substructures that are present in some trees with labeled edges due to the fact that the number of cells is greater than the number of dimensions. By analyzing this substructures, we will be able to compute how many of such trees actually represent polycubes. Then, in the following sections, we develop formulae for the numbers of all possible spanning trees of the polycubes, and then derive the actual number of polycubes.

3 Distinguished Structures

Our plan is to count polycubes by counting spanning trees of their adjacency graphs, taking into account possible multiplicities. In the reasoning below we shall consider several small structures, which may be contained in the spanning trees that we count. These structures are listed in Figure 2, and they are interesting for the following reason. For each labeled tree, we can attempt to build the corresponding polycube. Two things may happen: (a) We may get coinciding cells, like in patterns A or I (shown by dotted frames around these points). Such a tree is invalid and does not correspond to a polycube. (b) Two cells which are not connected by a tree edge may be adjacent, like in pattern B or C (indicated by dotted lines). Such a tree may correspond to a valid polycube, but it deserves special attention because the polycube has cycles in its cell-adjacency graph and, therefore, its spanning tree is not unique.

A *distinguished structure* is defined as a subtree that is responsible for the presence of two coinciding or adjacent cells, as explained above. More precisely, a structure is the union of all paths (edges plus incident vertices) that run between two coinciding or adjacent cells.

Consider a distinguished structure that leads to coinciding cells. Similarly to Observation 2, we see that for every label on the path between two vertices that correspond to such cells, repetitions of this label occur on this path an even number of times (which can be only 2). Moreover, in this case, these two edges are directed in opposite orientations. Due to the limited number of repeated labels, we get only two possibilities corresponding to a path of length 2 and a path of length 4 (see structures A and I in Figure 2).

Consider now a structure that leads to a non-existing adjacency. This clearly results in an (even) cycle with one edge removed. By Observation 3, the length of such a cycle can be only either 4 or 6. A reasoning similar to that above leads us to two possibilities: structures B and C in Figure 2.

[1] Recall that repeated labels are not distinguished in $\gamma(P)$, but they are distinguished in its spanning tree.

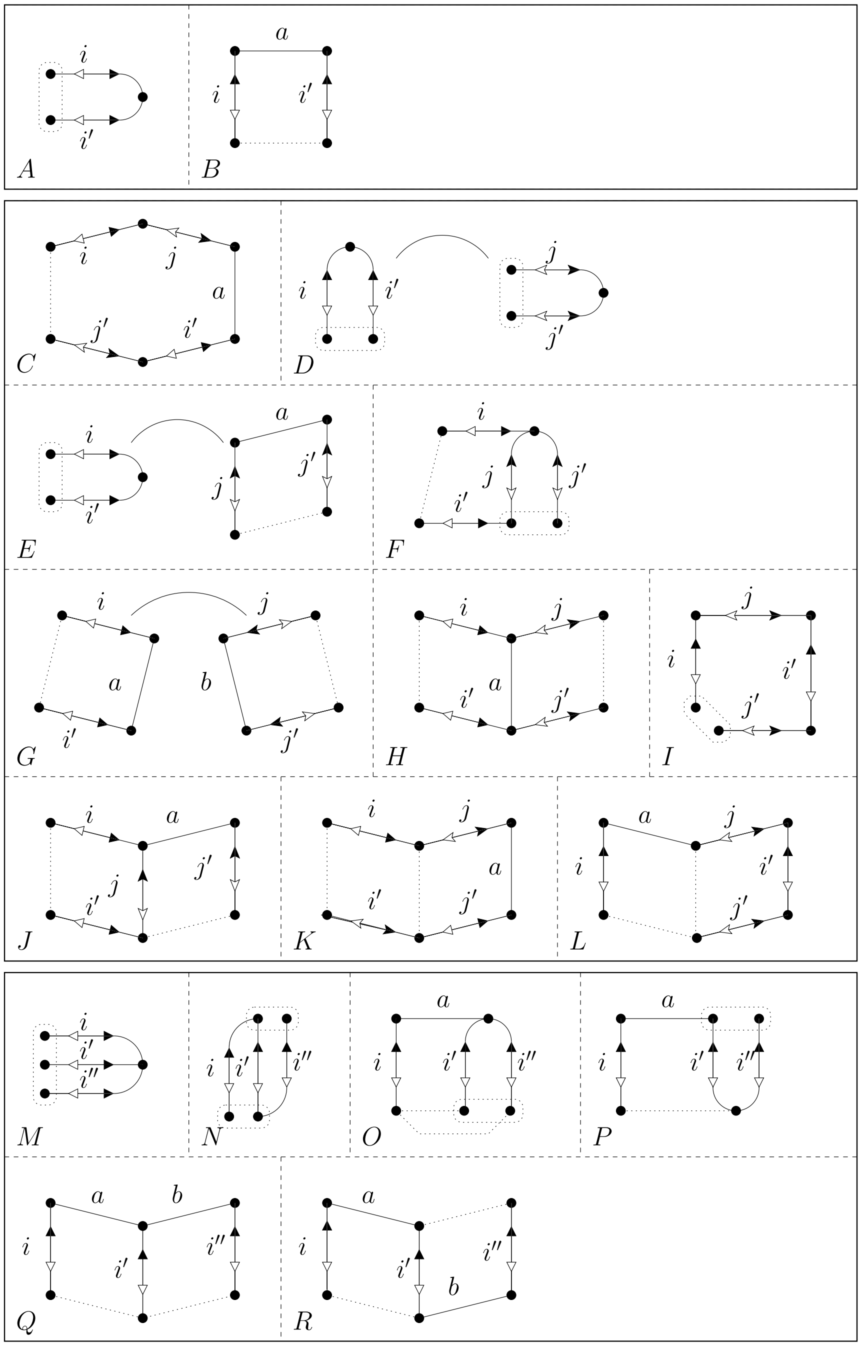


Figure 2: Distinguished structures used in the counting.

Thus, the distinguished structures are A, I, B, C, and other structures that contain several occurrences of these "basic" structures.[2] The number of occurrences is limited, since each occurrence uses up some repeated labels. The enumeration of the distinguished structures is, thus, a finite task. Figure 2 gives the complete list.

It may happen that a distinguished structure is disconnected, like D, E, or G. We consider the components of a structure as edge-connected components; thus, it is permitted that the two parts in D, E, or G share a vertex.

The structures $C, \dots, L$ occur only in trees with two pairs of repeated labels, while the structures $M, \dots, R$ occur only in trees with one triple of repeated labels. In contrast, A and B occur in both kinds of trees.

Next, we clarify the notation and conventions used in Figure 2. Each pattern in Figure 2 stands actually for several substructures that may differ in edge directions or the precise choice of labels from i, i', i'' or j, j'. The edges labeled i and i' (respectively, i, i', and i'') are all directed either according to black or to white arrows, and the same holds for j and j'. The directions of j, j' are independent of those of i, i'. The labels a, b are assumed to be unrelated to any of the other labels i, i', i'', j, j' appearing in each pattern. For example, a in pattern B is distinct from i and i' (but it could be another repeated label—say, j or j'). Where a or b appear in patterns $C, \dots, L$, they are automatically distinct from i, i', j, j'. Finally, in the remaining patterns $M, \dots, R$, a and b are automatically distinct from i, i', i''. Variations of the same label, say, i and i', can be permuted, or replaced by other variations of the same label, like i, i'' or i', i''.

In counting directed trees with $n-1$ labeled edges, which have subgraphs as in Figure 2, two lemmas will be used. Lemma 4 was proved earlier [2]; we will here relate it to a result from the literature.

Lemma 4. *[2, Lemma 4] The number of ordered sequences $T = (\tau_1, \dots, \tau_k)$ of $k \geq 1$ rooted trees with a total of $n-k$ edges and distinct edge labels $1, \dots, n-k$ is $n^{n-k-1}k$.*

Proof. Consider such a sequence T. It has a total of n vertices. For $i = 1, \dots, k$, denote the root of the component τ_i by $n-k+i$. For any vertex v which is not a root, set its label to be equal to the label of the first edge in the path from v to the root.

Now we have a forest T' on n labeled vertices, with roots labeled by $n-k+1, n-k+2, \dots, n$. The correspondence $T \leftrightarrow T'$ is clearly a bijection. The number of forests on n labeled vertices with k roots, whose labels belong to a specified set, is known to be $n^{n-k-1}k$, see Stanley [12, p. 25, Proposition 5.3.2]. (Stanley provides two proofs of this, both of them differing from our proof [2].) □

The other lemma is a direct application of the previous lemma.

Lemma 5. *The number of ordered sequences $\tilde{T} = (\tau_1, \dots, \tau_k)$ of $k \geq 1$ trees, such that τ_1 has two distinguished roots (which may coincide) and all other trees have one root, with a total of $n-k$ edges and distinct edge labels $1, \dots, n-k$, is n^{n-k}.*

Proof. Consider a sequence T as in Lemma 4, and mark an arbitrary vertex as the extra root. In this way $n^{n-k}k$ sequences $\tilde{T}$ are obtained. The component of $\tilde{T}$ that has two roots is

[2] Notice that I itself contains two occurrences of B.

any of $\tau_1, \ldots, \tau_k$, with equal probability. Therefore, in order to get the number of sequences $\tilde{T}$ in which the component with two roots is τ_1, we have to divide by k, obtaining n^{n-k}. $\square$

We now introduce two functions that count ordered sequences of directed trees. Let $F_1(k)$ denote the number of ordered sequences $(\tau_1, \ldots, \tau_k)$ of $k \geq 1$ *directed* rooted trees with a total of $n-k$ edges and distinct edge labels $1, \ldots, n-k$. Similarly, let $F_2(k)$ denote the number of ordered sequences $(\tau_1, \ldots, \tau_k)$ of $k \geq 1$ *directed* trees, such that τ_1 has two distinguished roots (which may coincide) and all other trees have one root, with a total of $n-k$ edges and distinct edge labels $1, \ldots, n-k$. By fixing directions to the edges of undirected trees, we obtain the following corollary of Lemmas 4 and 5.

Corollary 6.

1. $F_1(k) = 2^{n-k} n^{n-k-1} k$.
2. $F_2(k) = 2^{n-k} n^{n-k}$.

Finally, we will use Corollary 8, which follows directly from a previous result of ours [2].

Lemma 7. *[2, 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 $n \geq 2$.* $\square$

Let $\mathbf{T}_{22}$ denote the number of directed trees with n vertices and labeled edges, with two pairs of repeated labels.[3] Similarly, let $\mathbf{T}_3$ denote the number of directed trees with n vertices and labeled edges, with one triple of repeated labels.

Corollary 8.

1. $\mathbf{T}_{22} = \binom{n-3}{2} 2^{n-1} n^{n-3}$.
2. $\mathbf{T}_3 = (n-3) 2^{n-1} n^{n-3}$.

Let us turn to the enumeration of occurrences of $A, \ldots, R$ from Figure 2 in directed trees with n vertices and edges labeled $1, 2, \ldots, n-3$, and with repeated labels as explained above. For $C, D, E, \ldots$, we let $\mathbf{C}, \mathbf{D}, \mathbf{E}, \ldots$ denote the number of occurrences of these structures in such trees. Recall that the structures $C, \ldots, L$ occur only in trees with two pairs of repeated labels, while the structures $M, \ldots, R$ occur only in trees with one triple of repeated labels. In contrast, A and B occur in both kinds of trees. Therefore, for A and B we shall consider both cases (two pairs of repeated labels, and one triple of repeated labels), denoting the corresponding numbers by $\mathbf{A}_{22}$, $\mathbf{A}_3$, and $\mathbf{B}_{22}$, $\mathbf{B}_3$. For some cases we shall explain in detail how the formula is obtained; all other calculations are based on a similar reasoning.

We begin with counting *occurrences* of $A_{22}, B_{22}, C, \ldots, L$ in directed trees with two pairs of repeated labels. By an *occurrence* of a pattern U in a tree T, we mean a pair (S, T), where S is a subset of edges of T that form a the pattern U.

[3] Recall again that repeated labels in trees are distinguished.

1. $\mathbf{A}_{22} = (n-3) \cdot (n-4) \cdot 2 \cdot F_1(3) = 6(n-3)(n-4)2^{n-3}n^{n-4}$.
Here, we have the factors $(n-3)$ for choosing the repeated label (i, i') that makes the configuration and $(n-4)$ for choosing the second repeated label in the tree, a factor 2 for directing the edges i and i' (i.e., both according to the black arrows or the white arrows), and $F_1(3)$ for sequences of three trees that can be attached to the vertices. At this stage, the three vertices are distinguishable from each other, and therefore we count *sequences* of three trees. The same will be true for all other patterns.

2. $\mathbf{B}_{22} = (n-3) \cdot (n-4) \cdot 2 \cdot (n-3) \cdot 2 \cdot F_1(4) = (n-3)^2(n-4)2^n n^{n-5}$.
Here, the factors $(n-3)$, $(n-4)$, and 2 are the same as in the previous case, an additional factor $(n-3)$ is for choosing label a, an additional factor 2 is for directing the edge a, and $F_1(4)$ is for sequences of four trees that can be attached to the vertices.

3. $\mathbf{C} = \binom{n-3}{2} \cdot (n-5) \cdot 4 \cdot 2 \cdot 4 \cdot F_1(6) = 3(n-3)(n-4)(n-5)2^{n-1}n^{n-7}$.
Here, the factor $\binom{n-3}{2}$ is for choosing the repeated labels, $(n-5)$ is for choosing edge a, a factor 4 is for choosing which edge among i, i', j, j' is attached to the head of edge a, and, once this choice is made, a factor 2 is for choosing which of the complementary label is attached to the tail of a (e.g., if the first choice is i, then the second choice can be only j or j'), an additional factor 4 is for directing i, i' and j, j', and $F_1(6)$ is for sequences of six trees that can be attached to the vertices.

4. $\mathbf{D} = \binom{n-3}{2} \cdot 4 \cdot F_2(5) \cdot 9 = 9(n-3)(n-4)2^{n-4}n^{n-5}$.
Here, the factors $\binom{n-3}{2}$ and 4 are as in the previous case. The factor $3 \cdot 3 = 9$ stands for choosing the pair of vertices through which the components are connected: one vertex is chosen on each component. The factor $F_2(5)$ is for sequences of five trees that can be attached to the vertices (one of which connects the components of the configuration).

5. $\mathbf{E} = (n-3) \cdot (n-4) \cdot (n-5) \cdot 8 \cdot F_2(6) \cdot 12 = 3(n-3)(n-4)(n-5)2^{n-1}n^{n-6}$.
As above, we have the factors $(n-3)$, $(n-4)$, and $(n-5)$ for choosing the repeated labels and the edge a, a factor 8 for directing for directing i, i', j, j', and a. The factor $3 \cdot 4 = 12$ stands for choosing the pair of vertices through which the components are connected. The factor $F_2(6)$ stands for sequences of six trees that can be attached to the vertices (one of which connects the two components of the configuration).

6. $\mathbf{F} = (n-3) \cdot (n-4) \cdot 2 \cdot 2 \cdot 4 \cdot F_1(5) = 5(n-3)(n-4)2^{n-1}n^{n-6}$.

7. $\mathbf{G} = \binom{n-3}{2} \cdot (n-5) \cdot (n-6) \cdot 16 \cdot F_2(7) \cdot 16 = (n-3)(n-4)(n-5)(n-6)2^n n^{n-7}$.

8. $\mathbf{H} = \binom{n-3}{2} \cdot (n-5) \cdot 4 \cdot 4 \cdot F_1(6) = 3(n-3)(n-4)(n-5)2^{n-2}n^{n-7}$.

9. $\mathbf{I} = \binom{n-3}{2} \cdot 4 \cdot 4 \cdot F_1(5) = 5(n-3)(n-4)2^{n-2}n^{n-6}$.

10. $\mathbf{J} = (n-3) \cdot (n-4) \cdot (n-5) \cdot 2 \cdot 4 \cdot 4 \cdot F_1(6) = 3(n-3)(n-4)(n-5)2^n n^{n-7}$.

11. $\mathbf{K} = (n-3) \cdot (n-4) \cdot (n-5) \cdot 4 \cdot 4 \cdot F_1(6) = 3(n-3)(n-4)(n-5)2^{n-1}n^{n-7}$.

12. $\mathbf{L} = (n-3) \cdot (n-4) \cdot (n-5) \cdot 2 \cdot 4 \cdot 4 \cdot F_1(6) = 3(n-3)(n-4)(n-5)2^n \cdot n^{n-7}$.

Next, we count occurrences of $A_3, B_3, M, \ldots, R$ in directed trees with one triple of repeated labels.

13. $\mathbf{A}_3 = (n-3) \cdot 3 \cdot 2 \cdot F_1(3) = 9(n-3)2^{n-2}n^{n-4}$.
    Here, the factor $(n-3)$ is for choosing the repeated label (i, i', and i''), the factor 3 is for choosing two labels from $\{i, i', i''\}$, the factor 2 is for directing these edges, and $F_1(3)$ is for sequences of three trees attached to the vertices.

14. $\mathbf{B}_3 = (n-3) \cdot 3 \cdot (n-4) \cdot 2 \cdot 2 \cdot F_1(4) = 3(n-3)^2 2^n n^{n-5}$.
    Here, we have the factors $(n-3)$ and 3 as in the previous case, an additional factor $(n-4)$ for choosing the edge a (which is not a repetition of i), a factor 2 for directing the edges i and i', an additional factor 2 for choosing which edge with repeated label is attached to the head of edge a, and $F_1(4)$ for sequences of four trees attached to the vertices.

15. $\mathbf{M} = (n-3) \cdot 2 \cdot F_1(4) = (n-3)2^{n-1}n^{n-5}$.

16. $\mathbf{N} = (n-3) \cdot 3 \cdot 2 \cdot F_1(4) = 3(n-3)2^{n-1}n^{n-5}$.
    As above, the factor $(n-3)$ is for choosing the repeated label (i, i', and i''), the factor 3 is for choosing which of these labels is found between the two others, the factor 2 is for directing these edges, and $F_1(4)$ is for sequences of four trees attached to the vertices.

17. $\mathbf{O} = (n-3) \cdot (n-4) \cdot 2 \cdot 3 \cdot 2 \cdot F_1(5) = 15(n-3)(n-4)2^{n-3}n^{n-6}$.
    As in previous cases, the factor $(n-3)$ is for choosing the repeated label (i, i', and i''), $(n-4)$ is for choosing the edge a, the factor 2 is for choosing whether one or two edges with repeated labels are attached to the head of a, the factor 3 is for choosing which edge (or edges, depending on the previous choice) among i, i', i'' is (resp., are) attached to the head of a, the factor 2 is for directing the edges i, i', i'', and $F_1(5)$ is for sequences of five trees attached to the vertices.

18. $\mathbf{P} = (n-3) \cdot (n-4) \cdot 2 \cdot 3 \cdot 2 \cdot 2 \cdot F_1(5) = 15(n-3)(n-4)2^{n-2}n^{n-6}$.

19. $\mathbf{Q} = (n-3) \cdot 3 \cdot (n-4) \cdot (n-5) \cdot 2 \cdot 4 \cdot F_1(6) = 9(n-3)(n-4)(n-5)2^{n-2}n^{n-7}$.

20. $\mathbf{R} = (n-3) \cdot 3 \cdot (n-4) \cdot (n-5) \cdot 2 \cdot 4 \cdot F_1(6) = 9(n-3)(n-4)(n-5)2^{n-2}n^{n-7}$.

Armed with these formulae, we can now count the various possible types of polycubes of size n that are proper in $n-3$ dimensions.

4 Polycubes with a Tree Structure

We split the counting according to the combinations of repeated labels in the tree.

Two pairs of repeated labels. Denote by $\mathbf{X}_{22}$ the number of polycubes $P \in \mathcal{P}_n$, such that $\gamma(P)$ is a tree that has two pairs of repeated labels. By Corollary 8.1, the total number of directed trees with n vertices and directed labeled edges, with two pairs of repeated labels, is $\mathbf{T}_{22} = \binom{n-3}{2}2^{n-1}n^{n-3}$. Such a tree corresponds to a tree-like polycube in $\mathcal{P}_n$ unless it contains a subtree of type $A, \ldots, J$. Thus, subtrees of types A, B, or C are all that we need to exclude. However, each of $D, \ldots, J$ includes *two* subtrees of the type A, B, or C, and these are counted twice in the sum $\mathbf{A}_{22} + \mathbf{B}_{22} + \mathbf{C}$. Therefore, the number of trees which do not lead to tree-like polycubes is

$$\mathbf{A}_{22} + \mathbf{B}_{22} + \mathbf{C} - \mathbf{D} - \mathbf{E} - \mathbf{F} - \mathbf{G} - \mathbf{H} - \mathbf{I} - \mathbf{J}$$

(Each of the patterns K and L contains exactly one subtree of type A, B, or C,[4] and hence, they are correctly accounted for.)

Dividing by 4 (since each such polycube is represented by four trees), we obtain that

$$\begin{aligned}\mathbf{X}_{22} &= (\mathbf{T}_{22} - \mathbf{A}_{22} - \mathbf{B}_{22} - \mathbf{C} + \mathbf{D} + \mathbf{E} + \mathbf{F} + \mathbf{G} + \mathbf{H} + \mathbf{I} + \mathbf{J})/4 \\ &= 2^{n-6}n^{n-7}(n-3)(n-4)(4n^4 - 28n^3 + 97n^2 - 200n + 300). \qquad (1)\end{aligned}$$

One triple of repeated labels. Denote by $\mathbf{X}_3$ the number of polycubes $P \in \mathcal{P}_n$, such that $\gamma(P)$ is a tree that has one triple of repeated labels. By Corollary 8.2, the total number of directed trees with n vertices and directed labeled edges, with one triple of repeated labels, is $\mathbf{T}_3 = (n-3)2^{n-1}n^{n-3}$. Such a tree corresponds to a polycube in $\mathcal{P}_n$ unless it has a subtree of type either A or B. In addition, all of $M, \ldots, R$ include (at least) two subtrees of the types A or B. The types M and O even include three subtrees of the types A or B. Therefore, by applying the inclusion-exclusion principle, and dividing by 6 (since each such polycube is represented by six trees), we obtain that

$$\begin{aligned}\mathbf{X}_3 &= (\mathbf{T}_3 - (\mathbf{A}_3 + \mathbf{B}_3) + (3\mathbf{M} + \mathbf{N} + 3\mathbf{O} + \mathbf{P} + \mathbf{Q} + \mathbf{R}) - (\mathbf{M} + \mathbf{O}))/6 \\ &= 2^{n-3}n^{n-7}(n-3)(2n^4 - 21n^3 + 106n^2 - 282n + 360)/3. \qquad (2)\end{aligned}$$

In total, we have

$$\mathbf{X} = \mathbf{X}_{22} + \mathbf{X}_3.$$

5 Polycubes with One Cycle

As mentioned above, if $\gamma(P)$ has only one cycle, then the length of this cycle must be either 4 or 6.

Assume first that $\gamma(P)$ has one 4-cycle whose edges are labeled i, j, i, j. Then, either $\gamma(P)$ has another edge with the label i or j, or it has no such edge; in the latter case $\gamma(P)$ has another pair of edges with a repeated label which is distinct from i and j. Denote, therefore, the number of polycubes of the former type by $\mathbf{Y}_{23}$, and the latter type by $\mathbf{Y}_{222}$.

Each graph of the first type has two spanning trees with two pairs of repeated edge labels and two spanning trees with a triple of repeated edge labels. Each graph of the second type has four spanning trees with two pairs of repeated edge labels.

[4] In fact, the pattern C is contained in no other distinguished structure.

Denote by $\mathbf{T}_{22}^1$ the total number of spanning trees of these graphs (of both types), which have two pairs of repeated labels, and by $\mathbf{T}_3^1$ the number of spanning trees of these graphs (necessarily of the first type), which have one triple of repeated labels. Then, we have

$$\mathbf{T}_{22}^1 = 4(2\mathbf{Y}_{23} + 4\mathbf{Y}_{222}) \quad \text{and} \quad \mathbf{T}_3^1 = 6(2\mathbf{Y}_{23}). \tag{3}$$

Spanning trees with two pairs of repeated labels. All these spanning trees have a single occurrence of B as a subtree. Thus, the number of occurrences of B in all such trees is $\mathbf{B}_{22}$. From this number we have to subtract the number of occurrences of the forbidden subtrees E, F, and I; the number of spanning trees of graphs that have two edge-disjoint 4-cycles (that is, G); and also the number of spanning trees (with two pairs of repeated labels) of graphs that have two 4-cycles with a common edge (that is, H, J, K, and L). Notice that in $\mathbf{B}_{22}$, trees with I, G, H, or J are counted twice. Therefore,

$$\begin{aligned}\mathbf{T}_{22}^1 &= \mathbf{B}_{22} - \mathbf{E} - \mathbf{F} - 2\mathbf{I} - 2\mathbf{G} - 2\mathbf{H} - 2\mathbf{J} - \mathbf{K} - \mathbf{L} \\ &= 2^{n-1}n^{n-6}(n-3)(n-4)(2n^2 - 13n + 25).\end{aligned} \tag{4}$$

Spanning trees with one triple of repeated labels. In this case, possible spanning trees have a subtree B (recall that a is neither i, i', nor i''). The number of occurrences of B in all such trees is $\mathbf{B}_3$.

From this number we have to subtract the number of occurrences of the forbidden subtrees O and P, and the number of spanning trees (with a triple of repeated labels) of graphs that have two 4-cycles with a common edge (that is, Q and R). Note that in $\mathbf{B}_3$, trees with O, Q, or R are counted twice. Therefore,

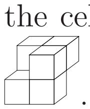
$$\begin{aligned}\mathbf{T}_3^1 &= \mathbf{B}_3 - 2\mathbf{O} - \mathbf{P} - 2\mathbf{Q} - 2\mathbf{R} \\ &= 3 \cdot 2^{n-1}n^{n-7}(n-3)(n-4)(2n^2 - 11n + 30).\end{aligned} \tag{5}$$

By solving the system (3–5), we obtain

$$\begin{aligned}\mathbf{Y}_{23} &= 2^{n-3}n^{n-7}(n-3)(n-4)(2n^2 - 11n + 30), \\ \mathbf{Y}_{222} &= 2^{n-5}n^{n-7}(n-3)(n-4)(n-5)(2n^2 - 7n + 12).\end{aligned}$$

Polycubes with a 6-cycle. Denote the number of polycubes in $\mathcal{P}_n$ that have only one cycle of length 6 by $\mathbf{Y}_{\hexagon}$. We have

$$\mathbf{Y}_{\hexagon} = \binom{n-3}{3} \cdot 4 \cdot F_1(6) = 2^{n-4}n^{n-7}(n-3)(n-4)(n-5).$$

In order to establish this quantity, we do not need to consider spanning trees. Notice that the cells of a 6-cycle always form a $2 \times 2 \times 2$ cube with two opposite cells removed, as in . Thus, $\binom{n-3}{3}$ counts the dimensions in which the cube lies, the factor 4 counts ways to remove two reciprocal cells, and $F_1(6)$ counts ways to attach six trees to the remaining six cells.

In total, we have $\mathbf{Y} = \mathbf{Y}_{23} + \mathbf{Y}_{222} + \mathbf{Y}_{\hexagon}$.

6 Polycubes with Two Cycles

Again, we do not consider spanning trees.

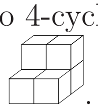
Two 4-cycles without a common edge. The number of polycubes $P \in \mathcal{P}_n$, such that $\gamma(P)$ has two 4-cycles without a common edge, is

$$\mathbf{Z}_{\diamond\!-\!\diamond} = \frac{1}{2} \cdot \binom{n-3}{2} \cdot \binom{n-5}{2} \cdot 16 \cdot F_2(7) = 2^{n-6} n^{n-7} (n-3)(n-4)(n-5)(n-6).$$

In this case, $\frac{1}{2}\binom{n-3}{2}\binom{n-5}{2}$ is the number of ways to choose the dimensions in which the "squares" lie (note that these pairs of dimensions are disjoint, otherwise $\gamma(P)$ would have a spanning tree with four repeated labels). The factor $4 \cdot 4 = 16$ is the number of ways to choose a pair of vertices through which the squares will be connected, and $F_2(7)$ is the number of ways to connect them by trees.

Two 4-cycles with a common edge. The number of polycubes $P \in \mathcal{P}_n$, such that $\gamma(P)$ has two 4-cycles with a common edge, is

$$\mathbf{Z}_{\square\square} = \binom{n-3}{3} \cdot 12 \cdot F_1(6) = 3 \cdot 2^{n-4} n^{n-7} (n-3)(n-4)(n-5).$$

The cells of two 4-cycles with a common edge form a $2 \times 2 \times 2$ cube with two adjacent cells removed, as in . Thus, $\binom{n-3}{3}$ counts the dimensions in which the cube lies. There are 12 ways to remove two adjacent cells, and there are $F_1(6)$ ways to attach six trees to the remaining six cells.

In total, we have $\mathbf{Z} = \mathbf{Z}_{\diamond\!-\!\diamond} + \mathbf{Z}_{\square\square}$.

7 Epilogue

Finally, we have

$$\begin{aligned}
\mathrm{DX}(n, n-3) = {} & \mathbf{X} + \mathbf{Y} + \mathbf{Z} \\
= {} & \mathbf{X}_{22} + \mathbf{X}_3 + \mathbf{Y}_{23} + \mathbf{Y}_{222} + \mathbf{Y}_{\hexagon} + \mathbf{Z}_{\diamond\!-\!\diamond} + \mathbf{Z}_{\square\square} \\
= {} & 2^{n-6} n^{n-7} (n-3)(n-4)(4n^4 - 28n^3 + 97n^2 - 200n + 300) \\
& + 2^{n-3} n^{n-7} (n-3)(2n^4 - 21n^3 + 106n^2 - 282n + 360)/3 \\
& + 2^{n-3} n^{n-7} (n-3)(n-4)(2n^2 - 11n + 30) \\
& + 2^{n-5} n^{n-7} (n-3)(n-4)(n-5)(2n^2 - 7n + 12) \\
& + 2^{n-4} n^{n-7} (n-3)(n-4)(n-5) \\
& + 2^{n-6} n^{n-7} (n-3)(n-4)(n-5)(n-6) \\
& + 3 \cdot 2^{n-4} n^{n-7} (n-3)(n-4)(n-5) \\
= {} & 2^{n-6} n^{n-7} (n-3)(12n^5 - 104n^4 + 360n^3 - 679n^2 + 1122n - 1560)/3,
\end{aligned}$$

which completes the proof of Theorem 1.

8 Conclusion

In this paper we provide a rigorous proof of the formula enumerating n-cell polycubes proper in $n-3$ dimensions, that is, $\mathrm{DX}(n, n-3)$. The proof is based on the same method as in the proof of $\mathrm{DX}(n, n-2)$ [2], but is significantly more involved. This reveals the difficulties expected with the general case, $\mathrm{DX}(n, n-k)$: The number of distinguished structures will grow rapidly, the inclusion relations between them will be much more complicated, and the ways in which they will be connected by forests will be much more varied. This will yield a large number of terms in the inclusion-exclusion analysis. Therefore, we believe that a complete manual analysis will not be feasible for the cases $k > 3$.

Peard and Gaunt [11, p. 6113, eq. (2.15)] predicted that the diagonal formula $\mathrm{DX}(n, n-k)$ has the pattern $2^{n-2k+1}\, n^{n-2k-1}\, g_k(n)$, 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. Luther and Mertens [8] provide the explicit polynomials $h_k(n)$ for $1 \le k \le 7$. Careful inspection of these polynomials reveals that the leading coefficient of $h_k(n)$ has the form $2^{k-1}/(k-1)!$. In conclusion, our refined conjecture is that

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

where $P_k(n)$ is a *monic* polynomial in n. It has also been conjectured [2, 8] that the degree of $P_k(n)$ is $3k-4$.

One can attempt to prove this pattern for the general case of k without actually writing down all the expressions (respective of the cases in our method), but rather by showing that one of them (containing the dominant term, that is, having the highest degree of n) has the form $2^n n^{n-2k-1}(n-k)$ times a polynomial in n of degree $3k-4$, and that all the other expressions have the same form multiplied by polynomials of at most the same degree. Then, the polynomial $P_k(n)$, and thus, $\mathrm{DX}(n, n-k)$, can be interpolated from any $3k-3$ known values of $\mathrm{DX}(n, n-k)$.

Recently, Luther and Mertens [8] provided an argument supporting this hypothesis. They showed that the highest-degree term in $\mathrm{DX}(n, n-k)$ is of the order $2^n n^{n+k-4}$. They also demonstrated (using a "physical argument") that, in fact, only $k+1$ known values suffice for recovering the formulae of $P_k(n)$ and $\mathrm{DX}(n, n-k)$.

We believe that the argument of Luther and Mertens can be refined as follows. The dominant term in $\mathrm{DX}(n, n-k)$ should correspond to the case in which $\gamma(P)$ is a tree (our $\mathbf{X}$). More precisely, it should correspond to the subcase in which all the repeated labels appear in pairs (our $\mathbf{X}_{22}$), that is, when there are $(n-1)-(n-k) = (k-1)$ pairs of repeated labels. Denote the corresponding term by $\mathbf{X}_{22\ldots2}$. In its turn, the dominant term of $\mathbf{X}_{22\ldots2}$ (denote it by $\mathbf{T}_{22\ldots2}$) should correspond to our $\mathbf{T}_{22}$: All the terms added or subtracted in the calculation of $\mathbf{X}_{22\ldots2}$ will have lower order. Note that the use of Lemma 7 will remain unchanged. The statement corresponding to Corollary 8.1 will then be

$$\mathbf{T}_{22\ldots2} = \binom{n-k}{k-1} 2^{n-1} n^{n-3} = \frac{2^{n-1} n^{n-3}}{(k-1)!}(n^{k-1} + \cdots).$$

Finally, we need to divide this expression by 2^{k-1}, the number of trees that represent each polycube in this case (cf. the factor 1/4 in Eq. (1)). Thus, the order of the dominant term is indeed $2^n n^{n+k-4}$, and the constant in this term is $\frac{1}{2^k(k-1)!}$.

Our next attempt will, thus, be to implement a computer program which will automatically generate, for a given value of k, all the "distinguished structures" and all the ways they can be attached to form trees, and will compute the inclusion-exclusion formula for $\mathrm{DX}(n, n-k)$.

References

[1] G. Barequet, M. Moffie, A. Ribó, and G. Rote, Counting polyominoes on twisted cylinders, *Integers* **6** (2006), #A22, http://www.emis.de/journals/INTEGERS/papers/g22/g22.pdf .

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

[3] S. R. Broadbent and J. M. Hammersley, Percolation processes: I. Crystals and mazes, *Math. Proc. Cambridge Philos. Soc.* **53** (1957), 629–641.

[4] D. S. Gaunt and P. J. Peard, $1/d$-expansions for the free energy of weakly embedded site animal models of branched polymers, *J. Phys. A* **33** (2000), 7515–7539.

[5] D. A. Klarner, Cell growth problems, *Canad. J. Math.* **19** (1967), 851–863.

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

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

[8] S. Luther and S. Mertens, Counting lattice animals in high dimensions, *J. Stat. Mech. Theory Exp.* **9** (2011), 546–565.

[9] N. Madras, A pattern theorem for lattice clusters, *Ann. Comb.* **3** (1999), 357–384.

[10] *The On-Line Encyclopedia of Integer Sequences*, http://oeis.org .

[11] 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. Phys. A* **28** (1995), 6109–6124.

[12] R. Stanley, *Enumerative Combinatorics*, Vol. 2, Cambridge University Press, 1999.

2010 *Mathematics Subject Classification*: Primary 05B50; Secondary 05A15, 05A16, 05C30.
Keywords: Lattice animals, polyominoes, directed trees.

(Concerned with sequences A127670, A171860, and A191092.)

Received June 6 2012; revised version received October 1 2012. Published in *Journal of Integer Sequences*, October 2 2012.

Return to Journal of Integer Sequences home page.