

On the combinatorics of crystal structures. II. Number of Wyckoff sequences of a given subdivision complexity

Wolfgang Hornfeck* and Kamil Červený

Received 1 July 2022

Accepted 13 March 2023

Edited by M. I. Aroyo, Universidad del País Vasco, Spain

Keywords: Wyckoff sequences; combinatorics; Shannon entropy; structural complexity.

FZU – Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 00 Prague 8, Czechia.

*Correspondence e-mail: hornfeck@fzu.cz

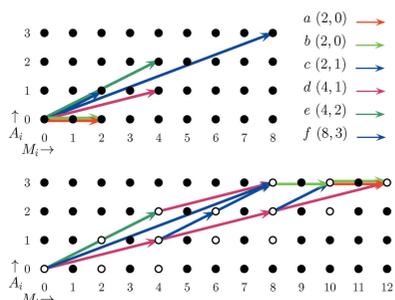
Wyckoff sequences are a way of encoding combinatorial information about crystal structures of a given symmetry. In particular, they offer an easy access to the calculation of a crystal structure's combinatorial, coordinational and configurational complexity, taking into account the individual multiplicities (combinatorial degrees of freedom) and arities (coordinational degrees of freedom) associated with each Wyckoff position. However, distinct Wyckoff sequences can yield the same total numbers of combinatorial and coordinational degrees of freedom. In this case, they share the same value for their Shannon entropy based subdivision complexity. The enumeration of Wyckoff sequences with this property is a combinatorial problem solved in this work, first in the general case of fixed subdivision complexity but non-specified Wyckoff sequence length, and second for the restricted case of Wyckoff sequences of both fixed subdivision complexity and fixed Wyckoff sequence length. The combinatorial results are accompanied by calculations of the combinatorial, coordinational, configurational and subdivision complexities, performed on Wyckoff sequences representing actual crystal structures.

1. Introduction

Any standardized crystal structure can be conveniently related to a descriptor uniquely encoding its combinatorial properties, namely its *Wyckoff sequence*: a string composed of the space group type number (sometimes the Hermann–Mauguin symbol is used instead) and followed by all the Wyckoff letters for each partially or fully occupied Wyckoff position in the crystal structure; the letters are put in reverse alphabetic order and augmented by their superscripted frequency of occurrence, in case a certain non-fixed Wyckoff position is occupied multiple times.

Standardization is necessary because many crystal structures do have equivalent descriptions in terms of their unit cell and atomic coordinates, depending either on matters of possible unit-cell choices or the symmetry properties of space groups [keywords: symmetry of symmetry, Cheshire groups, Euclidean normalizers; see Müller (2013), ch. 8]. A comprehensive scheme for crystal structure standardization has been theoretically developed by Parthé and coworkers (TYPIX) and implemented into the software *STRUCTURE TIDY* (Parthé & Gelato, 1984, 1985; Gelato & Parthé, 1987; Parthé *et al.*, 1993a,b). The uniqueness of the Wyckoff sequence thus depends on and follows from the uniqueness of the standardization.

The Wyckoff positions of a space group encompass all possible distinct sets of symmetry-equivalent sites within a unit cell. Or, to put it in rigorous mathematical terms: a Wyckoff position of a space group \mathcal{G} consists of all points X for



OPEN ACCESS

Published under a CC BY 4.0 licence

which the site-symmetry groups $\mathcal{S}(X)$ are conjugate subgroups of \mathcal{G} (see Hahn, 2005, ch. 8.3.2, p. 733; Aroyo, 2016, ch. 1.4.4.2, p. 62).

It is important to distinguish the notions of space group and space-group type. A space group encompasses all of the symmetry of an actual crystal structure, including its translation symmetry, specified by its lattice parameters. A space-group type is an abstract notion comprising all those, otherwise distinct, space groups which share the same representation by matrix-column pairs of symmetry operations (Nespolo *et al.*, 2018). This is of special importance in the study of group–subgroup relations, in which a group and one of its proper subgroups can be of the same space-group type, but not of the same space group. Another consequence is that the number of space groups is infinite, whereas the number of space-group types is finite: 230 distinct ones in three dimensions, including 11 enantiomorphous types. In particular, the Hermann–Mauguin symbol represents the space-group type, while concepts related to the Wyckoff positions, such as the Wyckoff sequences studied in this work, are related to the space group, in the sense that they are inseparably connected with the coordinate description of crystal structures. However, the number of distinct Wyckoff positions is considered to be finite, with a total of 1731 positions for all space groups, thus allowing for the combinatorial analysis to follow.

The Wyckoff positions are labelled by up to 27 possible Wyckoff letters (a to z and α), being devoid of any further meaning, but depending on the choice of space group. Since the Wyckoff sequence composed of these letters does not include any specific information about either the unit-cell parameters or the atomic coordinates, it is a more abstract notion as well, and useful as such for purposes of crystal structure classification and systematics (see, *e.g.*, Allmann & Hinek, 2007). Note that this abstraction means that actual crystal structures can share the same Wyckoff sequence, while being distinct with respect to the specific values of the free parameter(s) of their non-fixed Wyckoff position(s). Such crystal structures are called isopointal (Lima-de-Faria *et al.*, 1990).

Yet, Wyckoff sequences can be studied without referring to specific geometric crystal structures. Thus, any Wyckoff sequence really describes an infinite family of crystal structures, sharing a common parametrization in terms of their geometric degrees of freedom, according to the observed partial or full occupancy of their general position(s) and, if present, special position(s) of their corresponding space-group type. Indeed, the Wyckoff sequence has been used in the aforementioned sense, as a coordinate-free representation of crystal structures in modern machine learning approaches to materials discovery (Goodall *et al.*, 2022).

One particular advantage of this abstract, combinatorial point of view is due to the fact that the number of Wyckoff sequences of given length k is finite (Hornfeck, 2022a), making an exhaustive study possible for small values of k . In fact, most of the actual crystal structures found so far in nature have Wyckoff sequences of length below $k = 50$. For the space-group type $Pmmm$ (No. 47) with $\nu = 19$ non-fixed sites and $\varphi =$

8 fixed ones, constituting the case with the highest number of possible sequences, this would correspond to about 3.5×10^{18} distinct sequences to consider in total, up to the length of $k = 50$ inclusive (see Appendix A for the computation and the exact result).

Another advantage, and one focus of this work, is due to the fact that the Wyckoff sequence translates into the information about a crystal structure's combinatorial (M) and coordinational (A) collective degrees of freedom, associated with a weighted sum of each Wyckoff position's individual multiplicity (M_i) and arity (A_i), respectively, the latter being the number of independent coordinate parameters required to be specified in a standardized description of an actual crystal structure.

Both the combinatorial and coordinational degrees of freedom can be used to assess a crystal structure's combinatorial and coordinational complexity by means of an approach pioneered by Krivovichev (2012, 2014) and extended by Hornfeck (2020, 2022b) based on the utilization of the Shannon entropy as a complexity measure.

Taking a general point of view, the collective degrees of freedom represent a certain system (a macrostate), while the individual degrees of freedom each represent a certain subsystem (a microstate). This is true on different levels of hierarchy. For instance, on the crystal structure level, the collective configurational degrees of freedom, $M + A$, result as the sum of the individual combinatorial (M) and coordinational (A) degrees of freedom. In a similar way, on the Wyckoff position level, the collective combinatorial and coordinational degrees of freedom, M or A , result as the sum of the individual combinatorial and coordinational degrees of freedom, M_i or A_i , respectively.

In combinatorial terms this splitting of a system into subsystems corresponds to an integer partition. However, in their crystallographic application, the number of partitions is not simply given by the number-theoretic partition function, yet restricted by crystallographic symmetry and the coupling of combinatorial and coordinational degrees of freedoms as found within each Wyckoff position. In particular, the possible values of the Wyckoff multiplicities considering all space-group types are restricted to the set $M_i \in \{1, 2, 3, 4, 6, 8, 12, 16, 24, 48\}$ (assuming primitive unit cells, *i.e.* modulo centring translations, which, in the following, will be the preferred choice of description), while the possible values of the Wyckoff arities are restricted to the set $A_i \in \{0, 1, 2, 3\}$. Moreover, for a given choice of space-group type, the values of either set that can occur are restricted further (although not their frequencies of occurrence, except for fixed positions, which can occur only once), in accordance with the existing Wyckoff positions. A final restriction is imposed by each Wyckoff position introducing a coupling of values from both sets. The combinatorial problem under consideration thus is one of counting the number of restricted, coupled partitions.

Thus, due to these particular restrictions, it is a non-trivial task to describe a macrostate defined by the collective degrees of freedom (M, A) by the composition from or the subdivision

into its associated microstates (M_i, A_i). Foremost, it is a natural question to ask, given one macrostate, what is the number of microstates corresponding to it?

The following section will give the answer to this question, with the main results being a generating polynomial approach (Section 2.4, theory; Section 2.7, algorithm) and a dynamic programming approach (Section 2.8, theory and algorithm).

2. Combinatorics of Wyckoff sequences

To state the problem succinctly: how many distinct Wyckoff sequences (of optionally fixed length k) exist, given a pair of combinatorial and coordinational degrees of freedom (M, A) together with a choice of space group? (Note that a space group has to be fixed, since this determines the alphabet of Wyckoff letters which can appear in the Wyckoff sequence.)

2.1. A problem of crystals – exposition

The problem statement can be translated rather straightforwardly into some algebraic form. Note that two integer values exist for each individual Wyckoff position i , its multiplicity M_i and its arity A_i (coordinational degree of freedom). Both have to be considered in a coupled way, which will be done in the notation by the use of column vectors. Then, the total multiplicity and arity, M and A , respectively, corresponding to a certain set of Wyckoff sequences of crystal structures, are given as

$$\begin{bmatrix} M \\ A \end{bmatrix} = \sum_{i=1}^n \mu_i \begin{bmatrix} M_i \\ A_i \end{bmatrix}. \quad (1)$$

Here, the sum index i runs over all the n existing Wyckoff positions for a given space group, with the integer multipliers μ_i denoting the frequency of occurrence of a given site in the sum of individual multiplicities M_i and arities A_i of the combined sites. In one general point of view, as mentioned in the *Introduction*, one can call the pair (M, A) the system variables (describing the macrostate), while the M_i and arities A_i would be the subsystem variables (describing the microstates).

Information about the Wyckoff positions of the 230 three-dimensional space-group types is compiled in Vol. A of the *International Tables for Crystallography* (Aroyo, 2016), which is the authoritative source. Alternatively, it can also be retrieved online from the Bilbao Crystallographic Server (<https://www.cryst.ehu.es/>) using the routine *WYCKPOS* (Aroyo *et al.*, 2006a,b, 2011).

The values of the Wyckoff multiplicities M_i are explicitly stated as the numeral part of the Wyckoff symbol assigned to each Wyckoff position (the non-numeral part is given by the Wyckoff letter). Note, however, that these values are given for the centred unit cells, in which case M should also be specified with respect to the unit-cell content of the centred unit cell, in order to maintain the correct correspondence. Alternatively, the values of the Wyckoff multiplicities can be reduced by the division of an integer factor depending on the centring type (2 for C , A and I centring; 3 for R centring; 4 for F centring), if

the primitive unit cell, and the number of atoms it contains, is taken as a reference. Indeed, choosing the primitive unit cell as a reference is strongly recommended in the context of crystallographic complexity calculations (*cf.* Section 3), in order to make the results of these calculations comparable for crystal structures differing in their centring type, since any existing centring translation just repeats parts of a crystal structure inside a unit cell, thereby contributing no additional information to its description, and, accordingly, its complexity.

The values of the Wyckoff arities A_i , although not explicitly stated in the aforementioned sources, can be deduced from them, namely by means of visual inspection with respect to the number of positional variables x , y and z occurring in the listing of the general coordinates as provided for each Wyckoff position.

In general, this information can be obtained for any crystallographic space group, including higher-dimensional ones, namely by inspecting the number of symmetry-equivalent positions modulo translations, to obtain the multiplicities, as well as by elucidating the dependency of their general coordinates with respect to symmetry, to obtain the arities. In fact, all this information can be obtained in a purely algorithmic manner [see Brown *et al.* (1978) for the case of four dimensions].

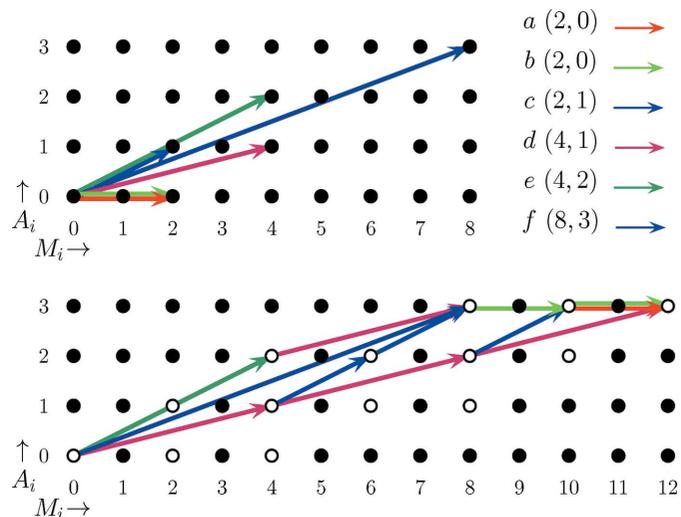


Figure 1 Geometric interpretation for the problem of finding the number of ways of combining individual Wyckoff positions of given multiplicities and arities, (M_i, A_i) , adding up to a given total multiplicity and arity (M, A) . In the illustration, the target vector $(M, A) = (12, 3)$, here written in row form, is denoting a point in the two-dimensional integer lattice \mathbb{Z}^2 (square lattice) in the upper-right corner. On the top, the individual vectors (M_i, A_i) corresponding to each Wyckoff position are shown: $a(2,0)$ red, $b(2,0)$ green, $c(2,1)$ blue, $d(4,1)$ dark red, $e(4,2)$ dark green, $f(8,3)$ dark blue. On the bottom, their combinations adding up to $(M, A) = (12, 3)$ are shown, with vectors composed in reverse lexicographic order. Other possible combinations, in which only the order of the vectors are changed, are not shown. However, all lattice points which can be reached by any possible combinations of vectors are highlighted as open circles instead of filled ones. To see the full graph one has to invert the depicted half of it in the point $(6, 1.5)$. In this interpretation the problem becomes a special case of a lattice path enumeration problem with the set of steps governed by the Wyckoff multiplicities and arities for a given choice of space-group symmetry.

It should be noted that the problem at hand can be given different mathematical interpretations and representations. Equation (1) already suggests a geometrical interpretation as two-dimensional vectors, which happen to live on the two-dimensional integer lattice \mathbb{Z}^2 (square lattice). The problem then appears as a special case of a lattice path problem (Fig. 1), in which the vectors associated with the Wyckoff positions of a space group define the set of steps.

Alternatively, the two-dimensional plane can be identified with the complex plane of Wessel, Argand and Gauß suggesting a change of notation, in which i is the imaginary unit ($i^2 = -1$):

$$M + Ai = \sum_{j=1}^n \mu_j (M_j + A_j i). \quad (2)$$

While these interpretations and representations are mathematically equivalent, and thus do not seem to make any difference *per se*, the knowledge of these alternatives can be of importance when it comes to searching for subfields of mathematics discussing already existing solutions to a given problem, or general methods to find them, and also in the case of implementation into computer code.

2.2. Conditions on the multipliers

For any given site the multipliers μ_i are restricted to discrete intervals

$$\{\mu_{i,\min}, \mu_{i,\min} + 1, \dots, \mu_{i,\max} - 1, \mu_{i,\max}\} \quad (3)$$

of potential integer values, which are denoted in shorthand as $[[\mu_{i,\min}, \mu_{i,\max}]]$ in the following. Naturally, by definition, the frequencies of occurrence are bounded from below, likewise for all sites, by all minimal multipliers $\mu_{i,\min} = 0$, meaning that a site is absent in this case. The variable upper bounds, the maximal multipliers, are determined according to the case distinction

$$\mu_{i,\max} = \begin{cases} r_i & \text{for } A_i = 1, 2, 3 = A_i^+ \\ 1 & \text{for } A_i = 0 \end{cases}, \quad (4)$$

differentiating between non-fixed and fixed sites. Collecting the terms for the non-fixed and fixed sites separately, the summation of equation (1) can be split into separate parts,

$$\begin{bmatrix} M \\ A \end{bmatrix} = \sum_{i=1}^{\nu} r_i \begin{bmatrix} M_i \\ A_i^+ \end{bmatrix} + \sum_{i=1}^{\varphi} \begin{bmatrix} M_i \\ 0 \end{bmatrix}, \quad (5)$$

in which ν and φ now denote the total number of non-fixed and fixed sites, respectively (compare Hornfeck, 2022a).

In any case, the repetition r_i is given as

$$r_i = \min \left\{ \left\lfloor \frac{M}{M_i} \right\rfloor, \left\lfloor \frac{A}{A_i} \right\rfloor \right\}, \quad (6)$$

with $\lfloor \cdot \rfloor$ denoting the floor function, and the smaller ratio restricting the number of possible occurrences of the corresponding site from above. Stated in a different way, each multiplier for any non-fixed site has to fulfil both of the conditions

$$\mu_i M_i \leq M \quad \text{and} \quad \mu_i A_i \leq A \quad (7)$$

simultaneously, with $\mu_{i,\max}$ being defined as the largest integer doing so.

These conditions originate from the fact that the Wyckoff multiplicities and arities are non-negative integers; thus any surpassing of either one of the limits M or A cannot be balanced by the addition of another Wyckoff position, hence cannot be a part of the solution, and thus signifies an end to the Wyckoff sequence construction process. This guarantees the existence of maximal values $\mu_{i,\max}$ and fixes the size of the search space of which the solution space is a subspace.

While the determination of the size of the solution space, the number of Wyckoff sequences existing for a given choice of (M, A) , is our main task, the determination of the size of the search space appears as a first step towards a result, in that it gives a numerical upper bound for the size of the solution space, a combinatorial overview about the Wyckoff sequences to expect, with respect to their length, as well as an estimate regarding the computational tractability of their actual construction based on an exhaustive exploration of the search space.

2.3. Size of the search space

For non-fixed sites the multiplier intervals are given as $[[0, r_i]] = \{0, 1, \dots, r_i\}$, with variable r_i , while for the fixed sites the multiplier intervals are given as $[[0, 1]] = \{0, 1\}$, invariably. The Cartesian product over all interval sets of multipliers of either type determines the search space:

$$\mathcal{S} = [[0, \mu_{1,\max}]] \times [[0, \mu_{2,\max}]] \times \dots \times [[0, \mu_{n,\max}]]. \quad (8)$$

This corresponds to the set of all possible Wyckoff letter sequences, being Wyckoff sequences without the space-group type symbol/number prefix explicitly stated, as constructed from the multiset $[\mu_{1,\max}\alpha_1, \mu_{2,\max}\alpha_2, \dots, \mu_{n,\max}\alpha_n]$, in which the α_i denote a general Wyckoff letter out of n possible letters for a given space group. Note that the choice of space group, while defining the alphabet of Wyckoff letters and limiting the number of terms to consider in the Cartesian product, does not determine the size of the search space by itself. This size, the cardinality

$$|\mathcal{S}| = \prod_{i=1}^n (\mu_{i,\max} + 1) \quad (9)$$

of the search space, in which individual solutions have to be found, if they exist, is determined by the values of the maximal multipliers. Thus, the size of the search space is determined only if both the space group and the total numbers of degrees of freedom M and A are fixed. Then, the search space size gives an absolute upper bound on the potential number of solutions, yet usually, and in anticipation of our following results, the actual number of solutions will be much lower or even zero. This difference in the number of solutions is due to the construction of the search space by means of the Cartesian product, namely because the restrictions imposed by the $\mu_{i,\max}$ values for individual Wyckoff positions do not take into account their cumulative, conditional interactions.

As is often the case in combinatorial problems, the same cardinality $|\mathcal{S}|$ can be obtained by an alternative counting method, namely as the result of the summation of all coefficients in the expansion of the univariate polynomial defined by

$$P(x) = \prod_{i=1}^n \sum_{k=0}^{\mu_{i,\max}} x^k. \quad (10)$$

This is the generating polynomial for a multiset with finite multiplicities [compare its description and, in particular, equation (4) in Hornfeck (2022a)]. Upon its expansion the coefficients c_k for each term x^k of this polynomial count the number of sequences of length k , thereby representing a more differentiated view of the search space's contents. Eventually,

$$|\mathcal{S}| = \sum_{k=0}^{\mu} c_k, \quad (11)$$

in which $\mu = \mu_{1,\max} + \mu_{2,\max} + \dots + \mu_{n,\max}$ is the multiset's cardinality.

The size of the search space gains some importance due to the fact that the combinatorial approach we will describe in the remainder of this work is non-constructive, as is commonly the case for such combinatorial questions, since it gives only the number of potential Wyckoff sequences matching with a given parameter pair (M, A) , but does not reveal the Wyckoff sequences themselves; these can be discovered by an exhaustive check of all admissible multiplier combinations existing within the combined multiplier intervals.

We envision that the search space size can be reduced to some degree by applying effective intermediate checks for multiplier combinations already violating the upper limit as imposed by the choice of the parameters (M, A) , overshooting either one parameter at a time or both simultaneously, possibly in combination with the use of a clever data structure such as pruned trees. However, search space sizes below $|\mathcal{S}| < 10^6$ are easily tractable on a standard desktop personal computer, which should encompass most tasks related to the comparison of the potential combinatorial solutions with Wyckoff sequences representing actual crystal structures.

2.4. A generating polynomial approach to find solutions

Our combinatorial problem stated above is solved in two steps: first, by reducing it, conceptually, to an analogous classical problem of combinatorics, the coin change problem, as can be found in many textbooks on the topic (for instance, Marcus, 1998, p. 89), and second, by adapting the classical problem to the crystallographic one.

The classical coin change problem is stated in Appendix B, and can be seen as an illustration of the use of generating polynomials, defined in some abstract variable. Notably, the variable is an indeterminate symbol only, entailing no specific meaning other than to allow algebraic operations performed on it; hence it merely acts as a placeholder and bookkeeping device, yet is the decisive one, in order to systematically find a solution.

This use of generating polynomials has already been described in the first entry of this series (Hornfeck, 2022a) to which the reader, interested in more detailed information, is referred. An introduction to the wider field of generating functions is given by Graham *et al.* (1994), and a more detailed exposition of the main ideas involved is given by Wilf (2006).

Now, adapting the classical coin change problem to our crystallographic one is carried out in three steps: (i) matching the number of distinct types of coins with the number of distinct Wyckoff positions; (ii) identifying the values of distinct types of coins with the *pair* of values of the Wyckoff position's multiplicity M_i and arity A_i ; and (iii) treating the pair of values (M_i, A_i) in a coupled way, by introducing two abstract variables (x, y) in the generating polynomials, instead of one.

Thus, in some way, the crystallographic problem is a coin change problem with a twist, based on imaginary coins with denominations on both their front and back sides and a pair of target values to reach upon summation. Similar combinatorial problems arise for the case of real cards, coupling values denoted by digits or letters with symbolic ones such as diamonds, hearts, spades and clubs.

In particular, the third adaptation step is crucial for the correct enumeration. As a consequence of it, generating polynomials of the kind

$$P_i(x, y) = \sum_{j=0}^{\mu_{i,\max}} x^j M_i y^{j A_i} \quad (12)$$

are assigned to each Wyckoff position, with their product over all n Wyckoff positions,

$$\mathcal{P}(x, y) = \prod_{i=1}^n P_i(x, y), \quad (13)$$

yielding the solution, namely by the value of the coefficient of $x^M y^A$ in the expanded form of the polynomial $\mathcal{P}(x, y)$.

As an aside, one can note that the product over all n sites can be split,

$$\mathcal{P}(x, y) = \left[\prod_{i=1}^v \sum_{j=0}^{r_i} x^j M_i y^{j A_i} \right] \left[\prod_{i=1}^{\varphi} (1 + x^{M_i}) \right], \quad (14)$$

according to the contributions of v non-fixed and φ fixed sites [compare equation (5)]. This splitting reduces the problem for the fixed sites to a univariate one.

It should be noted that this approach can be further generalized, in principle, by taking into account chemical degrees of freedom [as introduced by Hornfeck (2020)] in terms of atomic numbers of atoms occupying a given Wyckoff position as well. Then, one would have to introduce a third variable z into the respective polynomials with all other procedures considered to be analogously performed. However, there is a difference, in that the atomic numbers are not restricted in any way, that is to say there exists no natural coupling between them and the Wyckoff multiplicities and arities – they are a pure matter of choice. In contrast, the Wyckoff multiplicities and arities are coupled in their values for each Wyckoff position of a space group. Thus, regarding

this relative arbitrariness of choice and the relative unimportance of this general case, we refrain from expanding in this direction for the moment. However, our opinion on this topic might change in the future, if there should be an interesting application for fixing the total chemical degrees of freedom, that is the total electron count within a reduced unit cell, to a set value, together with the other degrees of freedom, and asking for the number of crystal structures fulfilling this condition. In this case, any generalization required can be obtained in a straightforward manner by following the same extension procedure as described in the following.

2.5. A generalization including the Wyckoff sequence length

Another and considerably more useful generalization is made by taking into account the length k of the Wyckoff sequence as a further restriction, which can be seen as an extension and a refinement to a previous result in the combinatorics of Wyckoff sequences (Hornfeck, 2022a). This can be achieved by adjusting equation (1) according to

$$\begin{bmatrix} M \\ A \\ k \end{bmatrix} = \sum_{i=1}^n \mu_i \begin{bmatrix} M_i \\ A_i \\ k_i \end{bmatrix} \quad (15)$$

in addition to

$$r_i = \min \left\{ \left\lfloor \frac{M}{M_i} \right\rfloor, \left\lfloor \frac{A}{A_i} \right\rfloor, k \right\}, \quad (16)$$

used for the determination of the $\mu_{i,\max}$ for each individual site i . Trivially, each site itself is of length $k_i = 1$ (thus, $\lfloor k/k_i \rfloor = k$), which gives the total length of the Wyckoff sequence as the sum of the multipliers μ_i : $k = \mu_1 + \mu_2 + \dots + \mu_n$. Again, the final result is given as the value of the coefficient of $x^M y^A z^k$ in the expanded polynomial

$$\mathcal{P}(x, y, z) = \prod_{i=1}^n P_i(x, y, z), \quad (17)$$

in which

$$P_i(x, y, z) = \sum_{j=0}^{\mu_{i,\max}} x^{jM_i} y^{jA_i} z^j \quad (18)$$

defines the individual terms.

As was the case before, the product over all n sites

$$\mathcal{P}(x, y, z) = \left[\prod_{i=1}^v \sum_{j=0}^{r_i} x^{jM_i} y^{jA_i} z^j \right] \left[\prod_{i=1}^{\varphi} (1 + x^{M_i} z) \right] \quad (19)$$

can be split according to the contributions of v non-fixed and φ fixed sites.

Finally, it should be noted that there exist general methods for obtaining explicit formulas for the coefficients of generating functions, for instance for the powers of bivariate generating functions (Kruchinin *et al.*, 2021), as this is an active field of mathematical research.

2.6. A problem of crystals – exemplification

In the following, an illustrative example is discussed in full calculational detail.

The tetragonal space-group type $P\bar{4}2_1m$ (No. 113) encompasses a total of six distinct Wyckoff positions,

$$\mathcal{W}_i(M_i, A_i) = \begin{bmatrix} M_i \\ A_i \end{bmatrix}, \quad (20)$$

here written in a more convenient in-line notation, indexed according to their Wyckoff letters, and with their corresponding multiplicities and arities stated:

$$\mathcal{W}^{113} = \left\{ \begin{array}{lll} \mathcal{W}_a(2, 0), & \mathcal{W}_b(2, 0), & \mathcal{W}_c(2, 1), \\ \mathcal{W}_d(4, 1), & \mathcal{W}_e(4, 2), & \mathcal{W}_f(8, 3). \end{array} \right. \quad (21)$$

The chosen space group is the one with the smallest number of Wyckoff positions for which all possible arity values $A_i = 0, 1, 2, 3$ are present. The six Wyckoff positions comprise a total of three and four distinct values for the Wyckoff multiplicity and arity, respectively, thereby allowing us to illustrate the combinatorial calculation in some detail.

Now, with some arbitrary chosen $M = 12$ and $A = 3$ given for this particular example, this results in the following maximal multipliers:

$$\begin{array}{l} \mu_{a,\max} = 1, \mu_{b,\max} = 1, \mu_{c,\max} = 3, \\ \mu_{d,\max} = 3, \mu_{e,\max} = 1, \mu_{f,\max} = 1, \end{array} \quad (22)$$

for each Wyckoff position, and consequently in the following generating polynomials:

$$\begin{array}{l} P_a(x) = 1 + x^2, \\ P_b(x) = 1 + x^2, \\ P_c(x, y) = 1 + x^2y + x^4y^2 + x^6y^3, \\ P_d(x, y) = 1 + x^4y + x^8y^2 + x^{12}y^3, \\ P_e(x, y) = 1 + x^4y^2, \\ P_f(x, y) = 1 + x^8y^3. \end{array} \quad (23)$$

Note how the two fixed sites with Wyckoff letters a and b give rise to exactly the same polynomial factor in one variable x only (namely the one used for the Wyckoff multiplicities), and how the polynomials corresponding to the non-fixed sites are restricted in terms of the monomials of highest degree in either x or y by the given values of either M or A or both.

Note also that the number of variables in each polynomial corresponds to the splitting of all n sites into v contributions from the non-fixed sites (bivariate case) and φ contributions from the fixed ones (univariate case) [compare equation (5)].

The expanded polynomial $\mathcal{P}(x, y)$ consisting of a total of 60 terms is given by

$$\begin{aligned}
 &1 + 2x^2 + x^4 + x^2y + 3x^4y \\
 &+ 3x^6y + x^8y + 2x^4y^2 + 5x^6y^2 + 5x^8y^2 \\
 &+ 3x^{10}y^2 + x^{12}y^2 + 2x^6y^3 + 7x^8y^3 + 9x^{10}y^3 \\
 &+ 6x^{12}y^3 + 3x^{14}y^3 + x^{16}y^3 + x^8y^4 + 5x^{10}y^4 \\
 &+ 10x^{12}y^4 + 10x^{14}y^4 + 5x^{16}y^4 + x^{18}y^4 + x^{10}y^5 \\
 &+ 5x^{12}y^5 + 10x^{14}y^5 + 12x^{16}y^5 + 9x^{18}y^5 + 3x^{20}y^5 \\
 &+ 3x^{14}y^6 + 9x^{16}y^6 + 12x^{18}y^6 + 10x^{20}y^6 + 5x^{22}y^6 \\
 &+ x^{24}y^6 + x^{16}y^7 + 5x^{18}y^7 + 10x^{20}y^7 + 10x^{22}y^7 \\
 &+ 5x^{24}y^7 + x^{26}y^7 + x^{18}y^8 + 3x^{20}y^8 + 6x^{22}y^8 \\
 &+ 9x^{24}y^8 + 7x^{26}y^8 + 2x^{28}y^8 + x^{22}y^9 + 3x^{24}y^9 \\
 &+ 5x^{26}y^9 + 5x^{28}y^9 + 2x^{30}y^9 + x^{26}y^{10} + 3x^{28}y^{10} \\
 &+ 3x^{30}y^{10} + x^{32}y^{10} + x^{30}y^{11} + 2x^{32}y^{11} + x^{34}y^{11}. \quad (24)
 \end{aligned}$$

The decisive term $x^M y^A = x^{12} y^3$ has 6 as its coefficient, which thus corresponds to the six existing solutions, looked for in the search space; they can be stated in terms of their Wyckoff letter sequence, with multiple occurrences of the same letter marked by superscripts:

$$fba, edba, d^3, d^2cb, d^2ca, dc^2ba. \quad (25)$$

Note, in particular, the cases d^2cb and d^2ca which differ only in the exact choice of the fixed site, being otherwise degenerate in their column vector summation [compare equation (1)].

In fact, the bivariate polynomial given in equation (24) contains the information about *all* possible solutions (M, A) which can be constructed from the multiset of Wyckoff letters $[a, b, 3c, 3d, e, f]$ with restricted multiset multiplicities, as determined by the maximal multipliers. Since the multiplier intervals are $[[0, 3]]$ for both the Wyckoff positions c and d , as well as $[[0, 1]]$ for all other sites, the search space has a size of $4^2 \times 2^4 = 256$ cases, whose distribution according to the length k of the sequence can be read off from the expansion (11 terms; not shown) of the generating polynomial

$$(1+x)^4(1+x+x^2+x^3)^2 \quad (26)$$

since $\mu_{i,\max} = 1$ for four out of the six Wyckoff sites and $\mu_{i,\max} = 3$ for the remaining two. The observed six solutions with the property $(M, A) = (12, 3)$ constitute only a tiny fraction of the search space, which is generally true, the size of the search space typically being overwhelmingly larger than the number of solutions.

In particular, the full set of solutions contains the singular occurrence of the empty sequence, $\{\}$, corresponding to the trivial monomial $x^0 y^0 = 1$, as well as that of the maximal length sequence, $abcccddef$, corresponding to the monomial $x^{34} y^{11}$ of highest degree, for which $M = 34$ and $A = 11$. Note how one can instantly check that there exists no solution for, say, the case $M = 7$ and $A = 1$, because no monomial of the form $x^7 y$ appears in $\mathcal{P}(x, y)$, or, to put it another way, the coefficient of this monomial in the expansion of $\mathcal{P}(x, y)$ is equal to zero. Note also how one specific solution, namely solution d^3 , corresponds to the occurrence of the monomial $x^{12} y^3$, the $j = 3$ case of the basic monomial $x^j y^j$ representing

the $\mathcal{W}_d(4, 1)$ Wyckoff position, in the generating polynomial $P_d(x, y)$ of equation (23).

Taking into account the length k as an additional parameter, one has to adjust the individual polynomial terms according to

$$\begin{aligned}
 P_a(x, z) &= 1 + x^2 z, \\
 P_b(x, z) &= 1 + x^2 z, \\
 P_c(x, y, z) &= 1 + x^2 yz + x^4 y^2 z^2 + x^6 y^3 z^3, \\
 P_d(x, y, z) &= 1 + x^4 yz + x^8 y^2 z^2 + x^{12} y^3 z^3, \\
 P_e(x, y, z) &= 1 + x^4 y^2 z, \\
 P_f(x, y, z) &= 1 + x^8 y^3 z. \quad (27)
 \end{aligned}$$

Expansion of their product yields a polynomial in 142 terms (not shown), with, for instance, the value of the coefficient for $x^M y^A z^k = x^{12} y^3 z^4$ being equal to 3, corresponding to the Wyckoff letter sequences

$$edba, d^2cb, d^2ca, \quad (28)$$

which form a subset of the six aforementioned solutions.

2.7. Summary of the generation function approach

To summarize the aforementioned results in an algorithmic form, using the generating polynomial approach one has to perform the following steps to obtain a solution (with the first two items in the enumeration defining the input to the algorithm and the last one its output):

- (i) Fix a space group and thereby the set of Wyckoff positions and the potential alphabet of Wyckoff letters from which a Wyckoff sequence can be formed.
- (ii) Fix the values for the total Wyckoff multiplicity M and the total Wyckoff arity A and (optionally) the Wyckoff sequence length k , all of which are expected to be integers.
- (iii) Retrieve all the individual values for the Wyckoff multiplicities and arities $\mathcal{W}_i(M_i, A_i)$.
- (iv) From the individual Wyckoff multiplicities and arities compute the individual maximal multipliers $\mu_{i,\max}$.
- (v) From the individual maximal multipliers construct the individual generating polynomials, $P_i(x, y)$ or $P_i(x, y, z)$.
- (vi) From the individual generating polynomials build their product, $\mathcal{P}(x, y)$ or $\mathcal{P}(x, y, z)$, and expand it.
- (vii) From the expansion read off the coefficient for the monomial of the form $x^M y^A$ or $x^M y^A z^k$; this is the solution.

2.8. A dynamic programming approach to find solutions

Finally, after illustrating the aforementioned combinatorial method on an explicit example, we want to highlight the possibility of an alternative algorithmic way of calculation which turns out to be very efficient, by making optimal use of the recursive nature and overlapping substructure of the problem, in such a way that subsolutions are only ever computed once and retrieved as needed for the calculation of the solution. This algorithmic way is known under the name dynamic programming. A simple example for the classical

univariate and infinite coin change problem is given in Appendix B. In the following, we show the adapted pseudocode for the bivariate and finite coin change problem in its crystallographic application, thus taking into account the case distinction for the multipliers of non-fixed and fixed Wyckoff sites (Fig. 2). The adaptation to the trivariate and finite case including the length k as a parameter follows a straightforward procedure (Fig. 3), as would be the case for any future multivariate generalization, for instance also taking into account chemical degrees of freedom.

In either case, the approach starts with the trivial base case (there is always one possibility for the null tuple), from which it proceeds bottom-up making use of a tabulation implementation of subsolution values cached into either a matrix (bivariate case) or a tensor (trivariate case), which is updated iteratively until the target case is reached and the solution is returned. Equation (29) shows the solution matrix \mathcal{T} for the case defined by the target tuple $(M, A) = (12, 3)$ and the multiset of tuples $\mathcal{W} = [(2, 0), (2, 0), (2, 1), (4, 1), (4, 2), (8, 3)]$ in which the values for M_i and A_i increase along the rows and the columns, respectively. Note that for a more economic display, the matrix is shown in its transposed form, with columns and rows interchanged, such that its upper-left corner represents the matrix element $\mathcal{T}_{0,0}$ and the lower-right corner represents the matrix element $\mathcal{T}_{12,3}$, from which the solution, $\mathcal{T}_{12,3} = 6$, can be read off:

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 5 & 0 & 5 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 7 & 0 & 9 & 0 & 6 \end{bmatrix}. \quad (29)$$

A direct comparison shows that all non-zero entries represent the non-vanishing coefficients as occurring in the polynomial expansion given in equation (24), yet with terms only up to the solution $x^{12}y^3$ monomial inclusive. Note that the algebraic structure of the Wyckoff positions is reflected in the matrix as well, since the reachable positions are determined by the smallest increments and their parity as observed for the Wyckoff multiplicities and arities. For instance, since all the Wyckoff multiplicities are even numbers, all odd columns of the above matrix are given by the null vector.

A Python implementation of the algorithm for the bivariate case is given in Appendix C, which also contains a *Mathematica* implementation of the generating polynomial approach.

Input: 1) Target tuple for the total multiplicity, arity, and length: (M, A, k)
 2) Multiset of Wyckoff position tuples for a chosen space group:
 $\mathcal{W} = [(M_1, A_1), (M_2, A_2), \dots, (M_i, A_i), \dots, (M_N, A_N)]$

Output: Number of Wyckoff sequences matching the target tuple

Input: 1) Target tuple for the total multiplicity and arity: (M, A)
 2) Multiset of Wyckoff position tuples for a chosen space group:
 $\mathcal{W} = [(M_1, A_1), (M_2, A_2), \dots, (M_i, A_i), \dots, (M_N, A_N)]$

Output: Number of Wyckoff sequences matching the target tuple

```

1  $\mathcal{T}$ : null matrix of dimension  $M + 1 \times A + 1$ 
2  $\mathcal{T}[0][0] \leftarrow 1$ 
3 for  $i = 0$  to  $N$  do
4      $M_i \leftarrow \mathcal{W}[i][0]$ 
5      $A_i \leftarrow \mathcal{W}[i][1]$ 
6     if  $A_i = 0$  then
7          $f_i \leftarrow 1$ 
8     else
9          $f_i \leftarrow \min(\lfloor M/M_i \rfloor, \lfloor A/A_i \rfloor)$ 
10    for  $u$  in  $M - M_i$  to 0 do
11        for  $v$  in  $A - A_i$  to 0 do
12            if  $\mathcal{T}[u][v] > 0$  then
13                for  $f$  in 1 to  $f_i$  do
14                     $M_f \leftarrow u + f \times M_i$ 
15                     $A_f \leftarrow v + f \times A_i$ 
16                    if  $M_f \leq M$  and  $A_f \leq A$  then
17                         $\mathcal{T}[M_f][A_f] += \mathcal{T}[u][v]$ 
17 return  $\mathcal{T}[M][A]$ 
    
```

Figure 2
 Dynamic programming algorithm (given in pseudocode) for the determination of the number of Wyckoff sequences of a given space group and subdivision complexity as determined by the total number of degrees of freedom for the Wyckoff multiplicity and arity (bivariate case).

```

1  $\mathcal{T}$ : null tensor of dimension  $M + 1 \times A + 1 \times k + 1$ 
2  $\mathcal{T}[0][0][0] \leftarrow 1$ 
3 for  $i = 0$  to  $N$  do
4      $M_i \leftarrow \mathcal{W}[i][0]$ 
5      $A_i \leftarrow \mathcal{W}[i][1]$ 
6      $k_i \leftarrow 1$ 
7     if  $A_i = 0$  then
8          $f_i \leftarrow 1$ 
9     else
10         $f_i \leftarrow \min(\lfloor M/M_i \rfloor, \lfloor A/A_i \rfloor, \lfloor k/k_i \rfloor)$ 
11    for  $u$  in  $M - M_i$  to 0 do
12        for  $v$  in  $A - A_i$  to 0 do
13            for  $w$  in  $k - k_i$  to 0 do
14                if  $\mathcal{T}[u][v][w] > 0$  then
15                    for  $f$  in 1 to  $f_i$  do
16                         $M_f \leftarrow u + f \times M_i$ 
17                         $A_f \leftarrow v + f \times A_i$ 
18                         $k_f \leftarrow w + f \times k_i$ 
19                        if  $M_f \leq M$  and  $A_f \leq A$  and  $k_f \leq k$  then
20                             $\mathcal{T}[M_f][A_f][k_f] += \mathcal{T}[u][v][w]$ 
20 return  $\mathcal{T}[M][A][k]$ 
    
```

Figure 3
 Dynamic programming algorithm (given in pseudocode) for the determination of the number of Wyckoff sequences of a given space group, subdivision complexity and length as determined by the total number of degrees of freedom for the Wyckoff multiplicity, Wyckoff arity and length (trivariate case). Note that $k_i = 1$ for all Wyckoff sites – this does not have to be specified for each Wyckoff position independently. However, possible simplifications due to this fact have not been included in the pseudocode in order to highlight its similarity with the bivariate case shown in Fig. 2.

3. Complexity of Wyckoff sequences

The aforementioned problem of crystals arose in the context of calculating Shannon entropy based complexity measures for crystal structures, taking into account a crystal structure's fundamental chemical, combinatorial and coordinational degrees of freedom (Hornfeck, 2020). Apart from its chemical degrees of freedom – its decoration (colouring) of Wyckoff sites with atoms – a crystal structure is geometrically defined by its combined combinatorial and coordinational degrees of freedom, the collective multiplicities M and arities A of its occupied Wyckoff positions.

3.1. Shannon entropy based complexity measures

In general, a given multiset $[X_1, X_2, \dots, X_n]$ of n individual degrees of freedom X_i defines a discrete probability distribution

$$\mathcal{X} = [X_1/X, X_2/X, \dots, X_n/X] = [x_1, x_2, \dots, x_n], \quad (30)$$

where $X = X_1 + X_2 + \dots + X_n$ denotes the collective number of degrees of freedom as obtained from the partition of the individual numbers of degrees of freedom. From this a Shannon entropy

$$H(\mathcal{X}) = \sum_{i=1}^n L(x_i) = \sum_{i=1}^n -x_i \log_2 x_i \quad (31)$$

can be obtained. Note that $L(0) = 0$, by definition.

As mentioned before, the general interpretation is that of a system with X degrees of freedom on a higher level of structural hierarchy being subdivided into n subsystems of X_i degrees of freedom, each on a lower level of structural hierarchy. For more details of the general theory, the reader is referred to previous work done by the author (Hornfeck, 2020, 2022b).

In particular, a crystal structure's M collective combinatorial degrees of freedom are associated with the multiset $[M_1, M_2, \dots, M_n]$ of individual Wyckoff multiplicities M_i , thereby yielding a fundamental combinatorial Shannon entropy:

$$H_{\text{comb}} = H(\mathcal{M}) = \sum_{i=1}^n L(M_i/M). \quad (32)$$

Note that in order to ensure the comparability of combinatorial complexity values between different crystal structures, the values of the Wyckoff multiplicities M_i have to refer to a primitive unit cell [although this is really important only for derived complexity values such as the maximal complexity, $H_{\text{comb,max}} = \log_2 M$, the normal complexity, $H_{\text{comb,norm}} = H_{\text{comb}}/H_{\text{comb,max}}$, or the total complexity, $H_{\text{comb,total}} = M \times H_{\text{comb}}$, which have been described by Hornfeck (2020)].

Now, in the same way, a crystal structure's A collective coordinational degrees of freedom are associated with the multiset $[A_1, A_2, \dots, A_n]$ of individual Wyckoff arities A_i , thereby yielding a fundamental coordinational Shannon entropy

$$H_{\text{coor}} = H(\mathcal{A}) = \sum_{i=1}^n L(A_i/A). \quad (33)$$

Now, proceeding in a completely analogous manner, a crystal structure's $F = M + A$ collective configurational degrees of freedom are associated with the combined multiset

$$[M_1, M_2, \dots, M_n, A_1, A_2, \dots, A_n] = [F_1, F_2, \dots, F_{2n}] \quad (34)$$

of individual Wyckoff multiplicities M_i and individual Wyckoff arities A_i , corresponding to the individual configurational degrees of freedom F_i , thereby yielding a composite configurational Shannon entropy

$$H_{\text{conf}} = H(\mathcal{F}) = \sum_{i=1}^{2n} L(F_i/F). \quad (35)$$

While the association with the combined multiset is natural, it is noteworthy to mention that this does not mean that the corresponding entropies are (simply) additive; on the contrary

$$H_{\text{conf}} \neq H_{\text{comb}} + H_{\text{coor}}. \quad (36)$$

Most importantly, however, by means of the strong additivity property of the Shannon entropy, the configurational entropy is equivalent to the general expression

$$H_{\text{conf}} = H_{\text{subdiv}} + w_M H_{\text{comb}} + w_A H_{\text{coor}}, \quad (37)$$

including appropriate weighting factors $w_M = M/(M + A)$ and $w_A = A/(M + A)$ and an additional subdivision (mixing) complexity based on them,

$$H_{\text{subdiv}} = L(w_M) + L(w_A). \quad (38)$$

This equivalence now allows an assessment of the relative complexity related to the splitting of a given collective number

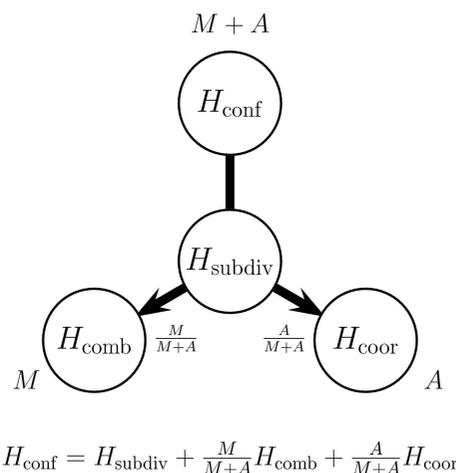


Figure 4 Schematic representation of the relation between the configurational complexity H_{conf} , calculated for $M + A$ degrees of freedom, and the combinatorial and coordinational complexities, H_{comb} and H_{coor} , calculated for M and A degrees of freedom, respectively. Shown also are the weighting factors $M/(M + A)$ and $A/(M + A)$ as well as the unit weight contribution of the invariant subdivision complexity $H_{\text{subdiv}}(M, A)$ which taken all together sum to the configurational complexity H_{conf} . For a given choice of M and A the subdivision complexity $H_{\text{subdiv}}(M, A)$ is a constant, while the combinatorial and coordinational complexities, H_{comb} and H_{coor} , depend on the respective partitions of M and A possible for a given space-group type.

of degrees of freedom, $M + A$, for the composite system, into individual numbers of degrees of freedom, M and A , for the fundamental (sub)systems.

The subdivision complexity H_{subdiv} is a Shannon entropy of the weighting factors occurring in equation (37) and for all the combinatorially enumerated cases it is an invariant characterizing the subdivision step $(M + A) \rightarrow (M, A)$ on the higher crystal structure level of hierarchy. It can be seen as connecting the collective treatment of degrees of freedom, within the combined configurational complexity H_{conf} , with the individual treatment of degrees of freedom, within the separated combinatorial and coordinational complexities, H_{comb} and H_{coor} , respectively. This allows for a complexity partition analysis (Hornfeck, 2022b). This central importance of the subdivision complexity H_{subdiv} and its relationship to the other complexity measures is graphically depicted in Fig. 4.

In the same manner, other subdivision complexities exist on the lower Wyckoff position level of hierarchy, characterizing the subdivision steps $(X_1 + X_2 + \dots + X_n) \rightarrow (X_1, X_2, \dots, X_n)$, with X denoting either M or A .

Furthermore, it should be noted that the strong additivity property also holds for the case of maximal Shannon entropies

$$H_{\text{conf,max}} = H_{\text{subdiv}} + w_M H_{\text{comb,max}} + w_A H_{\text{coor,max}}, \quad (39)$$

where H_{subdiv} takes on the same value as in equation (37). The maximal entropies are reached in the case of maximal subdivision and thus perfect equidistribution of degrees of freedom (corresponding to maximally expanded partitions $1 + 1 + \dots + 1$ of variable length equal to the number of degrees of freedom). This makes their calculation particularly simple, resulting eventually in $H_{\text{comb,max}} = \log_2 M$, $H_{\text{coor,max}} = \log_2 A$ and $H_{\text{conf,max}} = \log_2(M + A) = \log_2 F$.

Each maximal entropy can also be used in turn to define its corresponding non-maximal entropy, for instance

$$H_{\text{conf}} = \log_2 F - \sum_{i=1}^{2n} (F_i/F) \log_2 F_i \quad (40)$$

by means of taking the difference between the maximal entropies for the collective (here, F) and individual (here, F_i) degrees of freedom, the latter ones attributed with their appropriate weighting factors F_i/F .

All the aforementioned interrelations are depicted in Fig. 5.

4. Combining the combinatorics and complexity of Wyckoff sequences

Some elementary statistical results shall be stated about the magnitude of the collective degrees of freedom to expect, in extreme and on average, and with respect to actual crystal structure data as retrieved from the 20 040 unique Wyckoff sequences compiled in the Pearson's Crystal Data Crystal Structure Database for Inorganic Compounds (Villars & Cenzual, 2020).

The minimal observed combinatorial and coordinational degrees of freedom are, as determined for the individual, independent distributions, $M_{\text{min}} = 1$, $A_{\text{min}} = 0$, the maximal are $M_{\text{max}} = 5926$, $A_{\text{max}} = 1476$, the mean values are $M_{\text{mean}} =$

94.3, $A_{\text{mean}} = 53.1$, the median values are $M_{\text{median}} = 48$, $A_{\text{median}} = 20$, and the mode values are $M_{\text{mode}} = 24$, $A_{\text{mode}} = 6$, respectively.

Our combinatorial result now tells us how many of these partitions, being solutions to a combinatorial problem restricted by crystallographic symmetry, can be realized for a given space group, given the fixed number of degrees of freedom (M, A) and, potentially, the length k of the Wyckoff sequence. In the following we will illustrate this by performing explicit calculations for three examples.

It should be noted that the examples given here were chosen such that the discussion of the results could be made explicit, with the full set of potential Wyckoff sequence solutions being listed, and with some actual representatives being found among the known crystal structures. In general, the number of potential solutions might exhibit a rapid growth

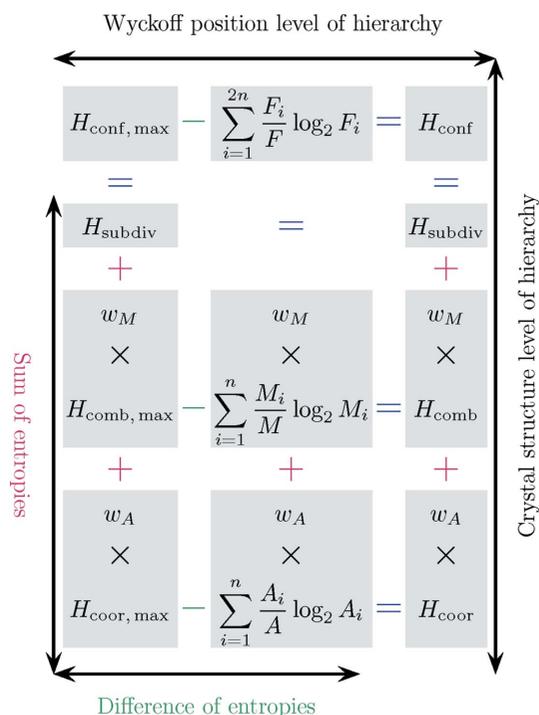


Figure 5

Schematic representation of the interrelation between various maximal and non-maximal entropies (combinatorial, coordinational, configurational) and the subdivision entropy on different levels of structural hierarchy. On the Wyckoff position level of hierarchy the difference of (maximal) entropies defines the conventional non-maximal entropies (from left to right), while on the crystal structure level of hierarchy the (partially weighted) sum of entropies defines the maximal and non-maximal configurational entropy (from bottom to top). Note that the distinction between differences and sums of entropies depends on a deliberate choice of how to distribute terms on either side of the equals sign. In the scheme as presented here, the case for differences of entropies is based on the choice of collecting all maximal entropies together, while distributing collective and individual contributions on opposite sides of the equation could highlight the fact that the non-maximal entropies fulfil the same role as a subdivision complexity on the Wyckoff position level as the subdivision complexity H_{subdiv} does on the crystal structure level, thus highlighting the applicability of the strong additivity property on both levels of hierarchy (in this case the common weighting factors w_M or w_A can be omitted on the Wyckoff position level of hierarchy).

Table 1

Configurational complexities for the six possible Wyckoff sequences (of length k) with 12 combinatorial and three coordinational degrees of freedom of space group type number 113.

Shannon entropies are stated in units of bit per freedom.

Wyckoff sequence	k	$(M, A) = (12, 3)$	H_{subdiv}	H_{comb}	H_{coor}	H_{conf}
113 fba	3	$(8 + 2 + 2, 3 + 0 + 0)$	0.722	1.252	0.000	1.723
113 d^3	3	$(4 + 4 + 4, 1 + 1 + 1)$	0.722	1.585	1.585	2.307
113 $edba$	4	$(4 + 4 + 2 + 2, 2 + 1 + 0 + 0)$	0.722	1.918	0.918	2.440
113 d^2cb	4	$(4 + 4 + 2 + 2, 1 + 1 + 1 + 0)$	0.722	1.918	1.585	2.574
113 d^2ca	4	$(4 + 4 + 2 + 2, 1 + 1 + 1 + 0)$	0.722	1.918	1.585	2.574
113 dc^2ba	5	$(4 + 2 + 2 + 2 + 2, 1 + 1 + 1 + 0 + 0)$	0.722	2.252	1.585	2.840

(combinatorial explosion), while the number of actual representatives lags behind greatly.

4.1. Example 1: space-group type $P\bar{4}2_1m$ (No. 113) revisited

As our first example, we revisit the case of space-group type $P\bar{4}2_1m$ (No. 113). Table 1 compiles some values of specific Shannon entropies for the Wyckoff sequences given in this example, taking into account the combinatorial, coordinational and configurational degrees of freedom, highlighting the constant term H_{subdiv} , with the other Shannon entropies related according to their strong additive sum:

$$H_{\text{conf}} = H_{\text{subdiv}} + (12/15)H_{\text{comb}} + (3/15)H_{\text{coor}}. \quad (41)$$

A search in the Pearson’s Crystal Data Crystal Structure Database for Inorganic Compounds (Villars & Cenzual, 2020) for the space-group type $P\bar{4}2_1m$ (No. 113) reveals a total of 759 Wyckoff sequences for which crystal structure prototypes have been assigned. Reducing for multiple entries arising from multiple crystal structure determinations sharing the same prototype yields 79 entries with a unique Wyckoff sequence/prototype combination. (Notably, 538 entries share the same Wyckoff sequence 113 fe^3ca , with seven distinct prototypes associated to it, of which the $tP24\text{-Ca}_2\text{MgSi}_2\text{O}_7$ prototype alone occurs 307 times, thereby explaining the heavy degree of reduction observed.) Reducing again for distinct prototypes sharing the same Wyckoff sequence yields 63 unique Wyckoff sequences.

Of these, 55 correspond to a unique pair (M, A) of combinatorial and coordinational degrees of freedom, with six pairs occurring two times, and one pair, $(28, 12)$, occurring three times. The associated Wyckoff sequences and prototypes are: fe^4cb (Th_2Se_5 prototype), fe^4ca [$(\text{Ca}_{0.25}\text{La}_{0.75})_2\text{Ga}_3\text{O}_{7.25}$ prototype] and e^4dc^3a ($\text{Ce}_2\text{CoAl}_7\text{Ge}_4/\text{LiSmTiO}_4$ prototypes).

On the basis of this work, one obtains a larger number of 44 possible Wyckoff letter sequence solutions [based on the Wyckoff multiplicities and arities, compare equation (21)], which can be constructed from an exhaustive exploration of a search space of much larger size of 11 648 multiplier choices (representatives with prototypes marked by an asterisk):

$$f^2e^3, fe^4d, e^5d^2, f^2e^2c^2, fe^3dc^2, e^4d^2c^2, f^2ec^4, fe^2dc^4, e^3d^2c^4, f^2c^6, fedc^6, e^2d^2c^6, fdc^8, ed^2c^8, d^2c^{10}, *fe^4cb, e^5dcb, fe^3c^3b, e^4dc^3b, fe^2c^5b, e^3dc^5b, fec^7b, e^2dc^7b, fc^9b, edc^9b, dc^{11}b, *fe^4ca, e^5dca, fe^3c^3a, *e^4dc^3a, fe^2c^5a, e^3dc^5a, fec^7a, e^2dc^7a, fc^9a, edc^9a, dc^{11}a, e^6ba, e^5c^2ba, e^4c^4ba, e^3c^6ba, e^2c^8ba, ec^{10}ba, c^{12}ba. \quad (42)$$

While the combinatorial approach is non-constructive, it might often suffice to only know the number of possible solutions.

4.2. Example 2: space-group type $Fm\bar{3}m$ (No. 225)

As another example, we consider the cubic space-group type $Fm\bar{3}m$ (No. 225) encompassing a total of 12 distinct Wyckoff positions of the following multiplicities and arities:

$$W^{225} = \begin{cases} W_a(1, 0), & W_b(1, 0), & W_c(2, 0), \\ W_d(6, 0), & W_e(6, 1), & W_f(8, 1), \\ W_g(12, 1), & W_h(12, 1), & W_i(12, 1), \\ W_j(24, 2), & W_k(24, 2), & W_l(48, 3). \end{cases} \quad (43)$$

Note that the multiplicities stated here are those for a reduced (primitive) unit cell, while the multiplicities for an F -centred unit cell would be four times larger. Again, all of the four possible values for the Wyckoff arity are present in this example, in addition to seven distinct values for the Wyckoff multiplicity. This example was chosen because it partly corresponds to actual crystal structures.

Now, for the given choice of $(M, A) = (28, 3)$ degrees of freedom one finds 17 matching Wyckoff letter sequences:

$$if^2, hf^2, *gf^2, f^2ed, *ifec, *hfec, gfec, *fe^2dc, *ifeba, hfeba, gfeba, fe^2dba, f^3cba, ie^2cba, he^2cba, *ge^2cba, *e^3dcb. \quad (44)$$

Here, the seven sequences preceded by an asterisk are realized as crystal structures. A compilation of their complexity values, which are interrelated according to

$$H_{\text{conf}} = H_{\text{subdiv}} + (28/31)H_{\text{comb}} + (3/31)H_{\text{coor}}, \quad (45)$$

is given in Table 2. Here, only the combinatorial contribution H_{comb} is responsible for the variable amount of configurational complexity H_{conf} ; the other terms, in particular the value for H_{subdiv} , being constant.

Table 2

Configurational complexities for the seven possible Wyckoff sequences (of length k) with 28 combinatorial and three coordinational degrees of freedom of space group type number 225 which have realizations as crystal structures in nature.

Shannon entropies are stated in units of bit per freedom. Information regarding the structure type for which the Wyckoff sequence is realized is taken from Pearson's Crystal Data Crystal Structure Database for Inorganic Compounds (Villars & Cenzual, 2020). If more than one structure type is associated with the same Wyckoff sequence a selection has been made.

Wyckoff sequence	k	$(M, A) = (28, 3)$	H_{subdiv}	H_{comb}	H_{coor}	H_{conf}	Structure type
225 gf^2	3	(12 + 8 + 8, 1 + 1 + 1)	0.459	1.557	1.585	2.018	$Y_{0.25}Bi_{0.75}O_{1.5}$
225 $ifec$	4	(12 + 8 + 6 + 2, 1 + 1 + 1 + 0)	0.459	1.788	1.585	2.227	$KY_3F_{10}, Zr_3PbO_4F_6$
225 $hfec$	4	(12 + 8 + 6 + 2, 1 + 1 + 1 + 0)	0.459	1.788	1.585	2.227	$Y_{0.27}Bi_{0.73}O_{1.5}$
225 fe^2dc	5	(8 + 6 + 6 + 6 + 2, 1 + 1 + 1 + 0 + 0)	0.459	2.217	1.585	2.615	$Cs_3Re_3S_4I_4$
225 $ifeba$	5	(12 + 8 + 6 + 1 + 1, 1 + 1 + 1 + 0 + 0)	0.459	1.860	1.585	2.292	$(Ag_{0.5}Pd_{0.5})_{11}Se_3$
225 ge^2cba	6	(12 + 6 + 6 + 2 + 1 + 1, 1 + 1 + 1 + 0 + 0 + 0)	0.459	2.092	1.585	2.501	$Cr_{0.8}Mn_{0.2}Mn(CN)_6(H_2O)_4$
225 e^3cba	7	(6 + 6 + 6 + 6 + 2 + 1 + 1, 1 + 1 + 1 + 0 + 0 + 0 + 0)	0.459	2.520	1.585	2.888	$K_{0.04}(VO)Co_{0.88}(CN)_{3.8}(H_2O)_{1.1}$

Table 3

Configurational complexities for the 11 possible Wyckoff sequences (of length k) with 22 combinatorial and ten coordinational degrees of freedom of space group type number 63 which have realizations as crystal structures in nature.

Shannon entropies are stated in units of bit per freedom. Information regarding the structure type for which the Wyckoff sequence is realized is taken from Pearson's Crystal Data Crystal Structure Database for Inorganic Compounds (Villars & Cenzual, 2020). If more than one structure type is associated with the same Wyckoff sequence a selection has been made.

Wyckoff sequence	k	$(M, A) = (22, 10)$	H_{subdiv}	H_{comb}	H_{coor}	H_{conf}	Structure type
63 hf^3c	5	(8 + 4 + 4 + 4 + 2, 3 + 2 + 2 + 2 + 1)	0.896	2.187	2.246	3.101	$SrGe_{5.5}$
63 g^2f^2ec	6	(4 + 4 + 4 + 4 + 4 + 2, 2 + 2 + 2 + 2 + 1 + 1)	0.896	2.550	2.522	3.438	$K_2Zn_5As_4$
63 g^2fec^3	7	(4 + 4 + 4 + 4 + 2 + 2 + 2, 2 + 2 + 2 + 1 + 1 + 1 + 1)	0.896	2.732	2.722	3.625	$[UO_2]Cl_2[NH_3]_6$
63 hf^2c^2	7	(8 + 4 + 2 + 2 + 2 + 2 + 2, 3 + 2 + 1 + 1 + 1 + 1 + 1)	0.896	2.550	2.646	3.476	$Yb_2Mn_{0.33}Si_{3.67}$
63 f^4c^2b	7	(4 + 4 + 4 + 4 + 2 + 2 + 2, 2 + 2 + 2 + 2 + 1 + 1 + 0)	0.896	2.732	2.522	3.562	$K_2Cu_3US_5$
63 gf^3c^2a	7	(4 + 4 + 4 + 4 + 2 + 2 + 2, 2 + 2 + 2 + 2 + 1 + 1 + 0)	0.896	2.732	2.522	3.562	$Zn(Zn_{0.45}Co_{0.55})Co_3Sn_4$
63 f^4c^2a	7	(4 + 4 + 4 + 4 + 2 + 2 + 2, 2 + 2 + 2 + 2 + 1 + 1 + 0)	0.896	2.732	2.522	3.562	$CaFe_4O_6$
63 hc^7	8	(8 + 2 + 2 + 2 + 2 + 2 + 2 + 2, 3 + 1 + 1 + 1 + 1 + 1 + 1 + 1)	0.896	2.732	2.846	3.664	$KNaOs[NO][F_3][H_2O]$
63 g^2fc^4a	8	(4 + 4 + 4 + 2 + 2 + 2 + 2 + 2, 2 + 2 + 2 + 1 + 1 + 1 + 1 + 0)	0.896	2.914	2.722	3.750	$\beta-U_3O_8$
63 f^3c^4a	8	(4 + 4 + 4 + 2 + 2 + 2 + 2 + 2, 2 + 2 + 2 + 1 + 1 + 1 + 1 + 0)	0.896	2.914	2.722	3.750	$Cs_2Cu_5Se_4$
63 f^2c^6a	9	(4 + 4 + 2 + 2 + 2 + 2 + 2 + 2 + 2, 2 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 0)	0.896	3.096	2.922	3.938	$Sr_2Ta_2O_7$

If one performs the same calculation with the additional restriction of the Wyckoff sequence length to the value $k = 5$ one obtains the number of four solutions, which the reader can easily check on the list given above.

4.3. Example 3: space-group type $Cmcm$ (No. 63)

As a final example, we consider the orthorhombic space-group type $Cmcm$ (No. 63) encompassing a total of eight distinct Wyckoff positions of the following multiplicities and arities:

$$\mathcal{W}^{63} = \begin{cases} \mathcal{W}_a(2, 0), & \mathcal{W}_b(2, 0), & \mathcal{W}_c(2, 1), \\ \mathcal{W}_d(4, 0), & \mathcal{W}_e(4, 1), & \mathcal{W}_f(4, 2), \\ \mathcal{W}_g(4, 2), & \mathcal{W}_h(8, 3). \end{cases} \quad (46)$$

Note that the multiplicities stated here are those for a reduced (primitive) unit cell, while the multiplicities for a C -centred unit cell would be two times larger. Again, all of the four possible values for the Wyckoff arity are present in this example, in addition to three distinct values for the Wyckoff multiplicity.

Now, for the given choice of $(M, A) = (22, 10)$ degrees of freedom, one finds 67 matching Wyckoff letter sequences:

$$\begin{aligned} &hg^3c, hg^2fc, hgf^2c, *hf^3c, g^4ec, g^3fec, *g^2f^2ec, \\ &gf^3ec, f^4ec, hg^2c^3, hgf^2c^3, hf^2c^3, g^3ec^3, *g^2fec^3, \\ &gf^2ec^3, f^3ec^3, hgc^5, *hfc^5, g^2ec^5, gfc^5, f^2ec^5, \\ &*hc^7, gec^7, fec^7, ec^9, g^5b, g^4fb, g^3f^2b, \\ &g^2f^3b, gf^4b, f^5b, g^4c^2b, g^3fc^2b, g^2f^2c^2b, gf^3c^2b, \\ &*f^4c^2b, g^3c^4b, g^2fc^4b, gf^2c^4b, f^3c^4b, g^2c^6b, gfc^6b, \\ &f^2c^6b, gc^8b, fc^8b, c^{10}b, g^5a, g^4fa, g^3f^2a, \\ &g^2f^3a, gf^4a, f^5a, g^4c^2a, g^3fc^2a, g^2f^2c^2a, *gf^3c^2a, \\ &*f^4c^2a, g^3c^4a, *g^2fc^4a, gf^2c^4a, *f^3c^4a, g^2c^6a, gfc^6a, \\ &*f^2c^6a, gc^8a, fc^8a, c^{10}a. \end{aligned} \quad (47)$$

Here, the 11 sequences preceded by an asterisk are realized as crystal structures. A compilation of their complexity values, which are interrelated according to

$$H_{\text{conf}} = H_{\text{subdiv}} + (22/32)H_{\text{comb}} + (10/32)H_{\text{coor}}, \quad (48)$$

is given in Table 3. Here, both the combinatorial and coordinational degrees of freedom are responsible for the variable amount of configurational complexity H_{conf} .

5. Conclusion

As a contribution to the combinatorics of Wyckoff sequences, we have presented two methods to calculate their number for a fixed space group, given a pair of combinatorial and coordinational total degrees of freedom, and, optionally, their length. The first method is based on a generating polynomial approach (see Sections 2.4 and 2.7 for the key results), while the second makes use of a dynamic programming algorithm (Section 2.8). While the generating polynomial approach appears to be conceptually easier to understand, the dynamic programming algorithm is considerably better in its computational performance. The methods have been exemplified on cases of ideal and actual crystal structures with invariant subdivision complexity and variable configurational complexity in the sense of Hornfeck (2020), thus relating the combinatorics of Wyckoff sequences to the complexities of crystal structures.

APPENDIX A

Number of Wyckoff sequences of length k

The number of Wyckoff sequences of length k is given as

$$W(\nu, \varphi, k) = \sum_{i=0}^k \binom{\nu+i-1}{i} \binom{\varphi}{k-i} \quad (49)$$

in which ν and φ , respectively, denote the number of non-fixed and fixed Wyckoff positions of a given space group [equation (7) in Hornfeck (2022a)]. In the worst case $\nu = 19$ and $\varphi = 8$, thus

$$\sum_{k=1}^{50} W(19, 8, k) = 3\,519\,297\,892\,616\,574\,305 \quad (50)$$

gives the number of distinct Wyckoff sequences up to $k = 50$ inclusive.

APPENDIX B

The classical coin change problem

Assume you are in the possession of the following multiset of coins: three coins of value $\nu = 1$, four coins of value $\nu = 2$, two coins of value $\nu = 5$. How many ways exist to pay a total price of value $V = 12$?

Let S_ν denote the individual sum of coins of value ν , hence

$$V = S_1 + S_2 + S_5 = 12. \quad (51)$$

The values which each individual sum may take are determined and limited in number by the number of available coins of each nominal value. Thus, one gets

$$\begin{aligned} S_1 &\in \{0, 1, 2, 3\}, \\ S_2 &\in \{0, 2, 4, 6, 8\}, \\ S_5 &\in \{0, 5, 10\} \end{aligned} \quad (52)$$

as all possible values. Now, the trick is that one identifies these values with the exponents of a couple of generating polynomials, one for each type of coin, namely

$$\begin{aligned} P_1(x) &= 1 + x + x^2 + x^3, \\ P_2(x) &= 1 + x^2 + x^4 + x^6 + x^8 \\ P_5(x) &= 1 + x^5 + x^{10} \end{aligned} \quad (53)$$

in which $x^0 = 1$ by definition. From these individual polynomials one forms their product

$$\mathcal{P}(x) = P_1(x)P_2(x)P_5(x) \quad (54)$$

and analyses its expansion

$$\begin{aligned} &1 + x + 2x^2 + 2x^3 + 2x^4 + 3x^5 + 3x^6 + 4x^7 \\ &+ 4x^8 + 4x^9 + 4x^{10} + 4x^{11} + 4x^{12} + 4x^{13} + 4x^{14} \\ &+ 3x^{15} + 3x^{16} + 2x^{17} + 2x^{18} + 2x^{19} + x^{20} + x^{21}. \end{aligned} \quad (55)$$

The coefficient of x^{12} in $\mathcal{P}(x)$ gives the desired solution, here it is equal to four, describing the four cases (i) $2 \times 1 + 2 \times 5$, (ii) $1 \times 2 + 2 \times 5$, (iii) $3 \times 1 + 2 \times 2 + 1 \times 5$ and (iv) $1 \times 1 + 3 \times 2 + 1 \times 5$, all amounting to 12.

A generalization is possible, in case the number of coins available for all given values ν is assumed to be infinite. Then, the generating polynomials are replaced by the respective generating functions of a geometric series,

$$F_\nu(x) = 1 + x^\nu + x^{2\nu} + x^{3\nu} + \dots = \frac{1}{1 - x^\nu}, \quad (56)$$

and the solution to the problem is given by their product

$$\mathcal{F}(x) = F_1(x)F_2(x)F_5(x) = \frac{1}{(1-x)(1-x^2)(1-x^5)} \quad (57)$$

in the same way as before, namely as the value of the coefficient $[x^V]$ in $\mathcal{F}(x)$, in which V denotes the total price. In the infinite case the result is $[x^{12}]\mathcal{F}(x) = 13$, necessarily containing at least as many or even more solutions than were found in the finite case. In this general form the problem has been addressed by Pólya (1956). A closed form solution for the univariate infinite case has been described by Graham *et al.* (1994, pp. 327–330 and pp. 344–346).

Note that

$$\prod_{i=1}^{\infty} \frac{1}{1 - x^i} \quad (58)$$

is the generating function of the number-theoretic integer partition function (Alfonsín, 2005, p. 71), and our problem asks for the number of restricted partitions into specified, possibly repeated parts (a finite or infinite number of coins with a finite set of distinct denominations). In this context the number of non-negative integer solutions is known as Sylvester's denumerant (Sylvester, 1857) which has a rich history in number theory, partly due to the difficulty of obtaining exact results (Alfonsín, 2005, ch. 4).

Finally, it should be emphasized that an elegant method of solving this problem in practice is given by the concept of dynamical programming. Since the problem again and again reduces to simpler subproblems of the same kind, the solu-

tions of these subproblems can be stored whenever they occur for the first time. Moreover, using a bottom-up approach (top-down would be equally feasible), the solutions for subproblems can be systematically generated from the already known solutions. For this purpose the function

```
def coins(n):
    values = [ 1, 2, 5 ]
    result = [ 1 ] + [ 0 ] * n
    for i in range( 0, len(values) ):
        for j in range( 0, ( n+1 ) - values[i] ):
            result[ j+values[i] ] += result[j]
    return result[n]
```

here given in a Python implementation is initialized with a result vector of length $n + 1$ of the form $(1, 0, 0, \dots, 0)$. This result vector is subsequently updated in a separate loop for every coin value. The final value at position n of the result vector then is the solution to the (infinite) coin problem described before (for $n = 12$ the result is equal to 13).

APPENDIX C Code examples

Figs. 6 and 7 give working examples for the calculation of the number of Wyckoff sequences with the same subdivision complexity by means of the generating polynomial and dynamic programming approaches. The input values are taken from the example described in Section 2.6, with the output being equal to six.

Acknowledgements

The author thanks Philipp Kuhn for discussions, one reviewer for helpful comments which greatly improved the manuscript, Michal Dušek for his kind endorsement, as well as the organizers of the conference *Mathematics and Computer Science for Materials Innovation* (MACSMIN), Vitaliy Kurlin and Mois Aroyo, for the invitation to present this work at MACSMIN-2022 (Crystal Lattice Classifications) on location at the Materials Innovation Factory, Liverpool, United Kingdom.

```
case[M_,A_,W_] :=
{
  Mi[i_] := W[[i]][[1]];
  Ai[i_] := W[[i]][[2]];
  ri[i_] := Min[{Floor[M/Mi[i]], Floor[A/Ai[i]]}];
  mui[i_] := If[Ai[i] == 0, 1, ri[i]];
  S[i_] := Sum[x^(j*Mi[i])*y^(j*Ai[i]), {j, 0, mui[i]}];
  P = Table[S[i], {i, 1, Length[W]}];
  coeff = Coefficient[Expand[Apply[Times, P]], x^M*y^A];
  MatrixForm[{M, A, W, coeff}]
}
case[12, 3, {{2, 0}, {2, 0}, {2, 1}, {4, 1}, {4, 2}, {8, 3}}]
```

Figure 6
Generating polynomial approach (*Mathematica*).

```
from math import floor
def case(target_tuple, wyckoff_tuples):
    M=target_tuple[0]
    A=target_tuple[1]
    Ir=range(0,M+1)
    Jr=range(0,A+1)
    T=[[0 for j in Jr] for i in Ir]
    T[0][0]=1
    N=len(wyckoff_tuples)
    for i in range(0,N):
        M_i=wyckoff_tuples[i][0]
        A_i=wyckoff_tuples[i][1]
        if A_i==0:
            f_i=1
        else:
            f_i=min(floor(M/M_i), floor(A/A_i))
        M_range=range(M-M_i,-1,-1)
        A_range=range(A-A_i,-1,-1)
        for u in M_range:
            for v in A_range:
                if T[u][v]>0:
                    for f in range(1,f_i+1):
                        M_f=u+f*M_i
                        A_f=v+f*A_i
                        if M_f<=M and A_f<=A:
                            T[M_f][A_f]+=T[u][v]
    return T[M][A]
```

Figure 7
Dynamic programming approach (Python). Here, the target tuple is given as $[M, A] = [12, 3]$ and the Wyckoff tuples are given as $\mathcal{W} = [[2, 0], [2, 0], [2, 1], [4, 1], [4, 2], [8, 3]]$.

Funding information

This work was supported by the Czech Science Foundation, project No. 21-05926X. We also acknowledge CzechNanoLab Research Infrastructure supported by MEYS CR (LM2018110).

References

- Alfonsín, J. L. R. (2005). *The Diophantine Frobenius Problem*. Oxford: Oxford University Press.
- Allmann, R. & Hinek, R. (2007). *Acta Cryst.* **A63**, 412–417.
- Aroyo, M. I. (2016). Editor. *International Tables for Crystallography, Vol. A, Space-group Symmetry*, 6th ed. Chichester: Wiley.
- Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M. & Wondratschek, H. (2006a). *Acta Cryst.* **A62**, 115–128.
- Aroyo, M. I., Perez-Mato, J. M., Capillas, C., Kroumova, E., Ivantchev, S., Madariaga, G., Kirov, A. & Wondratschek, H. (2006b). *Z. Kristallogr.* **221**, 15–27.
- Aroyo, M. I., Perez-Mato, J. M., Orobengoa, D., Tasci, E., de la Flor, G. & Kirov, A. (2011). *Bulg. Chem. Commun.* **43**, 183–197.
- Brown, H., Bülow, R., Neubüser, J., Wondratschek, H. & Zassenhaus, H. (1978). *Crystallographic Groups of Four Dimensional Space*. New York: Wiley.
- Gelato, L. M. & Parthé, E. (1987). *J. Appl. Cryst.* **20**, 139–143.
- Goodall, R. E. A., Parackal, A. S., Faber, F. A., Armiento, R. & Lee, A. A. (2022). *Sci. Adv.* **8**, eabn4117.
- Graham, R. L., Knuth, D. L. & Patashnik, O. (1994). *Concrete Mathematics – a Foundation for Computer Science*, 2nd ed. New York: Addison-Wesley.
- Hahn, T. (2005). Editor. *International Tables for Crystallography, Vol. A, Space-group symmetry*. 5th ed. Heidelberg: Springer.

- Hornfeck, W. (2020). *Acta Cryst.* **A76**, 534–548.
- Hornfeck, W. (2022a). *Acta Cryst.* **A78**, 149–154.
- Hornfeck, W. (2022b). *Z. Kristallogr.* **237**, 127–134.
- Krivovichev, S. (2012). *Acta Cryst.* **A68**, 393–398.
- Krivovichev, S. V. (2014). *Angew. Chem. Int. Ed.* **53**, 654–661.
- Kruchinin, D., Kruchinin, V. & Shablya, Y. (2021). *Mathematics*, **9**, 428.
- Lima-de-Faria, J., Hellner, E., Liebau, F., Makovicky, E. & Parthé, E. (1990). *Acta Cryst.* **A46**, 1–11.
- Marcus, D. A. (1998). *Combinatorics – a Problem Oriented Approach*. Washington, D. C.: The Mathematical Association of America.
- Müller, U. (2013). *Symmetry Relationships between Crystal Structures – Applications of Crystallographic Group Theory in Crystal Chemistry*. Oxford: Oxford University Press.
- Nespolo, M., Aroyo, M. I. & Souvignier, B. (2018). *J. Appl. Cryst.* **51**, 1481–1491.
- Parthé, E., Cenzual, K. & Gladyshevskii, R. E. (1993a). *J. Alloys Compd.* **197**, 291–301.
- Parthé, E., Gelato, L., Chabot, B., Penzo, M., Cenzual, K. & Gladyshevskii, R. (1993b). *TYPIX – Standardized Data and Crystal Chemical Characterization of Inorganic Structure Types* (Vol. 1). In *Gmelin Handbook of Inorganic and Organometallic Chemistry*, 8th ed. Berlin: Springer.
- Parthé, E. & Gelato, L. M. (1984). *Acta Cryst.* **A40**, 169–183.
- Parthé, E. & Gelato, L. M. (1985). *Acta Cryst.* **A41**, 142–151.
- Pólya, G. (1956). *Am. Math. Mon.* **63**, 689–697.
- Sylvester, J. J. (1857). *Q. J. Pure Appl. Math.* **1**, 141–152.
- Villars, P. & Cenzual, K. (2020). *Pearson's Crystal Data Crystal Structure Database for Inorganic Compounds*. ASM International, Materials Park, Ohio, USA. <https://www.crystalimpact.com/pcd/Default.htm>.
- Wilf, H. S. (2006). *Generatingfunctionology*, 3rd ed. Wellesley: A. K. Peters.