Automatic Proofs for Formulae Enumerating Proper Polycubes

Gill Barequet [1]

Dept. of Computer Science
The Technion—Israel Inst. of Technology,
Haifa 32000, Israel

Mira Shalah [2]

Dept. of Computer Science
The Technion—Israel Inst. of Technology,
Haifa 32000, Israel

Abstract

We develop a general framework for computing formulae enumerating polycubes of size n which are proper in $n{-}k$ dimensions (spanning all $n{-}k$ dimensions), for a fixed value of k. Besides the fundamental importance of knowing the number of these simple combinatorial objects, such formulae are central in the literature of statistical physics in the study of percolation processes and collapse of branched polymers. We re-affirm the already-proven formulae for $k \leq 3$, and prove rigorously, for the first time, that the number of polycubes of size n that are proper in $n{-}4$ dimensions is $2^{n-7}n^{n-9}(n-4)(8n^8 - 128n^7 + 828n^6 - 2930n^5 + 7404n^4 - 17523n^3 + 41527n^2 - 114302n + 204960)/6$.

---

[1] Email: `barequet@cs.technion.ac.il`
[2] Email: `mshalah@cs.technion.ac.il`

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 distinct of they differ in their shapes or orientations. A polycube is called *proper* in d dimensions if the convex-hull of the centers of its cubes is d-dimensional. Following Lunnon [7], we let $\mathrm{DX}(n, d)$ denote the number of fixed polycubes of size n that are proper in d dimensions.

Enumeration of polycubes and computing their asymptotic growth rate are important problems in combinatorics and discrete geometry, originating in statistical physics [4], where they play a fundamental role in the analysis of percolation processes and collapse of branched polymers. 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. The main interest in 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 [7]. In a matrix listing the values of DX, the top-right triangular half and the main diagonal contain only 0s. 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. In 1967 Klarner [5] showed in that the limit $\lambda_2 = \lim_{n\to\infty} \sqrt[n]{A_2(n)}$ exists. Only in 1999 Madras [9] proved that the asymptotic growth rate $\lambda = \lim_{n\to\infty} A(n+1)/A(n)$ exists. (Similarly, the growth-rate limit λ_d of polycubes is guaranteed to exist in any fixed dimension $d>2$.) λ_2 (also known as *Klarner's constant*) is the *growth rate* limit of polyominoes (2-dimensional polycubes). Its exact value has remained elusive till these days. The currently best known lower and upper bounds on λ_2 are roughly 4.0025 [2] and 4.6496 [6], respectively. Significant progress in estimating λ_d has been obtained in the literature of statistical physics, although the computations usually relied on unproven assumptions and on formulae for $\mathrm{DX}(n, n-k)$ interpolated empirically from known values of $A_d(n)$. Peard and Gaunt [11] predicted that for $k > 1$, the diagonal formula $\mathrm{DX}(n, n-k)$ has the pattern $2^{n-2k+1} n^{n-2k-1} (n-k) h_k(n)$, where $h_k(n)$ is a polynomial in n, and conjectured explicit formulae for $h_k(n)$ for $k \le 6$. Luther and Mertens [8] conjectured a formula for $k = 7$.

It is easy to show, using Cayley trees, that $\mathrm{DX}(n, n-1) = 2^{n-1} n^{n-3}$ (seq. A127670 in [10]). Barequet et al. [3] proved rigorously, for the first time, that $\mathrm{DX}(n, n-2) = 2^{n-3} n^{n-5} (n-2)(2n^2 - 6n + 9)$ (seq. 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. [1] proved that $\mathrm{DX}(n, n-3) = 2^{n-6} n^{n-7} (n-$

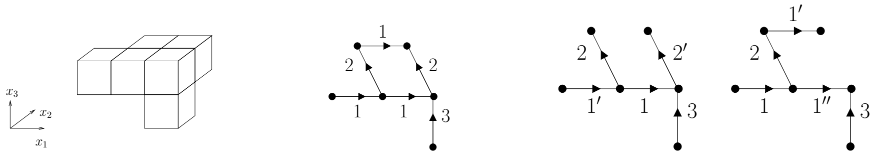
3)$(12n^5-104n^4+360n^3-679n^2+1122n-1560)/3$, again, by counting spanning trees of polycubes, yet the reasoning and calculations were significantly more involved. The inclusion-exclusion principle was applied in the proof in order to count correctly polycubes whose cell-adjacency graphs contained certain 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 are connected by forests are much more varied. As anticipated [1], it is totally impractical to manually achieve a similar proof for $k>3$.

In this paper we create a theoretical set-up for proving the formulae for $\mathrm{DX}(n, n-k)$, for a fixed value of k. Our method fully automates the manual method presented in [3,1]. For this nontrivial generalization, we need 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 and analyze 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 formula (which has never been proven before) for $\mathrm{DX}(n, n-4)$, stated in the following theorem.

Theorem 1.1 $\mathrm{DX}(n, n-4) = 2^{n-7}n^{n-9}(n-4)(8n^8-128n^7+828n^6-2930n^5+7404n^4-17523n^3+41527n^2-114302n+204960)/6$.

2 Overview of the Method

Denote by $\mathcal{P}_n$ the set of proper polycubes of size n in $n-k$ dimensions. Let $P\in\mathcal{P}_n$, and let $\gamma(P)$ denote the adjacency graph of P constructed as follows: Its vertices correspond to the cells of P; two vertices are connected by an edge if their corresponding cells are adjacent; an edge has label i ($1 \le i \le n-k$) if the corresponding cells have different i-coordinate. The direction of the edge is from the lower to the higher cell. See Figure 2 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 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 are labeled with the



(a) Polycube (b) Adjacency graph (c) Spanning trees

Fig. 1. A polycube P, $\gamma(P)$, and two spanning trees of $\gamma(P)$.

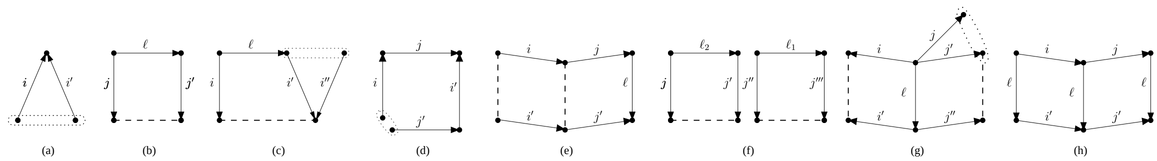


Fig. 2. (a–g) A few structures in $\mathcal{DS}_4$ (h) A cycle structure. A dotted line is drawn between every pair of neighboring cells and around every pair of coinciding cells.

labels $1, 2, ..., n-k$, and the remaining $k-1$ edges repeat labels from the same set. There is a bijection between the possibilities of repeated edge labels and the partitions of the integer $k-1$. Specifically, each partition $p=\{a_1, \ldots, a_h\}$ ($\sum_{i=1}^{h} a_i=k-1$), corresponds to h repeated labels in the spanning tree, such that the ith repeated label appears a_i+1 times. In such case, we say that the tree is "labeled according to p." When we consider a spanning tree of $\gamma(P)$, we distinguish a repeated label i that appears r times by $i, i', ..., i'^{(r-1)}$. However, when considering $\gamma(P)$, repeated labels are assumed not to be distinguished.

To compute $|\mathcal{P}_n|$, we consider all possible directed edge-labeled trees of size n with the possible repetitions of edge-labels, and count only those that represent valid polycubes, and derive the actual number of polycubes.

3 Distinguished Structures

For each directed edge-labeled tree, one can attempt to build the corresponding polycube. In this process there are two types of problems: (a) Cells may coincide (a tree with overlapping cells is invalid, see Figures 2(a,d)); and (b) Two cells which are not connected by a tree edge may be adjacent (such a tree corresponds to a polycube P with cycles in $\gamma(P)$, and therefore, its spanning tree is not unique, see Figures 2(b,e)). Thus, we consider the structures that cause the problems above. A *distinguished* structure is the union of all paths (edges and incident vertices) that connect two coinciding or adjacent cells. This characterization allows the design an algorithm for producing $\mathcal{DS}_k$: The set of all distinguished structures in $n-k$ dimensions. We begin by generating all "free trees" (non-isomorphic trees) of size at most the value specified in Lemma 3.1. Then, we process each free tree T by labeling its edges according to every partition $p \in \cup_{i=1}^{k-1}\Pi(i)$ so as to obtain a directed edge-labeled tree T', and then checking whether T' contains coinciding or neighboring cells (by a DFS traversal). T' is added to $\mathcal{DS}_k$ if it is *not* isomorphic to any structure of size t already in $\mathcal{DS}_k$, and at least one of the following conditions holds:

(i) T' contains two coinciding or neighboring cells which are connected by a path of $t-1$ edges (see, e.g., Figures 2(a,b,d,e));

(ii) T' is isomorphic to the union of $d_1, ..., d_m \in \mathcal{DS}_k$, such that the isomorphic copies of $d_1, ..., d_m$ in T' cover all its edges (see, e.g., Figures 2(c,g)).

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

Lemma 3.1 *A connected (resp., disconnected) distinguished structure in* $\mathcal{DS}_k$ *has at most* $3k-2$ *(resp.,* $4k$*) vertices.*

Lemma 3.2 *[1, Lemma 7] [3, 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* $n \geq 2$*.*

Let T_p denote the number of directed trees with n vertices labeled according to $p \in \Pi(k-1)$. Then, $T_p = \pi(p)\binom{n-k}{|p|}2^{n-1}n^{n-3}$.

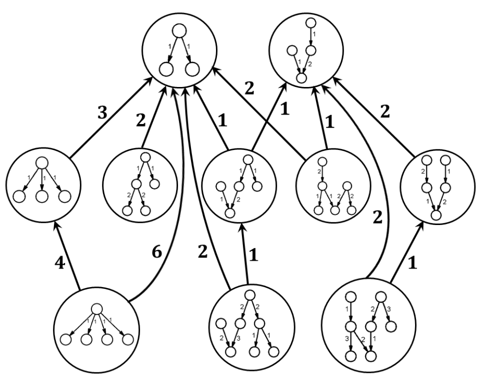
Lemma 3.3 *Let* $\sigma \in \mathcal{DS}_k$ *be composed of* $k^* \geq 1$ *trees* $s_1,\ldots,s_{k^*}$ *with a total of* n^* *vertices and distinct edge labels. The number of occurrences of* σ *in trees of size* n *with distinct edge labels is* $F_n(\sigma) = (\prod_{i=1}^{k^*}|s_i|)\frac{(n-n^*+k^*-1)!}{(n-n^*)!}n^{n-n^*+k^*-2}$.

Proof. By double counting, enumerating in two ways the different sequences of directed edges that can be added to a graph with $n-n^*$ vertices and the structure σ, so as to form a rooted tree. One way is to add the edges one by one: There are $\mathcal{N} = \prod_{i=1}^{k^*}|s_i|$ options to choose a root for each component. We begin with a forest with $n-n^*+k^*$ rooted trees. After adding a collection of edges forming a rooted forest with i trees, there are $n(i-1)$ choices for the next edge. Therefore, the total number of choices is $\mathcal{N}\prod_{i=2}^{n-n^*+k^*} n(i-1)$. Another way to count these edge sequences is to start with an unrooted edge-labeled tree which contains σ, choose one of its n vertices as a root, and choose one of the $(n-n^*)!$ possible sequences, say, η, label the $n-n^*$ vertices which do not belong to σ according to η, and "shift" each vertex label to the incident edge towards the root. The number of sequences formed this way is $nF_n(\sigma)(n-n^*)!$. It follows from that $F_n(\sigma) = \mathcal{N}\frac{(n-n^*+k^*-1)!}{(n-n^*)!}n^{n-n^*+k^*-2}$, as claimed. □

Let $\mathcal{F}_n(\sigma)$ denote the number of occurrences of σ in *directed* edge-labeled trees of size n. Obviously, $\mathcal{F}_n(\sigma) = 2^{n-n^*+k^*-1}F_n(\sigma)$. Let also $\sigma' \in \mathcal{DS}_k$ be 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 $p \in \Pi(k-1)$. Computing $\mathcal{O}_p(\sigma')$ involves choosing the $|p|$ repeated labels in the tree, choosing the $|p'|$ repeated labels of σ' out of those, choosing the unique labels (if there are any) of σ', and computing the number of automporphisms of σ'. Finally, multiplying by $\mathcal{F}_n(\sigma')$ completes the computation.

When counting the occurrences of a structure $\sigma \in \mathcal{DS}_k$, other distinguished structures which contain multiple occurrences of σ are counted multiple times. In order to count correctly, we build an inclusion-exclusion graph IE=$(\mathcal{V}, \mathcal{E})$, which 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, $I(\sigma_2)$ denote the set $\{\sigma_1 \in \mathcal{V} : (\sigma_1, \sigma_2) \in \mathcal{E}\}$, and $T_p(\sigma)$ denote the number of trees of size n labeled according to $p \in \Pi(k-1)$ that contain σ but no other structure $\sigma' \in I(\sigma)$ as a subtree. It is easy to prove by induction that $T_p(\sigma_2) = \mathcal{O}_p(\sigma_2) - \sum_{\sigma_1 \in I(\sigma_2)} \ell((\sigma_1, \sigma_2))T_p(\sigma_1)$. Hence, the roots of the IE graph are all the structures that are not contained in any other structure. In a sense, these are the "biggest" structures. Figure 3.1 shows a subgraph of the IE graph for k=4. A simple bottom-up procedure traverses the IE graph, and computes for every vertex $u \in \mathcal{V}$, $T_p(u), \forall p \in \Pi(k-1)$.

4 Counting Polycubes



For every $P \in \mathcal{P}_n$, there are several possibilities for the structure of $\gamma(P)$: It can be a tree (if P is a tree), or it can have cycles. We enumerate the polycubes which correspond to each possible structure, and sum up the results. The enumeration will be according to the different possibilities of repeated labels in the spanning tree of the polycube (the different partitions of $k-1$).

4.1 Trees

Every tree polycube gives rise to a unique spanning tree. For every possibility of repeated labels $p \in \Pi(k-1)$, let $\mathrm{DT}_p(n)$ denote the number of spanning trees of tree polycubes that are labeled according to p. The total number of directed trees with n vertices labeled according to p is T_p. Every such tree corresponds to a tree polycube in $\mathcal{P}_n$ unless it contains a structure $\sigma \in \mathcal{DS}_k$ as a subtree. (A spanning tree of a tree polycube can neither contain coinciding cells because these are illegal, nor can it contain neighboring cells). Thus, we exclude all the trees that contain every $\sigma \in \mathcal{DS}_k$ as a subtree. Hence, we have $\mathrm{DT}(n, n-k) = \sum_{p \in \Pi(k-1)} \mathrm{DT}_p(n) = \sum_{p \in \Pi(k-1)} \frac{T_p - \sum_{\sigma \in \mathcal{DS}_k} T_p(\sigma)}{\prod_{j=1}^{|p|} p[j]!}$. (The division by $\prod_{j=1}^{|p_i|}(p_i[j]!)$ is because each tree polycube is counted that many times.)

4.2 Nontrees

Let $\mathcal{C}(k)$ denote the set of all cycle structures of polycubes proper in $n-k$ dimensions. This set can be found using $\mathcal{DS}_k$: A distinguished structure is a spanning tree of a cycle if it contains only neighboring cells and no coinciding cells. For example, the structure shown in Figure 2(e) is a spanning tree of the cycle shown in Figure 2(h). For any $\mathcal{C}_i \in \mathcal{C}(k)$, let $P_{\mathcal{C}_i}$ denote the number of polycubes $P \in \mathcal{P}_n$ that contain $\mathcal{C}_i$ in $\gamma(P)$. Suppose that a distinguished structure $\sigma \in \mathcal{DS}_k$ has c occurrences in $\mathcal{C}_i$. Then, $P_{\mathcal{C}_i} = \sum_{p \in \Pi(k-1)} \frac{T_p(\sigma)}{c \prod_{j=1}^{|p|} p[j]!}$.

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

5 Results

The entire method was automated in a C++ program, using `Mathematica` for simplifying the final formula. Our results agree completely with the formulae conjectured in the literature of statistical physics. For k=3, the program found 147 distinguished structures and 13 cycle structures. For k=4, the program found 8,397 distinguished structures, and 179 cycle structures. The parallel computation took about 15 minutes on a supercomputer with 12 processors and 65 GB of RAM. Here are the main results in the computation:
$\mathrm{DT}(n, n-4) = \mathrm{DT}_{(2,2,2)}(n) + \mathrm{DT}_{(2,3)}(n) + \mathrm{DT}_{(4)}(n) = 2^{n-7}(n-4)n^{n-9}(8n^8 - 140n^7 + 1010n^6 - 3913n^5 + 9201n^4 - 15662n^3 + 34500n^2 - 120552n + 221760)/6$;
$\sum_{i=1}^{179} P_{\mathcal{C}_i} = 2^{n-7}(n-4)(n-5)n^{n-9}(12n^6 - 122n^5 + 373n^4 + 68n^3 - 1521n^2 - 578n + 3360)/6$; and finally,
$\mathrm{DX}(n, n-4) = \mathrm{DT}(n, n-4) + \sum_{i=1}^{179} P_{\mathcal{C}_i} = 2^{n-7}n^{n-9}(n-4)(8n^8 - 128n^7 +$

$828n^6 - 2930n^5 + 7404n^4 - 17523n^3 + 41527n^2 - 114302n + 204960)/6$. The program produced data files which document the entire computation and serve as a proof of the formula. This completes the proof of Theorem 1.1.

References

[1] 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.

[2] G. Barequet, G. Rote, and M. Shalah, $\lambda > 4$, *30th European Workshop on Computational Geometry*, Ein-Gedi, Israel, March 2014.

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

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

[5] D.A. Klarner, Cell growth problems, *Canadian J. of Mathematics*, 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, *Canadian J. of Mathematics*, 25 (1973), 585–602.

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

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

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

[10] *The On-Line Encyclopedia of Integer Sequences*, available at `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. Physics, A: Mathematical and General*, 28 (1995), 6109–6124.