On the use of symmetry in the configurational analysis for the simulation of disordered solids

Sami Mustapha[1], Philippe D'Arco[2], Marco De La Pierre[3], Yves Noël[2], Matteo Ferrabone[3] and Roberto Dovesi[3]

[1] Institut de Mathématiques de Jussieu (UMR 7586 UPMC-CNRS), UPMC, Sorbonne Universités, Paris, France

[2] Institut des Sciences de la Terre de Paris (UMR 7193 UPMC-CNRS), UPMC, Sorbonne Universités, Paris, France

[3] Dipartimento di Chimica, Università di Torino and NIS - Nanostructured Interfaces and Surfaces - Centre of Excellence, http://www.nis.unito.it, Via P. Giuria 7, 10125 Torino, Italy

E-mail: philippe.d_arco@upmc.fr

Abstract. The starting point for a quantum mechanical investigation of disordered systems usually implies calculations on a limited subset of configurations, generated by defining either the composition of interest or a set of compositions ranging from one end member to another, within an appropriate supercell of the primitive cell of the pure compound. The way symmetry can be used in the identification of symmetry independent configurations (SICs) is here discussed. First, Pólya's enumeration theory is adopted to determine the number of SICs, in the case of both varying and fixed composition, for colors in number of two or higher. Then, De Bruijn's generalization is presented, which allows to analyze the case where colors are symmetry related, e.g. spin up and down in magnetic systems. In spite of their efficiency in counting SICs, neither Pólya's nor De Bruijn's theories do help in solving the difficult problem of identifying the complete list of SICs. SICs representatives are here obtained by adopting an orderly generation approach, based on the lexicographic ordering, that offers the advantage of avoiding the (computationally expensive) analysis and storage of all the possible configurations. When the number of colors increases, this strategy can be combined with the surjective resolution principle, that permits to efficiently generate SICs of a problem in $|R|$ colors starting from the ones obtained for the $(|R|-1)$-colors case. The whole scheme is documented by means of three examples: the abstract case of the square with C_{4v} symmetry and the real cases of garnet and olivine mineral families.

PACS numbers: 02.20.-a, 61.43.-j, 61.50.Ah

Keywords: solid solution, configuration, symmetry exploitation, group theory, Pólya theory, De Bruijn theory, lexicographic order, surjective resolution, CRYSTAL code

Submitted to: *J. Phys.: Condens. Matter*

1. Introduction

Quantum mechanical computer simulation of the atomistic behavior of solids is one of the most successful methodologies employed in materials science. Nowadays this class of methods is routinely applied to ordered crystalline phases in their equilibrium states. The total and formation energies, the equilibrium geometry, the vibrational spectrum, the dielectric and polarizability tensors as well as many other properties can be evaluated routinely for periodic systems with unit cells containing up to 1000 atoms [1, 2] and even more [3].

However, the availability of the same properties for disordered systems and/or non-equilibrium states still remains an outstanding challenge. Yet these systems have a major importance for both Earth and materials science: nearly all rock-forming minerals are solid solutions (substitutional disorder); many technologically relevant materials are non stoichiometric (occupational and/or substitutional disorder) or magnetic (possible spin disorder).

In the past, various techniques of increasing complexity have been proposed and implemented for the simulation of disordered solids. The schemes that in the last two decades were shown to provide the most promising results are based on the description of these systems as a weighted average of ordered configurations. A further step forwards consists in using the energies of some of these configurations (obtained from "accurate", possibly *ab initio*, calculations) in a simple model that permits to estimate at very low cost the energies of many other configurations and to use them in a self-consistent manner to select other low energy configurations to be investigated quantum-mechanically. In this way, hopefully, a relatively low number of "accurate" calculations is sufficient for the complete description of the thermodynamic properties of the disordered system [4, 5, 6, 7, 8, 9, 10].

If the number of involved positions is $|D|$ (the positions being elements of the set D) and the number of species is $|R|$ (the species are the elements of the set R) the total number of configurations over the complete range of compositions is $N = |R|^{|D|}$ (see Section 2.1 for complete notations). This requires a large computer time in generating the set of configurations, and most of all a huge, unsustainable computational cost to treat the whole of this set at an *ab initio* level.

In crystalline structures (any dimensionality), the symmetry, whatever it is, induces a partition of the R^D set of configurations in equivalent classes. Then, in order to fully characterize the configurations, one needs to know the number of equivalence classes and one representative per class. The set of these representatives is the *smallest possible* subset of symmetry independent configurations (SICs).

The automatic generation of a set of SICs permits on the one hand to avoid the repetition of quantum mechanical calculations for equivalent configurations, on the other hand to automatically submit parallel quantum mechanical calculations and collect the related data, removing errors due to by hand data manipulation.

The simulation schemes mentioned above become reliable when large $|D|$ (that means

large cells) are used. However the enumeration of the SICs becomes rapidly a challenging problem when $|D|$ increases.

In the past, averaged structures described by small cells (very few atoms) have been considered, in most of the cases with reference to metallic alloys (see for example the excellent paper by Ferreira *et al.* [11], references therein and also Refs. [12, 13]). The too small number of atom(s) in the cell corresponding to the averaged structure calls for the use of larger derived cells (supercells). Ferreira *et al.* (1991)[11] considered supercells derived from FCC and BCC Bravais lattice (1 atom per primitive cell). The problem was split in two steps: 1) identify the independent sublattices corresponding to an index as large as possible, in order to consider a large number of involved positions; 2) for a given sublattice, enumerate symmetry independent structures corresponding to the complete set of SICs. In this analysis, the atoms are all equivalent by translation. The translational symmetry plays then the major role in finding SICs. The sublattices are generated using the geometric "smallest first" approach and for any generated sublattice the calculation of all the corresponding SICs is performed through an $\mathcal{O}(N^2)$ algorithm. For an example of implementation, the reader is referred to Van de Walle and Ceder [14].

Rutherford in a series of papers [15, 16, 17] evidentiated the role of Pólya's theory in the counting of independent coloring patterns of sublattices of a given index by means of the translation group.

Recently, Hart and Forcade (2008)[18] improved dramatically the approach by Ferreira *et al.* [11]. Making a systematic use of the so-called Hermite normal form (HNF) of integer matrices to identify the independent sublattices and the corresponding diagonal Smith normal form to determine the independent atomic configurations, they built a very fast algorithm which scales linearly with the number of unique structures. The rotational lattice symmetry allows to select inequivalent HNF matrices (or supercells) and the translation symmetry permits an easy identification of equivalent labeling for supercell of a given index. In 2009, they [19] (HF09) introduced a multilattice model in order to extend the applicability of their method to cases where the parent lattice is not a Bravais (i.e. simple) lattice. Both rotational symmetry of the lattice and symmetry of the multilattice are used to enumerate the SIC's up to a given index.

An alternative and more natural description of the symmetry acting on periodic structures uses space groups. For the solid state community (physicists, chemists and mineralogists) the space group is the standard starting point for describing the high symmetry structure and its disordered derivatives. Space groups are very useful for complex unit cells in which disorder (for example substitutional) occurs only at some sites, because they properly account for the details of the structures and for the neighborhoods of the involved sites. In other words, they account for positions not involved in the disorder. If one considers only the sites involved in the disorder (the set denoted D in HF09 [19]) the group of symmetry ($\mathcal{S}$, see appendix in Ref. [19]) could be higher than the actual one and non genuine equivalences would appear. On the contrary, if the not involved sites are considered to establish the group of symmetry,

then only the action of $\mathcal{S}$ on the involved sites has to be considered. This point is not discussed in HF09 [19].

For many crystalline compounds the cell contains a large number of atoms sitting in positions involved or not in the substitutions (see the garnet example below). For such compounds, the number of configurations is rapidly prohibitive (see Table 5 for the conventional cell of garnet $|D|$=16) also for supercells of modest index. In these cases, the relative weight of the two steps proposed by Ferreira *et al.* [11] differs significantly from simple cases, and the search for independent structures becomes dominant. This point of view was developed by Grau-Crespo *et al.* [20] and applied in a series of works to large cell minerals [21, 22, 23]. In their enumeration algorithm, the operations of the space group play a crucial role. The runtime scales however quadratically, $\mathcal{O}(N^2)$. These authors emphasize the importance of the multiplicity of the configurations to compute the average properties at given composition. This is achieved enumerating the equivalent configurations of any new SICs by applying every operation of the group to it. A crucial feature of their implementation is the use of two large tables containing the $|R|^{|D|}$ configurations. These authors underline that it is impractical to properly compute the average properties for a very large configuration space (at a quantum mechanical level and even with classical potentials). In such cases, they propose to use random sampling methods to study the configurations at a given composition [21]. The significance of the averaged values were checked with respect to the sample size.

The main goal of this work (whose formalism has been implemented in a development version of the CRYSTAL code [24]) is to enumerate the symmetrically independent configurations (SICs) within a given cell containing a set D of sites involved in solid solutions or disorder without restriction on the number of involved equivalent or independent positions ($|D| = 1, 2, 3, ...$) among other non involved positions and for any number of species (colors) $2 \leq |R| \leq |D|$. With respect to Hart and Forcade's algorithm [18], the present approach scales linearly with the size of the configuration space. At variance with respect to direct methods, long lists of configurations are not necessary. We have chosen to describe the symmetry of the parent structure using space groups (G) acting on D because, as previously explained, it accounts for positions not belonging to D. Configurations are identified by mappings from the set D to a set of atoms or colors. In the search of the SICs, a stopping rule provided by the Pólya – Redfield's theory is used, that reduces the runtime. Within this theoretical framework, the tools allowing efficient analysis of the configurations (orbits, stabilizers, cycle structure of symmetry operations, ...) are easily generated (e.g. the multiplicity of a configuration is obtained directly from its stabilizer). The robust theoretical foundations set up at this stage are necessary to understand the construction behind the "surjective resolution" principle used here to handle more than 2-colors configurations without exploring the full set of $|R|^{|D|}$ configurations. On this basis, the linear scaling of the present approach will be proved using Oberschelp's formula which gives the asymptotic behavior of the number of SICs for large cells. New developments such as symmetric selection (based on the structure of the stabilizers) and random sampling of the derivative structures

space (not included in the present paper, but very useful to explore tremendously large configuration spaces [25]) will appear as natural extensions of the present formalism.
The method here proposed is described in Section 2. The problem of the determination of the number of SICs is tackled in subsection 2.1, where Pólya's theory is shortly summarized, in the case a range of compositions needs to be analyzed, spanning from one end member to the other. Subsections 2.2 and 2.3 discuss two special cases, namely the fixed composition and the one where the two "colors" are symmetry related (De Bruijn's generalization). The algorithmic aspects are described in 2.4 . They combine the so-called lexicographic ordering, which allows orderly generation, with the surjective resolution principle to produce SICs representatives. The scaling behavior of the approach with respect to both the number of SICs and the size of the symmetry group is discussed in subsection 2.5. Section 3 applies the proposed scheme to two examples that are of great geochemical interest: garnets and olivines. Finally, the main conclusions are drawn in Section 4.

2. Method

In this Section the methodological aspects of group theory and combinatorics that are used to identify and characterize SICs are presented.
The first target is to evaluate the number of SICs. This problem can be solved by using Pólya's enumeration theory [26], that is based on the cycle structure of the symmetry operators acting on a set of objects D and gives a systematic method to count the number of non equivalent colorings of D.
Note that Pólya's theory refers to D as a finite set of objects, while crystalline solids are usually modeled as infinite periodic systems, i.e. they show an infinite set of atomic sites D'. However, thanks to periodicity, it is always possible to decompose a symmetry operation in a point symmetry operation plus a translation. Translational symmetry allows to describe all the point symmetry properties of the crystal by considering a finite, small subset D of the whole set of sites D', i.e. a unit cell. The simplest unit cell that can be chosen is the smallest translational set of sites, which is called primitive cell. Otherwise, a supercell of the primitive cell can be chosen as a reference.
The two sets of examples used to illustrate definitions and formulas appearing in the proposed algorithms are shown in Figure 1. Both of them are based on a two-dimensional square lattice under the action of the 8 symmetry operators of the C_{4v} group. In the case of Figure 1-A (model A), the unit cell contains 4 symmetry-related sites, lying on the σ_d reflection planes. By applying all the 8 symmetry operators to, for example, site "1", the full set of 4 sites is obtained. The full set of $2^4 = 16$ configurations that can be obtained for model A by using 2 colors is represented in Figure 2.
Model B (Figure 1-B) represents a more general case of C_{4v} symmetry acting on the square lattice. The unit cell contains 8 symmetry-related sites, which all lie in a general position, i.e. not on a symmetry element.

2.1. Pólya's enumeration theory

Let D denote a finite set of objects, for example the 4 sites of model A in Figure 1-A, $D = \{1, 2, 3, 4\}$, with $|D| = 4$, and let R denote a set of colors, e.g. R={blue,red}={b,r}, $|R| = 2$ (2-colors case). In what follows, the notion $|E|$ denotes the cardinal of a given set E. *Color* is the common terminology in combinatorics for the property that makes the elements of D distinguishable; in our framework it can be the chemical species or the spin status of the atoms occupying the lattice sites.
Let $S = R^D$ be the set of all mappings s from D to R that associate colors to objects. Each s is called a coloring of the set of objects, or a *configuration*. In the case of 2 colors over the 4 sites of model A, the $2^4 = 16$ possible configurations are shown in Figure 2. For $r \in R$ we shall denote $s^{-1}(r) = \{x \in D;\, s(x) = r\}$ the preimage of r, which associates to r the set of objects of D colored by this color. We shall refer to the $|R|$-plet $(|s^{-1}(r_1)|, .., |s^{-1}(r_{|R|})|)$ as the color-pattern, or *composition*, of the configuration s. Note that for a given s it comes $\sum_{j=1}^{|R|} |s^{-1}(r_j)| = |D|$. In the case of model A, for example for the first configuration in the second row of Figure 2 we have: $s^{-1}(b) = \{2, 3, 4\}$, $s^{-1}(r) = \{1\}$, the composition being $(3, 1)$.
Now, let G denote a group of symmetry operations g acting on the set D. In the examples of Figure 1, this group (C_{4v}) consists of the rotations and reflections which leave the set of 4 (8) sites of model A (B) invariant.
The action of G on D induces an action of G on S (called the Pólya's action) defined by:

$$(g \cdot s)(x) = s(g^{-1}x)\ ;\quad g \in G\ ,\quad x \in D. \tag{1}$$

In order to analyze this action, G must be seen as a subgroup of the permutations of D and the cycle decomposition of the symmetry operations $g \in G$ acting on D must be performed. The cycle decomposition induces a partition of D, that is a division of D into non-overlapping and non-empty parts, that covers all of D. We identify this partition with the cycle structure of g acting on D, and denote it $Cyc_D(g)$. The number of cycles describing the action of g on D is the number of elements of the partition: $|Cyc_D(g)|$. As an example, we consider the effect of the C_2 operator on model A (third line in Table 1). By applying C_2 to site "1", it goes to "3"; by applying C_2 to "3", it goes to "1". By applying C_2 to "2", it goes to "4"; finally, "4" goes to "2" ("Moves" in Table 1). We say that the cycle decomposition of C_2 acting on the set D of model A results in 2 cycles ("Cycles" in Table 1).
Pólya's enumeration theory relies on the following general concepts.

Orbits The group orbit of an element $s \in S$ is the set of configurations obtained by applying all the elements $g \in G$ to s:

$$\Omega(s) = \{g \cdot s \in S; g \in G\} \tag{2}$$

The set of all orbits $\Omega(s)$ of S under the action of G forms a partition of S. In the configurational analysis of disordered solids, an orbit is a class of symmetry equivalent

configurations. In Figure 2, all the 16 configurations of model A are shown. Each row represents an orbit (they are 6 in total). For each orbit, it is convenient to select one of its elements as its "canonical" representative[27]; the first found configuration of each orbit is here selected to play this role.

Stabilizers The stabilizer of an element s consists of the set of all the operators $g \in G$ that send s to itself:

$$G_s = \{g \in G; g \cdot s = s\} \tag{3}$$

Any stabilizer G_s is a subgroup of G. Note that all elements of a given orbit have conjugated stabilizers. For the 6 orbits obtained for model A, stabilizers are shown to the right in Figure 2.

Fixed points Fixed or invariant configurations of an operation $g \in G$ are the elements of a subset of S:

$$S_g = \{s \in S; g \cdot s = s\} \tag{4}$$

The fixed points of g are all the configurations whose stabilizer contains g. For example, for model A the fixed points of C_2 are the configurations belonging to orbits 1,4 and 6 (see Figure 2).

A general property of an orbit $\Omega(s)$ of a configuration s is that it can be mapped to the set of left cosets of the stabilizer G_s in a bijective way [28, 29]. This permits to easily know the length of the orbit, once the cardinal of G_s is known:

$$|\Omega(s)| = \frac{|G|}{|G_s|} \tag{5}$$

For model A (see Figure 2), the stabilizers of the six orbits contain 8, 2, 2, 4, 2 and 8 elements, respectively. Thus, the corresponding orbit lengths $|\Omega(s)|$ are 1, 4, 4, 2, 4 and 1, respectively.

Now, assume that a procedure for selecting canonical representatives is available. Let $\Delta(S)$ denote the set of canonical representatives for the orbits $\Omega(s)$ of the action of G on S, and let us introduce $W(s)$, a G-invariant weight defined on S (this means that W is constant on each orbit). Definitions (3) and (4) imply:

$$\sum_{s \in S} \sum_{g \in G_s} W(s) = \sum_{g \in G} \sum_{s \in S_g} W(s) \tag{6}$$

Using the fact that the elements of an orbit have stabilizers with the same cardinal and share the same value for W, we can factorize by orbits (i.e. by canonical representatives). Exploiting (5), we deduce:

$$\sum_{s' \in \Delta(S)} W(s') = \frac{1}{|G|} \sum_{g \in G} \sum_{s \in S_g} W(s) \tag{7}$$

Note that the l.h.s. of Eq. (7) is a sum of $|\Delta(S)|$ terms, i.e. it has as many terms as the number of canonical representatives, thus of orbits, of S. We can obtain $|\Delta(S)|$ by taking $W(s) = 1, \forall\ s \in S$:

$$|\Delta(S)| = \frac{1}{|G|}\sum_{g\in G}|S_g|. \tag{8}$$

This is the Cauchy-Frobenius Lemma often named the Burnside Lemma. In order to evaluate $|S_g|$ we observe that there is a natural correspondence between the set of fixed points $s \in S_g$ and the set of all mappings from $Cyc_D(g)$ to R. The reason for this correspondence is that, considering the cycle structure of g, we can state that s is stabilized by g if and only if every cycle of g has all its elements mapped to one and only one color. This implies that $S_g \cong R^{Cyc_D(g)}$, so the cardinal of S_g is:

$$|S_g| = |R|^{|Cyc_D(g)|} \tag{9}$$

In order to illustrate this result, in our model A we consider the set S_{C_2} of all the configurations unchanged by the operator C_2: they are four in orbits 1, 4 and 6 (see Figure 2). The cycle structure of C_2 from Table 1 shows 2 cycles: (13)(24). To build a stabilized configuration the elements of cycles (13) and (24) must be mapped onto the same color, i.e. in this 2-colors case: (bb),(bb); (bb),(rr); (rr),(bb); (rr),(rr), that are 4 configurations as obtained applying Eq. (9).
Now, substituting Eq. (9) in Eq. (8) we deduce

$$|\Delta(S)| = \frac{1}{|G|}\sum_{g\in G}|R|^{|Cyc_D(g)|} \tag{10}$$

which is the Pólya's counting formula for the SICs. For convenience, in the following applications we will use $N_{|R|}^{|D|}$ for $|\Delta(S)|$, in order to make more evident the dependence on $|D|$ and $|R|$.
Tables 1 and 2 provide the set of $|Cyc_D(g)|$ values in the case of models A and B, respectively, from which the number of configurations can be calculated for any number of colors. As an example, for model A (4 sites) in the case of 2 and 3 colors we have, respectively :

$$N_2^4 = \frac{1}{8}(\underbrace{2^4}_{E} + \underbrace{2\cdot 2^1}_{C_4} + \underbrace{2^2}_{C_2} + \underbrace{2\cdot 2^2}_{\sigma_v} + \underbrace{2\cdot 2^3}_{\sigma_d}) = 6 \tag{11}$$

$$N_3^4 = \frac{1}{8}(\underbrace{3^4}_{E} + \underbrace{2\cdot 3^1}_{C_4} + \underbrace{3^2}_{C_2} + \underbrace{2\cdot 3^2}_{\sigma_v} + \underbrace{2\cdot 3^3}_{\sigma_d}) = 21 \tag{12}$$

Note that $|Cyc_D(g)|$ is the same for all operators belonging to the same conjugacy class. Moreover, the identity always bears $|D|$ unitary cycles, thus providing the largest contribution to the number of SICs. This observation plays a significant role in the proof of the Oberschelp's formula [30, 31], that states

$$|\Delta(S)| = \frac{|R|^{|D|}}{|G|}(1 + o(1)), \quad o(1) \to 0 \ \text{ when } \ |D| \to \infty \tag{13}$$

This formula, obtained for unlabeled graphs, can be extended to the case of large unit cells, in which $|D| >> 1$. It will be at the base of the scaling analysis of our approach, presented in Section 2.5.

2.2. Pólya's theory at fixed composition

Pólya's theory allows to count SICs for a given composition. This is achieved by viewing the colors as variables $z_1, z_2, .., z_{|R|}$ and considering the weight function $W(s)$ defined by

$$W(s) = \prod_{1 \leq j \leq |R|} z_j^{n_j} \tag{14}$$

where $n_j = |s^{-1}(z_j)|$ is the number of elements of D mapped on color z_j. Note that W is constant at fixed composition. For example, in model A with 2 colors, all configurations on orbits 3 and 4 show 2 b and 2 r sites (see Figure 2), thus share the same W value: b^2r^2.

Summing on both sides of Eq. (14) over the elements stabilized by a given g, and using the fact that $S_g \cong R^{Cyc_D(g)}$, we obtain

$$\sum_{s \in S_g} W(s) = \sum_{t \in R^{Cyc_D(g)}} \prod_{c \in Cyc_D(g)} t(c)^{|c|} \tag{15}$$

where $t(c)$ is the variable corresponding to the color taken by t on the cycle $c \in Cyc_D(g)$. Eq. (15) can be rewritten as

$$\sum_{s \in S_g} W(s) = \prod_{c \in Cyc_D(g)} \sum_{j=1}^{|R|} z_j^{|c|} \tag{16}$$

To establish (16) one observes that expanding the product in the r.h.s. of (16) gives a sum of terms of the form $z_{i_1}^{|c_1|} z_{i_2}^{|c_2|} ... z_{i_N}^{|c_N|}$ (where $N = |Cyc_D(g)|$) which is precisely the weight of a mapping $t \in R^{Cyc_D(g)}$ that takes the color corresponding to the variable z_{i_1} on the cycle c_1, the color corresponding to the variable z_{i_2} on the cycle c_2 and so on. This sum contains exactly the same number of terms as the sum in the r.h.s. of (15) (i.e. $|R|^N$ terms) and each one of its terms corresponds to a unique term in the r.h.s of (15).

Using Eqs. (7) and (16) one obtains:

$$\sum_{s' \in \Delta(S)} W(s') = \frac{1}{|G|} \sum_{g \in G} \prod_{c \in Cyc_D(g)} \sum_{j=1}^{|R|} z_j^{|c|} \tag{17}$$

Let us denote the r.h.s. of this identity as $PP_{|R|}^{|D|}(z_1, .., z_{|R|})$. Expanding this polynomial

$$PP_{|R|}^{|D|}(z_1, .., z_{|R|}) = \sum_{n_1 + ... + n_{|R|} = |D|} k_{(n_1, .., n_{|R|})} z_1^{n_1} .. z_{|R|}^{n_{|R|}} \tag{18}$$

yields the desired information about the number of the orbits with a given composition. The $k_{(n_1, .., n_{|R|})}$ coefficient associated with the monomial $z_1^{n_1} .. z_{|R|}^{n_{|R|}}$ gives the number of orbits corresponding to the composition $(n_1, \ldots, n_{|R|})$.

In order to compute the number of SICs (orbits) at fixed composition, the "type" of each $g \in G$ is required. It is the $|D|$-plet $\mathcal{T}_D(g) = (l_1, l_2, .., l_{|D|})$, whose i^{th} value l_i indicates the number of cycles of length i (obviously, $\sum_i l_i \cdot i = |D|$). Note that the type is the same for all operators belonging to the same conjugacy class.
In Tables 1 and 2, the type of the operators is given for models A and B, respectively. As an example of polynomial, for model A (4 sites) and 2 colors we obtain:

$$PP_2^4(b, r) = b^4 \ + \ b^3 r \ + \ 2 \cdot b^2 r^2 \ + \ b r^3 \ + \ r^4 \tag{19}$$

This last equation indicates that each end-member b^4 or r^4 corresponds to 1 SIC, as well as compositions $b^3 r$, $b r^3$, while $b^2 r^2$ is split on 2 SICs. Analogously, for model A with 3 colors:

$$\begin{aligned} PP_3^4(b, r, g) = {} & b^4 \ + \ b^3 r \ + \ b^3 g \ + \ 2 \cdot b^2 r^2 \ + \ 2 \cdot b^2 r g \\ & + \ 2 \cdot b^2 g^2 \ + \ b r^3 \ + \ 2 \cdot b r^2 g \ + \ 2 \cdot b r g^2 \ + \ b g^3 \\ & + \ r^4 \ + \ r^3 g \ + \ 2 \cdot r^2 g^2 \ + \ r g^3 \ + \ g^4 \end{aligned} \tag{20}$$

2.3. De Bruijn's generalization: spin counting

Sections 2.1 and 2.2 dealt with a group G acting on a set D, and with the induced action of this group on the set of configurations $S = R^D$ (i.e. the Pólya's action). A more general action can be introduced when a second group H acts on the set of colors R. We shall indicate it as De Bruijn's action [32].
Following De Bruijn we shall say that two configurations $s_1, s_2 \in S$ are equivalent if there exist elements $g \in G$ and $h \in H$ such that

$$s_1(g \cdot x) = h \cdot s_2(x) \ ; \quad x \in D \tag{21}$$

This amounts to consider the direct product $G \times H$, consisting of all products $g \times h$, with $g \in G$, $h \in H$, and to see it as acting on S via

$$(g \times h) \cdot s(x) = h \cdot s(g^{-1} \cdot x) \ ; \quad x \in D \tag{22}$$

The results of Section 2.1 permit to assert in this case

$$|\Delta(S)| = \frac{1}{|G \times H|} \sum_{g \times h \in G \times H} |S_{g \times h}| \tag{23}$$

where $\Delta(S)$ denotes, as in Section 2.1, a set of representatives of the action (22).
Pólya's formula (10) was derived exploiting the fact that $S_g \cong R^{Cyc_D(g)}$. What we need here is to find a similar characterization which permits to find the number of configurations s that satisfy

$$s(g \cdot x) = h \cdot s(x) \ ; \quad x \in D \tag{24}$$

De Bruijn [33] succeeded in characterizing these configurations in terms of the cycle structure of g and h. Let us summarize his argument.
Assume for the types that $\mathcal{T}_D(g) = (l_1, l_2, .., l_{|D|})$ and that $\mathcal{T}_R(h) = (m_1, m_2, .., m_{|R|})$. Let s be a configuration that satisfies (24). Let x denote some element of D. Assume

that this element belongs to a cycle of g of length i. This cycle can be described through the elements

$$x,\ g\cdot x,\ g^2\cdot x,\ \cdots,\ g^{i-1}\cdot x \tag{25}$$

The crucial observation is that (24) implies that s must map the elements (25) on:

$$s(x),\ h\cdot s(x),\ h^2\cdot s(x),\ \cdots,\ h^{i-1}\cdot s(x) \tag{26}$$

and the following condition should be satisfied:

$$h^i\cdot s(x) = s(g^i\cdot x) = s(x) \tag{27}$$

This means that the length of the cycle of h to which $s(x)$ belongs should divide i.
Using this observation, we can easily compute the number of possibilities we have for $s\in S_{h\times g}$ (i.e. the analog of the quantity $|R|^{|Cyc_D(g)|}$ of Pólya's formula (10)). For each cycle of g the number of possibilities for the element of R on which the element x of (25) can be mapped is

$$\sum_{j|i} j\cdot m_j \tag{28}$$

where i is the length of the cycle (25) and where the m_j's are determined by the type of h; $j|i$ refers to the j divisors of i.
Since there are l_i cycles of length i in the decomposition of g we obtain

$$|S_{g\times h}| = \prod_i (\sum_{j|i} j\cdot m_j)^{l_i} \tag{29}$$

Combining (23) and (29) we deduce

$$|\Delta(S)| = \frac{1}{|G|\cdot|H|}\sum_{g\in G}\sum_{h\in H}\prod_i (\sum_{j|i} j\cdot m_j)^{l_i} \tag{30}$$

where (as indicated above) (l_1,l_2,..,$l_{|D|}$) is the type of g and (m_1,m_2,..,$m_{|R|}$) is the type of h.
Let us now focus on the case where R reduces to two elements $R=\{\uparrow,\downarrow\}$ and H to the group with two operators, namely identity (E_R) and exchange (X_R) of $\uparrow$ and $\downarrow$; the cycle structure of E_R and X_R is $(\uparrow)$, $(\downarrow)$ and $(\uparrow,\downarrow)$), respectively. Then, Eq. (30) reduces to

$$|\Delta(S)| = \frac{1}{2|G|}\sum_{g\in G}\left(\prod_i 2^{l_i} + \chi(g)\prod_{i|2} 2^{l_i}\right) \tag{31}$$

where $\chi(g)=0$ if g contains a cycle with an odd length and $\chi(g)=1$ if this is not the case.
Introducing the notation

$$G^e = \{g\in G, \chi(g)=1\} \tag{32}$$

we can rewrite (31) as

$$|\Delta(S)| = \frac{1}{2|G|}\sum_{g\in G} 2^{|Cyc_D(g)|} + \frac{1}{2|G|}\sum_{g\in G^e} 2^{|Cyc_D(g)|} \tag{33}$$

The first term of the sum is equivalent to the Pólya's counting formula (up to the factor $\frac{1}{2}$). The second term gives a zero contribution if every operation contains odd cycles. Counting orbits with a given spin composition is not as easy as for the classical Pólya's action. This is discussed by de Bruijn [33] and by Harary and Palmer [34]. In order to overcome this difficulty, we propose to proceed as follows. The first step consists in constructing the analog of the polynomials (18) attached to the group G acting individually on R^D. This gives a polynomial expression in $\uparrow$ and $\downarrow$ of the form

$$PP^{|D|}_{\uparrow\downarrow}(\uparrow,\downarrow) = \sum_{n_1+n_2=|D|} k_{(n_1,n_2)} \uparrow^{n_1}\downarrow^{n_2} \tag{34}$$

where the coefficients $k_{(n_1,n_2)}$ satisfy

$$k_{(n_1,n_2)} = k_{(n_2,n_1)} \ , \ \sum_{n_1+n_2=|D|} k_{(n_1,n_2)} = \frac{1}{|G|}\sum_{g\in G} 2^{|Cyc_D(g)|} \tag{35}$$

It is easy to see that the number of orbits corresponding to composition ($n_1 \uparrow$,$n_2 \downarrow$) with $n_1 \neq n_2$ is directly given by $k_{(n_1,n_2)}$. When $|D|$ is odd, these compositions are the only possible and the polynomial (34) encodes all the required information about the spin composition. One should note that in this case $G^e = \emptyset$ and formula (33) reduces to the Pólya term. When $|D|$ is even the second term in (33) contributes in a subtle way. The right counting is obtained by combining (10), (33) and (35):

$$k_{(n\uparrow,n\downarrow)} = \frac{1}{2|G|}\sum_{g\in G^e} 2^{|Cyc_D(g)|} + \frac{k_{(n,n)}}{2} \tag{36}$$

For models A and B (Tables 1 and 2), Formula (33) yields 4 and 27 spin SICs, respectively:

$$N^4_{\uparrow\downarrow} = \frac{1}{8}(\underbrace{2\cdot 2^1}_{C_4}+\underbrace{2^2}_{C_2}+\underbrace{2\cdot 2^2}_{\sigma_v}) + \frac{1}{16}(\underbrace{2^4}_{E}+\underbrace{2\cdot 2^3}_{\sigma_d}) = 4 \tag{37}$$

$$N^8_{\uparrow\downarrow} = \frac{1}{8}(\underbrace{2\cdot 2^2}_{C_4}+\underbrace{2^4}_{C_2}+\underbrace{2\cdot 2^4}_{\sigma_v}+\underbrace{2\cdot 2^4}_{\sigma_d}) + \frac{1}{16}(\underbrace{2^8}_{E}) = 27 \tag{38}$$

The 4 spin representatives of model A correspond to the first 4 representatives in Figure 2, provided that b is substituted by $\uparrow$ and r by $\downarrow$. Only compositions in the range 0 - 50 % have representatives; the reason is that $(0\uparrow, 4\downarrow)$ and $(1\uparrow, 3\downarrow)$ are equivalent to $(4\uparrow, 0\downarrow)$ and $(3\uparrow, 1\downarrow)$, respectively, due to the spin exchange symmetry. As regards the 50 % composition $(2\uparrow, 2\downarrow)$, there are 2 spin SICs, as in the case of the two colors b and r. However, the last statement is not true in general, as we will discuss below in the case of model B.

In the cases of model A and B (Tables 1 and 2), formula (35) gives, respectively:

$$k_{(2\uparrow,2\downarrow)} = \frac{1}{16}(\underbrace{2\cdot 2^1}_{C_4}+\underbrace{2^2}_{C_2}+\underbrace{2\cdot 2^2}_{\sigma_v}) + \frac{2}{2} = 2 \tag{39}$$

$$k_{(4\uparrow,4\downarrow)} = \frac{1}{16}(\underbrace{2\cdot 2^2}_{C_4}+\underbrace{2^4}_{C_2}+\underbrace{2\cdot 2^4}_{\sigma_v}+\underbrace{2\cdot 2^4}_{\sigma_d}) + \frac{13}{2} = 12 \tag{40}$$

Note that in the case of model A (4 sites), the number of spin SICs at 50 % composition is the same than in the two colors (b, r) case. On the contrary, for the 50% model B (8 sites) composition, the spin case yields 12 spin SICs, to be compared with 13 SICs for the two colors (b, r) case. The reason is that there are two SICs of the latter case that become symmetry equivalent in the former, due to the additional exchange operator X_R acting on the two spin states $\uparrow$ and $\downarrow$. The full sets of configurations corresponding to these two SICs are illustrated in Figure 3. The two sets are symmetry independent in the b, r-colors case, but become symmetry equivalent under the action of the spin exchange operator X_R on the set of colors; couple of configurations related by this operator are shown in the same row in the Figure.

2.4. Algorithmic aspects: lexicographic ordering and surjective resolution

In this Section, we are concerned with the difficult problem of finding the complete set of SICs. Direct methods require lists of independent configurations together with their equivalent configurations that must be stored for subsequent use. To decide whether a new configuration is independent or not from those already produced (isomorphism test), the complete list must be spanned.

Orderly generation methods [35, 28, 29] provide practical algorithms that do not require long lists, perform efficiently with more than two colors and reduce drastically the cost of isomorphism tests. They are based on the fact that orders on D and R induce a canonical order on the set R^D: the *lexicographic order* (see Figure 4). Providing the set of configurations with this canonical order permits to form a system of canonical representatives by taking the smallest element in each orbit. Within this framework, the generalization of de Bruijn can be easily implemented, thanks to the direct product group structure of the symmetry operation involved in the de Bruijn's action (22).

In order to explain the implementation of the orderly generation, we consider the simple case of a set D mapped on two colors $|R| = 2$ represented by 0 and 1. Each configuration can be represented as a 0-1 sequences of length $|D|$. For convenience, this sequences can be identified with a $|D|$-number in base 2, but this identification plays no role in the enumeration of the SICs. Starting from the "first" configuration $\ell_1 = (0 \cdots 0)$, one produces the sequence of successive configurations by increasing the $|D|$-digits number by 1 at each step. Application to model A is illustrated in Figure 4, where the various configurations are labelled to the left from ℓ_1 to ℓ_{16}. Increasing proceeds from right to left; a configuration of the list is higher than another when the corresponding $|D|$-digits number is larger. On this basis, canonical representatives of the orbits are efficiently selected. Being the first of the list, ℓ_1 is obviously a representative. ℓ_2 is compositionally different from ℓ_1, so no operator transforms it into ℓ_1. It does not belong to the orbit of ℓ_1 and is the smallest element of a new class of configurations. As such it is stored as the second representative. In the case of model A, configurations ℓ_3 and ℓ_2 have the same composition, so they could be equivalent by symmetry, that is it may exist an operator transforming ℓ_3 into ℓ_2. This is the case since the clockwise 4-fold rotation transformed

ℓ_3 into ℓ_2. Then ℓ_3 is symmetry equivalent to ℓ_2 and discarded. ℓ_4 is then considered. Applying every symmetry operator it is never transformed into a configuration equal or smaller than ℓ_3. It is then recorded as a representative. It is easy to see that if there exists $g \in G$ such that $g \cdot \ell_4$ is lexicographically smaller than or equal to ℓ_2 then either $\ell_4 \in \Omega(\ell_1)$ or $\ell_4 \in \Omega(\ell_2)$. On the basis of this remark, there is no need to hold the complete list of the elements of orbits already classified to decide if a new configuration ℓ_n belongs to one of these orbits. If there exists $g \in G$ transforming ℓ_n into a configuration smaller or equal to the last found canonical representative, then it belongs to one of these orbits and is not the canonical representative of a new orbit. The process continues along the same lines for the following members of the list and yields the next canonical representatives: ℓ_6, ℓ_8 and ℓ_{16} (Figure 4).

Such an implementation does not require long lists and the canonicity test is reduced to the comparison of G-equivalent configurations of the one under consideration with the last identified canonical representative. These features considerably speed up the selection with respect to direct schemes (it should be noticed that the cost of the selection of SICs, also when the direct strategy is adopted, may correspond in many cases to a small fraction of the overall cost of the calculation [21, 22, 23]). The possibility of reducing the number of canonicity tests using an augmentation procedure, as proposed by Read [35], has not been implemented here because it would require to hold a sub-list of canonical representatives. Furthermore, Goldberg [36] noted that in such case there is no efficient method to determine whether a configuration is canonical or not and no mechanism ensures that the algorithm does not consider an exponentially long list of unsuccessful augmentations.

Orderly generation of configurations applies equally for more than two colors and could provide the complete list of SICs. However, it can be improved by combining it with a recursion procedure. Before introducing this procedure, we show how the previous considerations allow to identify the SICs corresponding to a fixed composition. This is useful for situations where only one composition is of interest, for example in the study of inverse spinels or disordered systems in general.

All we need, starting from the configuration having the required composition and the lowest lexicographic rank, is to generate the lexicographically ordered list of $|D|$ sequences corresponding to the composition and test the canonicity of each new generated configuration as previously described. These $|D|$-sequences can be interpreted either as anagrams or as $|D|$-digit numbers in base $|R|$. If one converts configurations ($|D|$-sequences, or $|D|$-digit numbers in base $|R|$) into base-10 integers, one obtains a list of increasing but not consecutive integers. In the computer science language, this corresponds to a hashing scheme using a perfect, but not minimal hash table. In contrast, Hart *et al.* [37], looking for derivative structures at fixed composition, proposed an approach demanding minimality of the hash table. To index the configurations in minimal mode, they cleverly introduced a mixed-radix number. Minimality is not taken as a condition in the approach here presented. However an improvement is offered by the use of the Pólya's polynomial, whose coefficients provide a stopping condition. As

soon as the number of classes is found, the search is interrupted. Figure 5 illustrates the scheme in the case of model A with 3 colors.

In the general case where more than two colors are considered, we combine lexicographic ordering with a recursion procedure: the so-called *surjective resolution*. In order to explain this procedure let us assume that the group G acts on two sets of configurations $S = R^D$ and $S_1 = R_1^D$, where R_1 is obtained by adding a new color $z_{|R|+1}$ to the set R:

$$R_1 = R \cup \{z_{|R|+1}\} = \{z_1, \ldots, z_{|R|}, z_{|R|+1}\}. \tag{41}$$

A natural mapping Θ can be defined from R_1^D onto R^D

$$\begin{aligned} &\Theta : R_1^D \longrightarrow R^D \\ &\quad s \longrightarrow \Theta s \end{aligned} \tag{42}$$

by setting

$$\begin{cases} (\Theta s)_i = s_i & \text{if } s_i \neq z_{|R|+1} \\ (\Theta s)_i = z_{|R|} & \text{if } s_i = z_{|R|+1} \end{cases} \tag{43}$$

This mapping is surjective and "compatible" with the actions of G on R^D and R_1^D. This means that

$$\Theta(g \cdot s) = g \cdot \Theta(s) \ ; \ s \in R_1^D, g \in G. \tag{44}$$

In particular the orbit of an element $s \in R_1^D$ projects on the orbit of $\Theta(s)$ in R^D.

For model A and three colors ($R_1 = \{r, b, g\}$, where g stands for "green"), the projection Θ from R_1^D onto R^D corresponds to substitute r for g in each configuration s containing green color. The fact that every orbit of the action of $G = C_{4v}$ on $\{g, r, b\}^{\{1,2,3,4\}}$ is projected on the orbit of the projection of one of its elements is illustrated in Figure 6.

The surjective resolution principle asserts that it is possible to construct a system of representatives of the action of G on R_1^D from a set of representatives of the action of G on R^D and their stabilizers in G.

More precisely let Ω_1 denote an orbit of the action of G on R_1^D and let s_1 an element of this orbit. Let ω denote the canonical representative of the orbit $\Omega(\Theta(s_1))$ and let $g \in G$ be the operator defined by $\omega = g \cdot \Theta(s_1)$.

The compatibility property (44) implies that $\omega = \Theta(g \cdot s_1)$, which means that $g \cdot s_1 \in \Theta^{-1}(\omega)$. In other words the orbit Ω_1 intersects a set of form $\Theta^{-1}(\omega)$ for an ω belonging to the set of canonical representatives of the action of G on R^D. Figure 6 shows for example that orbit 13 intersects the set $\Theta^{-1}(10)$ and orbit 19 intersects the set $\Theta^{-1}(16)$.

It is easy to see that such a canonical representative is unique: if we assume that ω' is another canonical representative such that $\Omega_1 \cap \Theta^{-1}(\omega') \neq \varnothing$, then there exists $s_1' \in \Omega_1$ such that:

$$\omega' = \Theta(s_1') = \Theta(g' \cdot s_1) = g' \cdot \Theta(s_1) = g' \cdot g^{-1} \cdot \omega \tag{45}$$

for a certain $g' \in G$, which means that ω and ω' are on the same orbit.

It follows that each orbit of the action of G on R_1^D intersects one and only one preimage $\Theta^{-1}(\omega)$ of the set of canonical representatives of the action of G on R^D. On Figure 6,

these representatives are indicated by an asterisk.
The second important fact is that for each canonical representative ω, orbits of the action of G on $\Theta^{-1}(\omega)$ are exactly the orbits of the action of the stabilizer G_ω on $\Theta^{-1}(\omega)$. To see this, consider two elements $s, s' \in \Theta^{-1}(\omega)$ which are on the same orbit, i.e. $s' = g \cdot s$, for some $g \in G$, and note that

$$\omega = \Theta(s') = \Theta(g \cdot s) = g \cdot \Theta(s) = g \cdot \omega \tag{46}$$

which shows that the operator g is an element of the stabilizer G_ω. In model A, the 3-colors orbits are derived from 2-colors ones. As an example, consider class (orbit) 7 and its canonical representative. Classes 8 and 9 are obtained using the stabilizer of 7: C_{2v}.
The upshot is that once we have produced a set of canonical representatives ω of the action of G on R^D, it is enough to compute the preimages $\Theta^{-1}(\omega)$ and then the set of canonical representatives of the orbits of the action of the stabilizers G_ω on $\Theta^{-1}(\omega)$. Proceeding in this way permits to reduce the number of isomorphisms tests, because canonicity tests are performed within shorter lists and under the action of smaller groups. Figures 6 and 7 can help to illustrate this point, again with reference to model A. The latter shows to the left the generation of the representatives for $|R|$=2; to the right, on the contrary, the generation of the representatives for $|R|$=3 is shown, where the branching for the third color starts from the $|R|$=2 representative only. The more explicit generation of orbits with three colors, blue, red and green is shown in Figure 6. From Figures 6 and 7, one can note that after the representative of the 15^{th} class has been found, there is no need to go on. More precisely, the process is stopped when all the descendants of the last but one 2-colors only representative have been found. As shown by the presented example, we would then be looking for configurations built on 2 colors, labelled 1 and 2, instead of 0 and 1. The labeling of colors being irrelevant, the next representatives can be obtained from previous one by properly re-labeling the colors.

2.5. Orderly generation: scaling with number of SICs

In this Section we discuss time scaling aspects of the orderly generation algorithm. We start our analysis of the lexicographic ordering scheme for the 2-colors case, as described in Section 2.4. The full list of configurations $\{\ell_i\}$ is explored (providing a factor $2^{|D|}$); to each of them at most $|G|$ symmetry operators are applied to test the canonicity. The required total time (T_{LO}) is then:

$$T_{LO} \approx 2^{|D|} \cdot |G| \tag{47}$$

In contrast, we can estimate the scaling behavior of the direct method (see the flow charts reported in Figure 2 of Ref. [20]) for the 2-colors case. This method requires to construct the full list of the $2^{|D|}$ configurations. Within this list the first occurrence of a new orbit is determined by comparing each element of the list with all the elements of the already identified symmetry classes of configurations stored in a second vector. The

process is initiated by taking the first element of the full list as the representative of the first class and applying to it all the $|G|$ operators; the obtained configurations belonging to its class are stored in the second vector. Every time an element of the long list is not in the second vector, then it is taken as the representative of a new class and all its equivalent are added to the second vector. All the symmetry classes are found when the length of the two vectors are equal. This approach scales as $2^{2|D|}$. The application of the symmetry operators adds a contribution proportional to $|\Delta(S)| \cdot |G|$ that is negligible with respect to $2^{2|D|}$. So the required time is

$$T_{dir} \approx 2^{2|D|} \tag{48}$$

Relation (47) gives

$$T_{dir} \approx T_{LO} \cdot \frac{2^{|D|}}{|G|} \tag{49}$$

from which we deduce, by using the Oberschelp's formula (13):

$$\frac{T_{dir}}{T_{LO}} \approx |\Delta(S)| \tag{50}$$

In order to illustrate the relations (47) and (49), we considered the Mg sites in the tetragonal MgO (n,1,1) supercells built from the primitive cell, with n ranging from 12 to 32 and $|R| = 2$. The number of sites $|D|$ equals n and $|G|$ ranges from 48 to 128 (in steps of 4). The results are plotted in Figure 8. The CPU time needed by the algorithm to generate all the representatives of the SICs is linear on $|R|^{|D|}$. The quadratic behavior of the direct method on $|R|^{|D|}$ with $|R| = 2$ is also shown for comparison. For large cells ($|D|$ large), the number of SICs agrees with the relation (13) and supports the approximations used previously.

Let us now discuss the case of $|R|$ colors, with $|R| > 2$, which is handled by means of the surjective resolution. The 3-colors case will be considered, the obtained results being applicable to larger numbers of colors. Let N and $\ell_1, \ell_2, \ldots$ be the number and the representatives of the 2-colors SICs, respectively. From the definition and properties of the Θ mapping (42)- (43), the sum of the lengths of the preimages of all the ℓ_j is the number of the 3-colors configurations $3^{|D|}$:

$$\sum_{j=1}^{N} \left|\Theta^{-1}(\ell_j)\right| = 3^{|D|} \tag{51}$$

For large $|D|$, and obviously large N (approximated by the Oberschelp's formula), the mean length of the preimages of the ℓ_j is given by

$$< \left|\Theta^{-1}(\ell_j)\right| > \frac{2^{|D|}}{|G|} \approx 3^{|D|} \tag{52}$$

$$< \left|\Theta^{-1}(\ell_j)\right| > \approx |G| \left(\frac{3}{2}\right)^{|D|} \tag{53}$$

For each $\ell_j, j = 1, \ldots, N$, the time needed to explore the set $\Theta^{-1}(\ell_j)$ is in the order of $|\Theta^{-1}(\ell_j)| \cdot \left|G_{\ell_j}\right|$
The total time to explore the 3-colors SICs is then

$$T_{surj} = \sum_{j=1}^{N} \left|\Theta^{-1}(\ell_j)\right| \cdot \left|G_{\ell_j}\right| \approx< \left|\Theta^{-1}(\ell_j)\right| > \sum_{j=1}^{N} \left|G_{\ell_j}\right| \tag{54}$$

As shown by Goldberg (see Lemma 1 in Ref. [36]), an important consequence of Oberschelp's formula is

$$\sum_{j=1}^{N} \left|G_{\ell_j}\right| = (1 + o(1))N \approx N \approx \frac{2^{|D|}}{|G|} \tag{55}$$

Combining Eqs. (53), (54) and (55)

$$T_{surj} \approx |G| \left(\frac{3}{2}\right)^{|D|} \times \frac{2^{|D|}}{|G|} = 3^{|D|} \tag{56}$$

In the case of $|R|$ colors, the previous formula is written

$$T_{surj} \approx |R|^{|D|} \tag{57}$$

This result is well illustrated in Figure 9, where the CPU time needed to produce the representatives of SICs with 3 colors for MgO supercell of type (n,1,1), with $7 \leq n \leq 22$, appears to be linearly dependent on $3^{|D|}$. With respect to lexicographic approach, the surjective resolution is $|G|$ times more efficient.
This rather favorable scaling behavior permits to explore a relatively high number of colors. As an example, up to 5 and 6 colors are considered in Figure 10, with two different values for $|D|$, 12 and 16. Note that the CPU time depends on both $|R|$ and $|D|$, as expected from Eq. (57).

3. Examples from geochemistry: garnets and olivines

In this Section, we apply Pólya's and De Bruijn's theories introduced in the previous Sections to two real systems, namely garnets and olivines.
Garnets are orthosilicates with general chemical formula $X_3Y_2Si_3O_{12}$, where X^{2+} and Y^{3+} are divalent and trivalent cations, respectively. The primitive cell contains four formula units, for a total of 80 atoms; the space group G is cubic ($Ia\bar{3}d$) with 48 symmetry operators. Natural garnets form substitutional solid solutions extending over a broad chemical range, and involving up to 12 end members[38, 39]. The most common cases refer to substitutions of either trivalent cations at the Y octahedral site (1 orbit containing 8 symmetry equivalent sites in the primitive cell) or divalent cations at the X dodecahedral site (1 orbit with 12 equivalent sites).
In Table 3 we reported the analysis of the action of G on the set $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ of octahedral sites in the primitive cell. To make the Table more compact, symmetry operators g were grouped in conjugacy classes CC (10 in total); for each class CC, the number of cycles $|Cyc_D(CC)|$ and the type $\mathcal{T}_D(CC)$ are shown.

Various trivalent cations can occur in the garnet octahedral site, such as Fe^{3+} (iron), Al^{3+} (aluminium), Cr^{3+} (chromium). For the corresponding solid solutions, Pólya's formulas (10) and (17)-(18) can then be used to compute the total number of SICs and the number of SICs at fixed composition. For example, in the case of 2-colors binary systems, we have:

$$N_2^8 = \frac{1}{48}(\underbrace{2^8}_{E} + \underbrace{3\cdot 2^4}_{C_2} + \underbrace{6\cdot 2^4}_{C_2'} + \underbrace{8\cdot 2^4}_{C_3} + \underbrace{6\cdot 2^2}_{C_4} + \underbrace{2^8}_{i} + \underbrace{6\cdot 2^2}_{S_4} + \underbrace{8\cdot 2^4}_{S_6} + \underbrace{3\cdot 2^4}_{\sigma_h} + \underbrace{6\cdot 2^4}_{\sigma_d}) = 23 \tag{58}$$

$$PP_2^8(b,r) = b^8 + b^7 r + 3\cdot b^6 r^2 + 3\cdot b^5 r^3 + 7\cdot b^4 r^4 + 3\cdot b^3 r^5 + 3\cdot b^2 r^6 + b r^7 + r^8 \tag{59}$$

In the case of magnetic trivalent cations, such as Fe^{3+}, the system is usually a magnetic "solid solution", involving Fe(↑) and Fe(↓) species. Formulas (33) and (36), from De Bruijn's approach, become then useful to calculate the number of spin SICs:

$$N_{\uparrow,\downarrow}^8 = \frac{1}{48}(\underbrace{3\cdot 2^4}_{C_2} + \underbrace{6\cdot 2^4}_{C_2'} + \underbrace{6\cdot 2^2}_{C_4} + \underbrace{6\cdot 2^2}_{S_4} + \underbrace{3\cdot 2^4}_{\sigma_h} + \underbrace{6\cdot 2^4}_{\sigma_d}) + \frac{1}{96}(\underbrace{2^8}_{E} + \underbrace{2^8}_{i} + \underbrace{8\cdot 2^4}_{C_3} + \underbrace{8\cdot 2^4}_{S_6}) = 15 \tag{60}$$

$$k_{(4\uparrow,4\downarrow)} = \frac{1}{96}(\underbrace{3\cdot 2^4}_{C_2} + \underbrace{6\cdot 2^4}_{C_2'} + \underbrace{6\cdot 2^2}_{C_4} + \underbrace{6\cdot 2^2}_{S_4} + \underbrace{3\cdot 2^4}_{\sigma_h} + \underbrace{6\cdot 2^4}_{\sigma_d}) + \frac{7}{2} = 7 \tag{61}$$

note that, in the case of garnets, the number of spin SICs at 50 % composition is the same than the number of "chemical" SICs: compare $k_{(4\uparrow,4\downarrow)}$ with the coefficient of the b^4r^4 term in polynomial (59).

To explore a bigger set of SICs, it is necessary to take a larger set of sites D, by using a supercell of the primitive cell as a reference. As an example, we performed the Pólya's analysis in the case of the garnet conventional cell (160 atoms instead of 80), for both octahedral (1 orbit containing 16 equivalent sites in the conventional cell) and dodecahedral sites (1 orbit with 24 equivalent sites). The sets of $|Cyc_D(CC)|$ and the total numbers of SICs $N_{|R|}^D$ are reported in Tables 4 and 5, respectively.

Note that, when building a supercell, the symmetry group of the system must be enlarged: it is obtained as product group of the space group and of the translational vectors used to build up the supercell from the primitive one. The garnet conventional cell is double than the primitive one, so that there are 96 operators instead of 48, grouped in 20 (instead of 10) conjugacy classes. The additional classes result from the composition of each class of the primitive case with the centering vector $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ of the conventional cell. The $|Cyc_D(CC)|$ values for the conventional cell were reported in Table 4 in a compact form (see caption to Table).

Now, let us comment on the number of SICs (Table 5), which was also studied as a function of the number of colors $|R|$. Taking the 2-colors case and starting from the

primitive cell, when going from the octahedral ($|D| = 8$) to the dodecahedral ($|D| = 12$) sites, the number of configurations goes from 23 to 154; This number increases up to 179'444 for the 24 dodecahedral sites in the conventional cell. The number of SICs increases enormously with the number of colors: yet for 3 colors it can reach the order of billions (dodecahedral sites in conventional cell). Note that these numbers can be obtained with a very limited number of operations with the present scheme.
The second mineralogical example refers to olivines. They are orthosilicates, too, with chemical formula X_2SiO_4 (X^{2+} are divalent cations). Their primitive cell contains four formula units and 28 atoms in total; The space group G is orthorhombic ($Pbnm$) with 8 symmetry operators. The binary system Mg_2SiO_4-Fe_2SiO_4 is very common in nature [38, 40]. The X octahedral site, involved in this solid solution, shows 2 orbits in the primitive cell, each of them containing 4 equivalent sites (8 sites in total).
The action of G on the set of octahedral sites $D = \{1, 2, 3, 4, 5, 6, 7, 8\}$ is analyzed in Table 6. Sites 1-4 and 5-8 belong to the 2 separate orbits; this implies that in every cycle decomposition they are always found in different cycles. Similar to the case of garnets, we can apply Pólya's theory to get the number of SICs of the binary system:

$$N_2^8 = \frac{1}{8}(\underbrace{2^8}_{E} + \underbrace{3 \cdot 2^4}_{C_2} + \underbrace{2^6}_{i} + \underbrace{2^6}_{\sigma_h} + \underbrace{2 \cdot 2^4}_{\sigma_v}) = 58 \tag{62}$$

$$\begin{aligned} PP_2^8(b, r) = {} & b^8 \ + \ 2 \cdot b^7 r \ + \ 8 \cdot b^6 r^2 \ + \ 10 \cdot b^5 r^3 \ + \ 16 \cdot b^4 r^4 \\ & + \ 10 \cdot b^3 r^5 \ + \ 8 \cdot b^2 r^6 \ + \ 2 \cdot b r^7 \ + \ r^8 \end{aligned} \tag{63}$$

The iron end member Fe_2SiO_4 has magnetic Fe^{2+} cations in the octahedral sites, which result in the occurrence of magnetic solid solutions. A De Bruijn's analysis yields

$$N_{\uparrow,\downarrow}^8 = \frac{1}{8}(\underbrace{3 \cdot 2^4}_{C_2} + \underbrace{2 \cdot 2^4}_{\sigma_v}) + \frac{1}{16}(\underbrace{2^8}_{E} + \underbrace{2^6}_{i} + \underbrace{2^6}_{\sigma_h}) = 34 \tag{64}$$

$$k_{(4\uparrow,4\downarrow)} = \frac{1}{16}(\underbrace{3 \cdot 2^4}_{C_2} + \underbrace{2 \cdot 2^4}_{\sigma_v}) + \frac{16}{2} = 13 \tag{65}$$

note that, in the case of 50 % spin composition there are 13 spin SICs, against the 16 SICs of the "chemical" case.

4. Conclusions

In the present study it has been shown that the problem of the automatic and efficient identification of the symmetry independent configurations can be successfully solved for (formally) any number of involved positions and species (colors) by using Pólya's and de Bruijn's formalisms and, for the explicit generation of SICs representatives, an orderly generation approach based on lexicographic ordering combined with the surjective resolution principle.
The proposed algorithm (that presents many evident advantages with respect to the direct scheme [21, 22, 23]), has been implemented in a development version of the

CRYSTAL code and represents a contribution to the automatic investigation of solid solutions, nowadays (and even more in the near future) at hand, as high performance computing provides thousands of processors whose use imposes to minimize the number of manual operations at the various stages of the calculation.
Despite common features with a previously proposed method [18, 19, 37], the present approach offers a new lightening on the relevant problem of enumerating structures, and is prone to further developments which will be part of future work. Just as an example, we mention the problem of the automatic identification of the minimal cell that provides the required information (that is that contains all the required two-body interactions) with the constraint of the maximum symmetry.

Acknowledgments

The authors are very grateful to the anonymous reviewers for their valuable suggestions to improve the readibility of the text and the comparison with other published works, as well as for pointing us some pertinent references. They also acknowledge Compagnia di San Paolo for financial support (Progetti di Ricerca di Ateneo-Compagnia di San Paolo-2011-Linea 1A, progetto ORTO11RRT5).

g	Moves				Cycles	$\lvert Cyc_4(g)\rvert$	$\mathcal{T}_4(g)$	$pp_2^4(g)$	$pp_3^4(g)$
E	1	2	3	4	(1)(2)(3)(4)	4	(4,0,0,0)	$(b+r)^4$	$(b+r+g)^4$
C_4	2	3	4	1	(1234)	1	(0,0,0,1)	(b^4+r^4)	$(b^4+r^4+g^4)$
C_2	3	4	1	2	(13)(24)	2	(0,2,0,0)	$(b^2+r^2)^2$	$(b^2+r^2+g^2)^2$
C_4^{-1}	4	1	2	3	(1432)	1	(0,0,0,1)	(b^4+r^4)	$(b^4+r^4+g^4)$
σ_{v_1}	4	3	2	1	(14)(23)	2	(0,2,0,0)	$(b^2+r^2)^2$	$(b^2+r^2+g^2)^2$
σ_{v_2}	2	1	4	3	(12)(34)	2	(0,2,0,0)	$(b^2+r^2)^2$	$(b^2+r^2+g^2)^2$
σ_{d_1}	3	2	1	4	(13)(2)(4)	3	(2,1,0,0)	$(b+r)^2\cdot(b^2+r^2)$	$(b+r+g)^2\cdot(b^2+r^2+g^2)$
σ_{d_2}	1	4	3	2	(1)(3)(24)	3	(2,1,0,0)	$(b+r)^2\cdot(b^2+r^2)$	$(b+r+g)^2\cdot(b^2+r^2+g^2)$

Table 1. Action of the 8 symmetry operators g of the C_{4v} group on the 4-sites square (see Figure 1-A). "Moves" column gives the one-line notation for the permutation of the $|D|$=4 sites under the effect of each g; e.g. in the case of C_2, "3412" reads as "1 goes to 3, 3 to 1, 2 to 4 and 4 to 2". The cycles generated by each g are given in the third column, while their number $|Cyc_4(g)|$ is in the fourth column. $\mathcal{T}_4(g)$ is a 4-plet whose i^{th} value l_i indicates the number of cycles of length i resulting in the "Cycles" column. $pp_2^4(g)$ gives the contribution of each g to the Pólya's polynomial $PP_2^4(b,r)$ (Eqs. (17)-(18)); the subscript and superscript (2 and 4 in this case) are the number of colors and the label of the set of sites, respectively. $pp_3^4(g)$ gives the contributions to $PP_3^4(b,r,g)$.

g	Cycles	$\lvert Cyc_8(g)\rvert$	$\mathcal{T}_8(g)$	$pp_2^8(g)$
E	(1)(2)(3)(4)(5)(6)(7)(8)	8	(8,0,0,0,0,0,0,0)	$(b+r)^8$
C_4	(1234)(5678)	2	(0,0,0,2,0,0,0,0)	$(b^4+r^4)^2$
C_2	(13)(24)(57)(56)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
C_4^{-1}	(1432)(5876)	2	(0,0,0,2,0,0,0,0)	$(b^4+r^4)^2$
σ_{v_1}	(18)(27)(36)(45)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
σ_{v_2}	(16)(25)(38)(47)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
σ_{d_1}	(17)(26)(35)(48)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
σ_{d_2}	(15)(26)(37)(48)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$

Table 2. Action of the 8 symmetry operators of the C_{4v} group on the 8-sites square (see Figure 1-B). Symbols as in Table 1.

CC		Cycles	$\lvert Cyc_8(CC)\rvert$	$\mathcal{T}_8(CC)$	$pp_2^8(CC)$
E	(1)	(1)(2)(3)(4)(5)(6)(7)(8)	8	(8,0,0,0,0,0,0,0)	$(b+r)^8$
C_2	(3)	(12)(34)(56)(78)	4	(0,4,0,0,0,0,0,0)	$3\cdot(b^2+r^2)^4$
C_2'	(6)	(15)(26)(37)(48)	4	(0,4,0,0,0,0,0,0)	$6\cdot(b^2+r^2)^4$
C_3	(8)	(1)(234)(5)(687)	4	(2,0,2,0,0,0,0,0)	$8\cdot(b+r)^2\cdot(b^3+r^3)^2$
C_4	(6)	(1827)(3645)	2	(0,0,0,2,0,0,0,0)	$6\cdot(b^4+r^4)^2$
i	(1)	(1)(2)(3)(4)(5)(6)(7)(8)	8	(8,0,0,0,0,0,0,0)	$(b+r)^8$
S_4	(6)	(1827)(3645)	2	(0,0,0,2,0,0,0,0)	$6\cdot(b^4+r^4)^2$
S_6	(8)	(1)(234)(5)(687)	4	(2,0,2,0,0,0,0,0)	$8\cdot(b+r)^2\cdot(b^3+r^3)^2$
σ_h	(3)	(12)(34)(56)(78)	4	(0,4,0,0,0,0,0,0)	$3\cdot(b^2+r^2)^4$
σ_d	(6)	(15)(26)(37)(48)	4	(0,4,0,0,0,0,0,0)	$6\cdot(b^2+r^2)^4$

Table 3. Action of the space group $Ia\overline{3}d$ (48 symmetry operators) on the 8 octahedral sites of the garnet structure (primitive cell). Symbols as in Table 1. The first column lists the conjugacy classes CC of symmetry operators; their cardinal is given in brackets. $pp_2^8(CC)$ gives the contribution of each CC to the Pólya's polynomial $PP_2^8(b,r)$ (Eqs. (17)-(18)); note that it has the cardinal of the class as a multiplying factor.

		Primitive cell		Conventional cell				
CC		$\lvert Cyc_8\rvert$	$\lvert Cyc_{12}\rvert$	$\lvert Cyc_{16}\rvert$		$\lvert Cyc_{24}\rvert$		
E	(1)	8	12	16	8	24	12	
C_2	(3)	4	8	8	8	16	12	
C_2'	(6)	4	8	8	8	14	14	
C_3	(8)	4	4	8	4	8	4	
C_4	(6)	2	4	4	4	6	6	
i	(1)	8	6	12	12	12	12	
S_4	(6)	2	4	4	4	8	8	
S_6	(8)	4	2	6	6	4	4	
σ_h	(3)	4	6	8	8	12	12	
σ_d	(6)	4	6	4	4	6	6	

Table 4. Number of cycles $\lvert Cyc_D\rvert$ resulting from the action of the space group $Ia\overline{3}d$ on four sets of sites of the garnet structure. $\lvert Cyc_8\rvert$ and $\lvert Cyc_{12}\rvert$ refer to the 8 octahedral and 12 dodecahedral sites, respectively, of the primitive cell. $\lvert Cyc_{16}\rvert$ and $\lvert Cyc_{24}\rvert$ refer to the same 16 and 24 sites of the conventional cell. The first column lists the conjugacy classes CC of symmetry operators g; their cardinal is given in brackets. In the conventional cell case, in each $\lvert Cyc_D\rvert$ column there are two sets of values: the first refers to the classes CC in the first column, the second to the classes CC', resulting from the composition of each CC with the centering vector $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ of the conventional cell.

	Primitive Cell		Conventional Cell	
$\|R\|$	$N^8_{\|R\|}$	$N^{12}_{\|R\|}$	$N^{16}_{\|R\|}$	$N^{24}_{\|R\|}$
2	23	154	874	179'444
3	333	12'489	461'889	2'943'985'419
4	2'916	362'776	45'112'096	2'932'200'891'456
5	16'725	5'163'025	1'594'680'625	620'887'278'324'375
6	70'911	45'674'826	29'432'496'906	49'358'237'168'514'996

Table 5. Total number of SICs $N^D_{|R|}$ resulting from the action of the space group $Ia\bar{3}d$ on four sets of sites of the garnet structure, as a function of the number of colors $|R|$. $N^8_{|R|}$ and $N^{12}_{|R|}$ refer to the 8 octahedral and 12 dodecahedral sites, respectively, of the primitive cell. $N^{16}_{|R|}$ and $N^{24}_{|R|}$ refer to the same sites of the conventional cell.

g	Cycles		$\|Cyc_8(g)\|$	$\mathcal{T}_8(g)$	$pp^8_2(g)$
E	(1)(2)(3)(4)	(5)(6)(7)(8)	8	(8,0,0,0,0,0,0,0)	$(b+r)^8$
C_2^a	(13)(24)	(57)(68)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
C_2^b	(14)(23)	(58)(67)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
C_2^c	(12)(34)	(56)(78)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
i	(1)(2)(3)(4)	(56)(78)	6	(4,2,0,0,0,0,0,0)	$(b+r)^4\cdot(b^2+r^2)^2$
σ_h	(12)(34)	(5)(6)(7)(8)	6	(4,2,0,0,0,0,0,0)	$(b+r)^4\cdot(b^2+r^2)^2$
σ_{v_1}	(13)(24)	(58)(67)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$
σ_{v_2}	(14)(23)	(57)(68)	4	(0,4,0,0,0,0,0,0)	$(b^2+r^2)^4$

Table 6. Action of the space group *Pbnm* (8 symmetry operators) on the 8 octahedral sites of the olivine structure (primitive cell). Symbols as in Table 1.

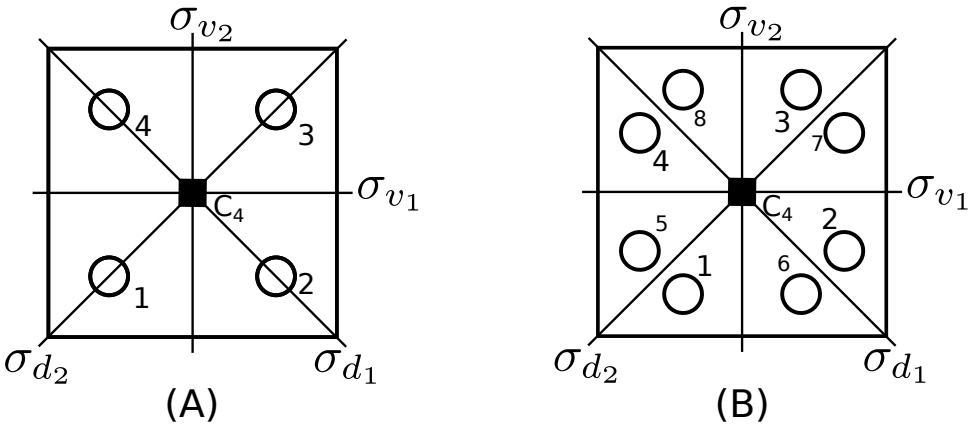

Figure 1. Action of the symmetry group C_{4v} (8 symmetry operators) on two sets of objects in a square: (A) 4 symmetry equivalent sites lying on the σ_d reflection planes (diagonals); (B) 8 symmetry equivalent sites in general position. The full list of the C_{4v} symmetry operators is given in the first column of Table 1.

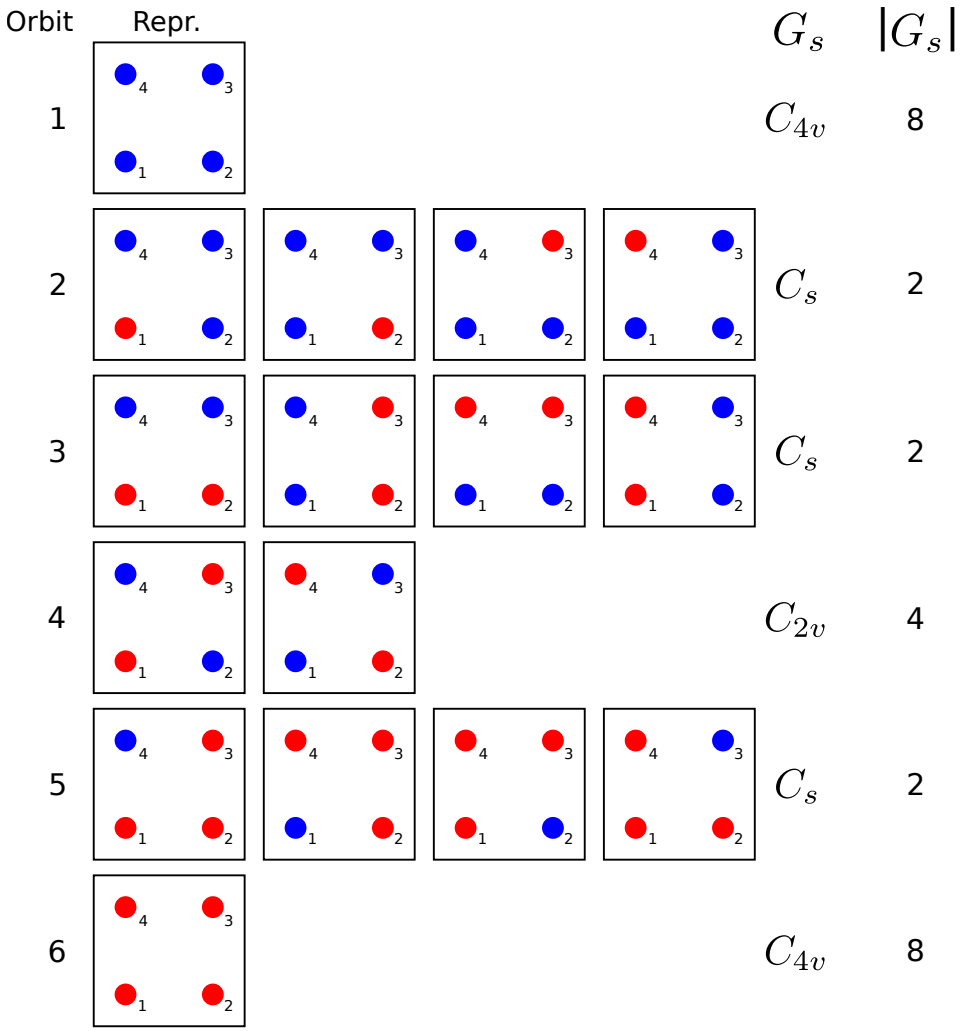

Figure 2. C_{4v} group acting on the 4-sites, 2-colors square (see Figure 1-A): the set of $2^4 = 16$ configurations, grouped in the 6 orbits. Each row corresponds to an orbit; the first configuration of each orbit has been chosen as its "canonical" representative. The orbit stabilizer G_s and its cardinal $|G_s|$ are given in the last two columns (Schoenflies notation).

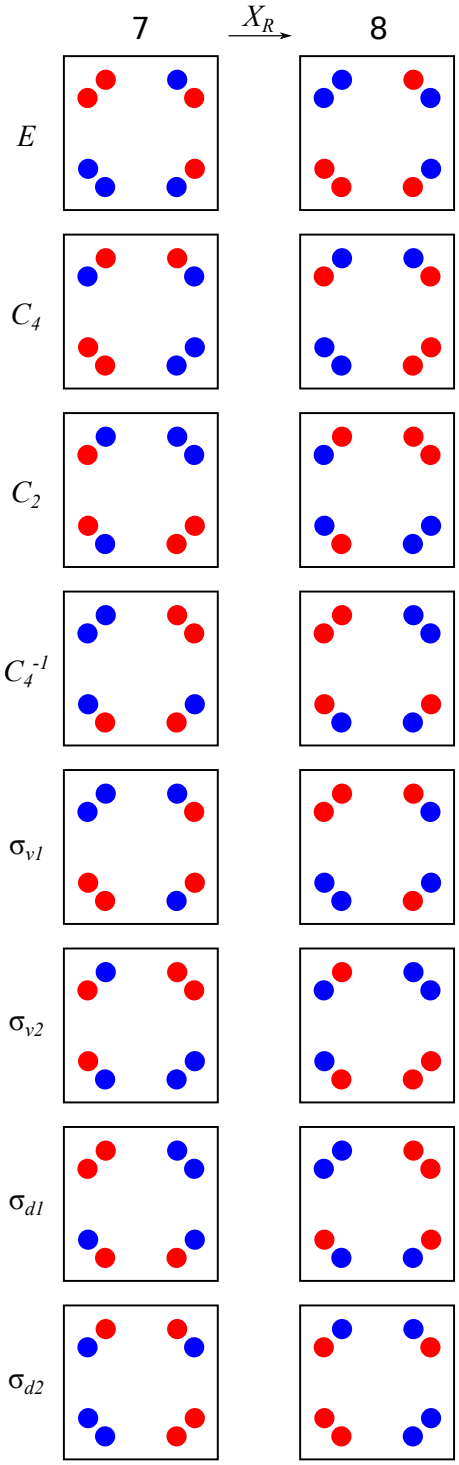

Figure 3. C_{4v} group acting on the 8-sites, 2-colors square (see Figure 1-B): effect of the additional action of $H = S_2$ group on the set of colors. The configurations belonging to orbits 7 and 8 for the 50% composition $(4b, 4r)$ (13 orbits in total) are reported in two columns. These two orbits are symmetry independent in the general b, r-colors case. On the contrary, they compose a single orbit of symmetry equivalent configurations in the $\uparrow, \downarrow$-colors case, i.e. when the $H = S_2$ group acts on the 2 colors. Couples of configurations that become equivalent due to the action of the exchange operator $X_R \in S_2$ are on the same row. The symmetry operator $g \in C_{4v}$, that must be applied to the first-row configurations to obtain the configurations in each row, is reported in the left column.

	Sites				
ω	4	3	2	1	
1	0	0	0	0	ℓ_1
2	0	0	0	1	ℓ_2
	0	0	1	0	ℓ_3
3	0	0	1	1	ℓ_4
	0	1	0	0	ℓ_5
4	0	1	0	1	ℓ_6
	0	1	1	0	ℓ_7
5	0	1	1	1	ℓ_8
	1	0	0	0	ℓ_9
	1	0	0	1	ℓ_{10}
	1	0	1	0	ℓ_{11}
	1	0	1	1	ℓ_{12}
	1	1	0	0	ℓ_{13}
	1	1	0	1	ℓ_{14}
	1	1	1	0	ℓ_{15}
6	1	1	1	1	ℓ_{16}

Figure 4. C_{4v} group acting on the 4-sites, 2-colors square (see Figure 1-A): generation of the canonical representatives through orderly generation; the left column gives the lexicographic order (LO) of the representatives.

	Sites				
ω	4	3	2	1	
1	0	1	1	2	f_1
2	0	1	2	1	f_2
	0	2	1	1	f_3
	1	0	1	2	f_4
	1	0	2	1	f_5
	1	1	0	2	f_6
	1	1	2	0	f_7
	1	2	0	1	f_8
	1	2	1	0	f_9
	2	0	1	1	f_{10}
	2	1	0	1	f_{11}
	2	1	1	0	f_{12}

Figure 5. C_{4v} group acting on the 4-sites, 3-colors square (see Figure 1-A): generation of the canonical representatives at fixed composition $(b, 2r, g)$ through orderly generation; the left column gives the LO of the representatives. All configurations at constant composition $(b, 2r, g)$ are listed in lexicographic order. The first one, the lowest, (f_1), is canonical. The second one (f_2) cannot be transformed in the previous one by any operator, it is then canonical. Pólya's polynomial (Table 1) indicates that there are only 2 representatives for this composition, then the search is stopped.

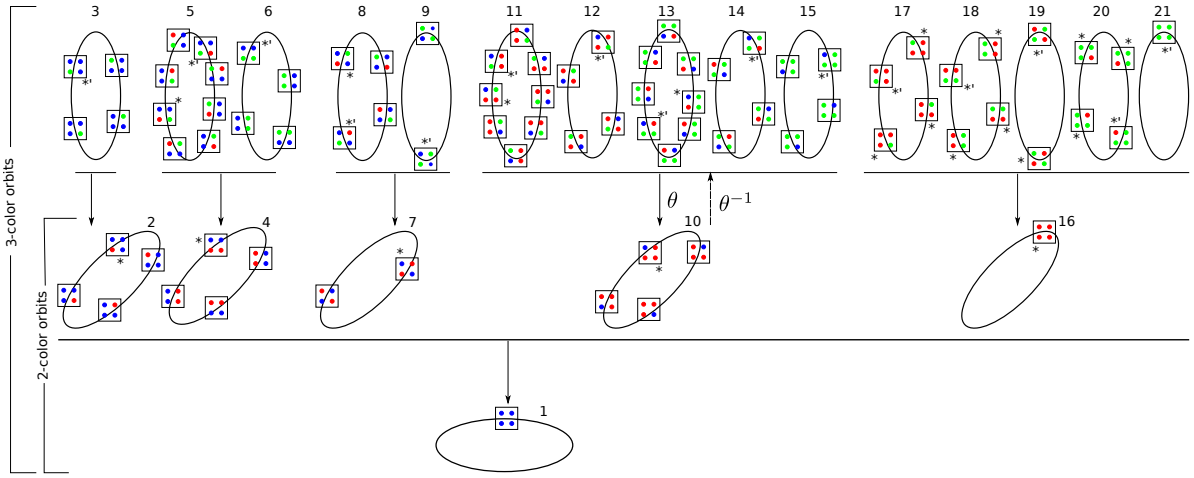

Figure 6. C_{4v} group acting on the 4-sites square (see Figure 1-A): orbits *vs* number of colors $|R|$, from one (blu, bottom layer), to two (red, middle layer), to three (green, top layer). The number of orbits increases from 1 to 6 to 21. The canonical representatives (SICs) for the $|R|+1$ case are generated from the set of SICs for the $|R|$ case by using the surjective resolution principle. The arrows represent the θ mapping, according to Eqs. (42)-(43), Section 2.4. Asterisks in the 2-colors (middle) layer label the 2-colors SICs. Asterisks in the 3-colors (top) layer label the configurations belonging to the θ-preimages of the 2-colors SICs; among them, primed asterisks label the 3-colors SICs.

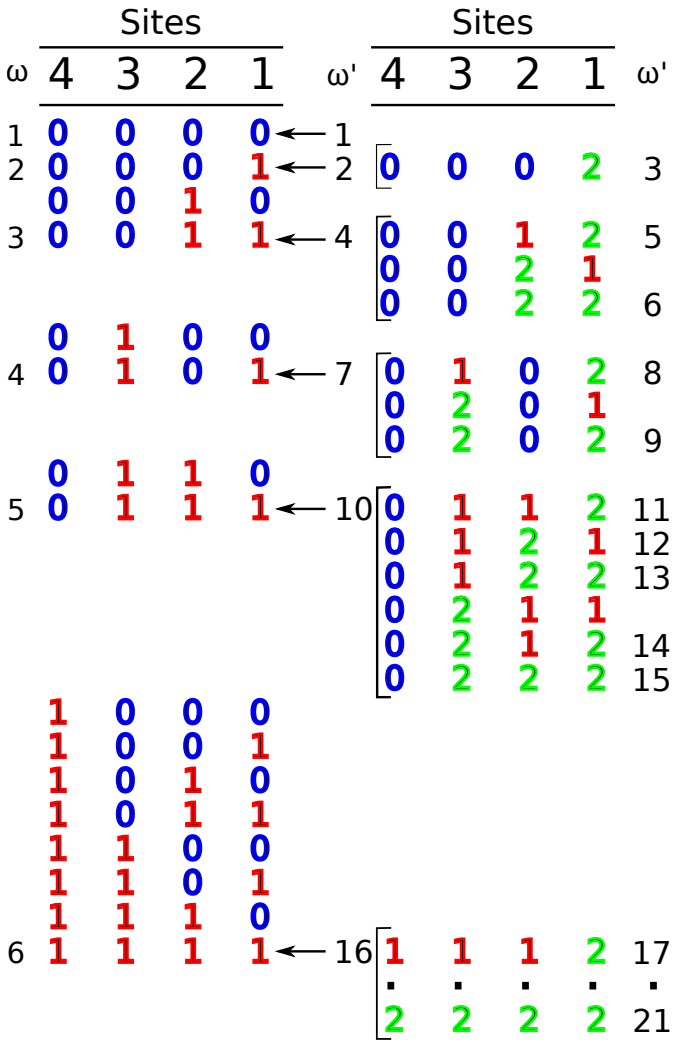

Figure 7. C_{4v} group acting on the 4-sites, 3-colors square (see Figure 1-A): generation of the canonical representatives (SICs) through lexicographic ordering (LO) combined with surjective resolution. On the left side, the 2-colors configurations are given. The left-most column gives the LO of the corresponding SICs in the 2-colors only case (see also Figure 4), while the central column indicates their LO in the 3-colors case. On the right side, the additional 3-colors configurations are listed. The arrows in the central column represent the θ mapping according to the surjective resolution principle (see Eqs. (42)-(43), Section 2.4). In the right-most column the LO of the additional 3-colors SICs is reported.

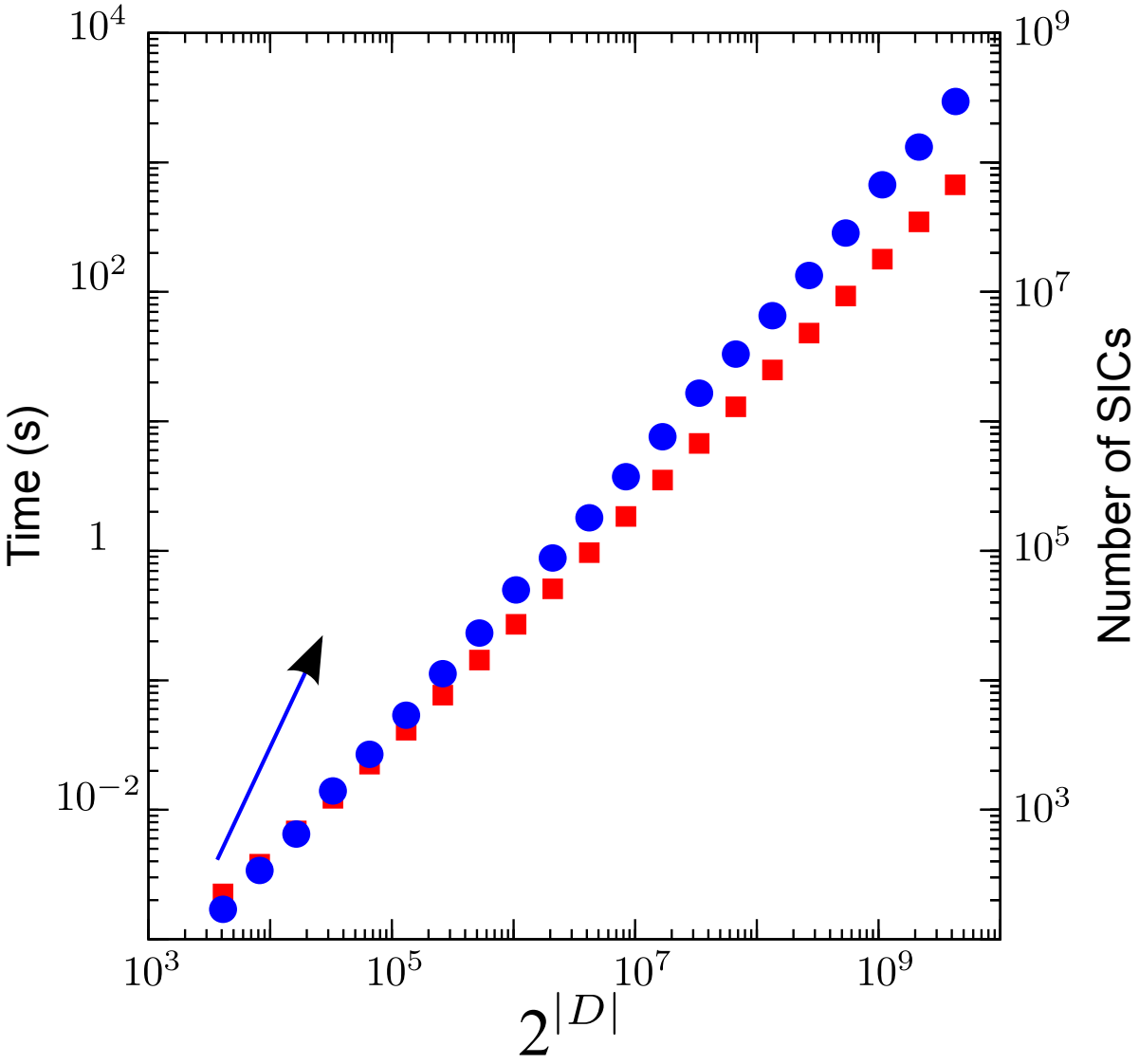

Figure 8. CPU time (circles, logarithmic scale) needed to generate the SICs using lexicographical ordering, as a function of the size of the system $|R|^{|D|}$, for $|R| = 2$. The reference system is a tetragonal MgO (n,1,1) supercell, with $12 \leq n \leq 32$. There are $12 \leq |D| \leq 32$ sites and $48 \leq |G| \leq 128$ symmetry operators. The number of SICs is represented by squares. The arrow indicates the time growth for direct method (see Eq. (48)). The scaling difference between the time and SICs curves is in the order of $|G|^2$ (see Eqs. (13) and (47)).

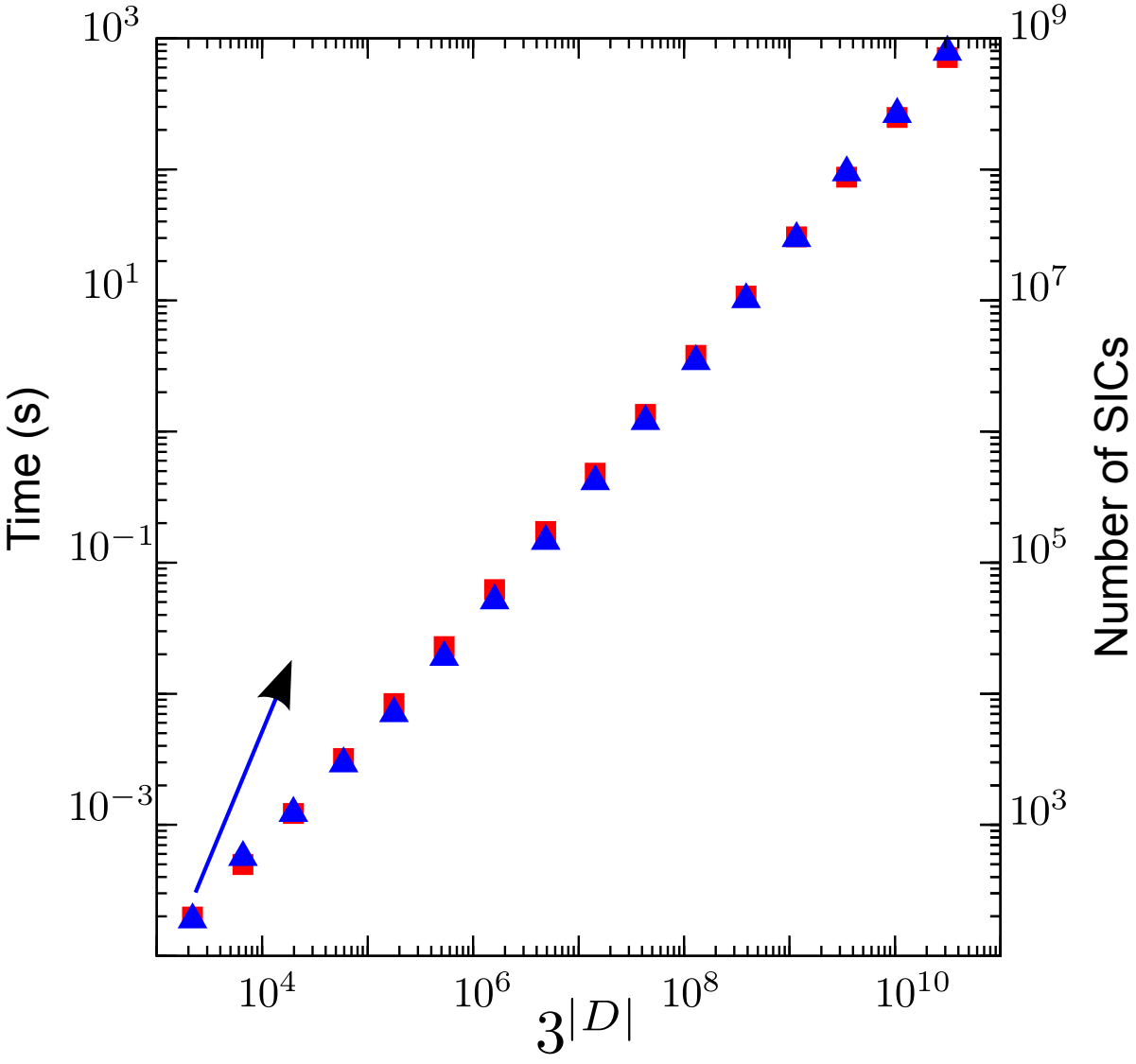

Figure 9. CPU time (triangles, logarithmic scale) needed to generate the SICs using surjective resolution, as a function of the size of the system $|R|^{|D|}$, for $|R| = 3$. The reference system is a tetragonal MgO (n,1,1) supercell, with $7 \leq n \leq 22$. There are $7 \leq |D| \leq 22$ sites and $28 \leq |G| \leq 88$ symmetry operators. The number of SICs is represented by squares. The arrow indicates the time growth for direct method (see Eq. (48)).

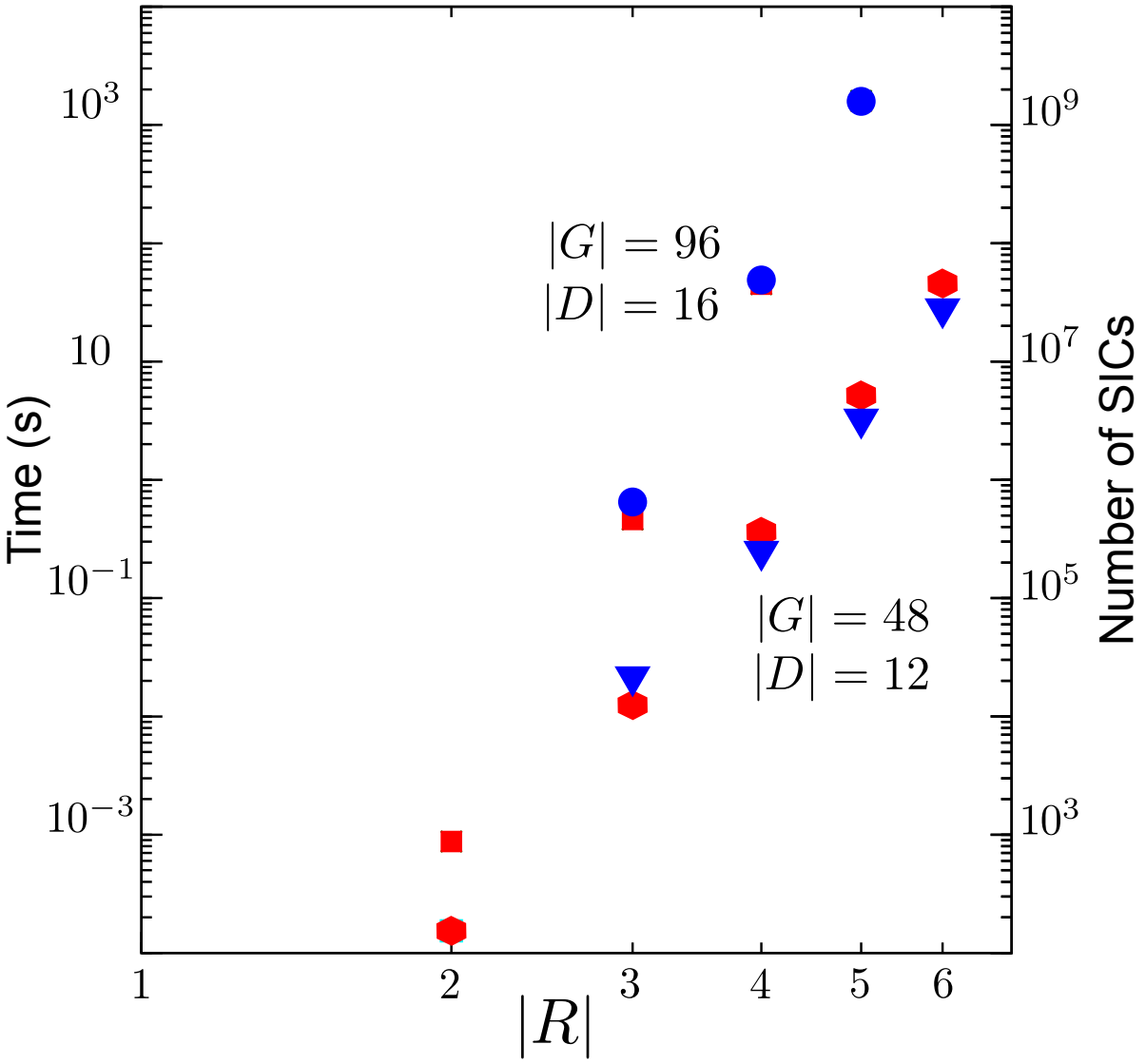

Figure 10. CPU time (circles and triangles, logarithmic scale) needed to generate the SICs using surjective resolution, as a function of the number of colors $|R|$. The reference system is the garnet structure, either octahedral sites in the conventional cell (circles, $|D| = 16$, $|G| = 96$ symmetry operators) or dodecahedral sites in the primitive cell (triangles, $|D| = 12$, $|G| = 48$). The number of SICs is represented by squares and hexagons for the two cases, respectively. For details on the garnet system, see Section 3.

References

[1] R. Dovesi, B. Civalleri, R. Orlando, C. Roetti, and V. R. Saunders. *Ab initio* Quantum Simulation in Solid State Chemistry. volume 21 of *Reviews in Computational Chemistry*, pages 1–125. John Wiley and Sons, 2005.

[2] P. Ugliengo, M. Sodupe, F. Musso, I. J. Bush, R. Orlando, and R. Dovesi. Realistic Models of Hydroxylated Amorphous Silica Surfaces and MCM-41 Mesoporous Material Simulated by Large-scale Periodic B3LYP Calculations. *Adv. Mater.*, 20:4579–4583, 2008.

[3] J. VandeVondele, U. Borštnik, and J. Hutter. Linear Scaling Self-Consistent Field Calculations with Millions of Atoms in the Condensed Phase. *J. Chem. Theory Comput.*, 8:3565–3573, 2012.

[4] D. B. Laks, L. G. Ferreira, S. Froyen, and A. Zunger. Efficient cluster expansion for substitutional systems. *Phys. Rev. B*, 46:12587–12605, 1992.

[5] M. H. F. Sluiter and Y. Kawazoe. Cluster expansion method for adsorption: Application to hydrogen chemisorption on graphene. *Phys. Rev. B*, 68:085410, 2003.

[6] V. L. Vinograd, M. H. F. Sluiter, B. Winkler, A. Putnis, U. Halenius, J. D. Gale, and U. Becker. Thermodynamics of mixing and ordering in pyrope-grossular solid solution. *Mineral. Mag.*, 68:101–121, 2004.

[7] M. H. F. Sluiter, V. Vinograd, and Y. Kawazoe. Intermixing tendencies in garnets: Pyrope and grossular. *Phys. Rev. B*, 70:184120, 2004.

[8] C. L. Freeman, N. L. Allan, and W. van Westrenen. Local cation environments in the pyrope-grossular $Mg_3Al_2Si_3O_{12}$-$Ca_3Al_2Si_3O_{12}$ garnet solid solution. *Phys. Rev. B*, 74:134203, 2006.

[9] Y. G. Yu, R. M. Wentzcovitch, V. L. Vinograd, and R. J. Angel. Thermodynamic properties of MgSiO3 majorite and phase transitions near 660 km depth in MgSiO3 and Mg2SiO4: A first principles study. *J. Geophys. Res.*, 116:B02208, 2011.

[10] A. Meyer, Ph. D'Arco, R. Orlando, and R. Dovesi. Andradite-Uvarovite Solid Solutions. An ab Initio All-Electron Quantum Mechanical Simulation with the CRYSTAL06 Code. *J. Phys. Chem. C*, 113:14507–14511, 2009.

[11] L. G. Ferreira, S.-H. Wei, and A. Zunger. Stability, Electronic Structure, and Phase Diagrams of Novel Inter-Semiconductor Compounds. *Int. J. Supercomput. Ap.*, 5:34–57, 1991.

[12] J. M. Sanchez and D. de Fontaine. The fcc Ising model in the cluster variation approximation. *Phys. Rev. B*, 17:2926–2936, 1978.

[13] R. Magri, J. E. Bernard, and A. Zunger. Predicting structural energies of atomic lattices. *Phys. Rev. B*, 43:1593–1597, 1991.

[14] A. van de Walle and G. Ceder. Automating first-principles phase diagram calculations. *J. Phase Equilib.*, 23:348359, 2002.

[15] J. S. Rutherford. The enumeration and symmetry-significant properties of derivative lattices. *Acta Crystallogr. Sec. A*, 48:500–508, 1992.

[16] J. S. Rutherford. The enumeration and symmetry-significant properties of derivative lattices. II. Classification by colour lattice group. *Acta Crystallogr. Sec. A*, 49:293–300, 1993.

[17] J. S. Rutherford. The enumeration and symmetry-significant properties of derivative lattices. III. Periodic colourings of a lattice. *Acta Crystallogr. Sec. A*, 51:672–679, 1995.

[18] G. L. W. Hart and R. W. Forcade. Algorithm for generating derivative structures. *Phys. Rev. B*, 77:224115, 2008.

[19] G. L. W. Hart and R. W. Forcade. Generating derivative structures from multilattices: Algorithm and application to hcp alloys. *Phys. Rev. B*, 80:014120, 2009.

[20] R. Grau-Crespo, S. Hamad, C. R. A. Catlow, and N. H. de Leeuw. Symmetry-adapted configurational modeling of fractional site occupancy in solids. *J. Phys: Condens. Matter*, 19:256201, 2007.

[21] Q. Wang, R. Grau-Crespo, and N.H. de Leeuw. Mixing Thermodynamics of the Calcite-Structured (Mn,Ca)CO3 Solid Solution: A Computer Simulation Study. *J. Phys. Chem. B*, 115:13854–13861, 2011.

[22] S. Haider, R. Grau-Crespo, A. J. Devey, and N. H. de Leeuw. Cation distribution and mixing thermodynamics in Fe/Ni thiospinels. *Geochimica et Cosmochimica Acta*, 88:275–282, 2012.

[23] M. Habgood, R. Grau-Crespo, and S. L. Price. Substitutional and orientational disorder in organic crystals: a symmetry-adapted ensemble model. *Phys. Chem. Chem. Phys.*, 13:9590–9600, 2011.

[24] R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, Ph. D'Arco, and M. Llunell. *CRYSTAL 2009 User's Manual*. University of Torino, Torino, 2009.

[25] S. Mustapha, Ph. D'Arco, M. de la Pierre, Y. Noel, M. Ferrabone, and R. Dovesi. Random selection of symmetry independent configurations for the simulation of disordered solids. *In preparation*, 2012.

[26] G. Pólya and R. C. Read. *Combinatorial enumeration of groups, graphs, and chemical compounds*. Springer-Verlag, New York, 1987.

[27] Note that here the term "canonical" should not be confused with the similar term used for the statistical mechanics classification of the ensembles.

[28] A. Kerber. *Applied finite group action*. Springer-Verlag, New York, 1999.

[29] S. G. Williamson. *Combinatorics for computer science*. Computer Science Press, New York, 1985.

[30] W. Oberschelp. Kombinatorische Anzahlbestimmungen in Relationen. *Math. Ann.*, 174:53–78, 1967.

[31] J. D. Dixon and H. S. Wilf. The random selection of unlabeled graphs. *J. Algorithms*, 4:205–213, 1983.

[32] N. G. De Bruijn. Pólya's theory of counting. In E. Beckenbach, editor, *Applied Combinatorial Mathematics*, pages 144–184. John Wiley, New York-London-Sidney, 1964.

[33] N. G. de Bruijn. Generalization of Pólya's fundamental theorem in enumerative combinatorial analysis. *Indag. Math.*, 21:59–69, 1959.

[34] F. Harary and E. Palmer. The power group enumeration theorem. *J. Combin. Theory*, 1:157–173, 1966.

[35] R. C. Read. Every One a Winner or How to Avoid Isomorphism Search when Cataloguing Combinatorial Configurations. *Annals Discrete Math.*, 2:107–120, 1978.

[36] L. A. Goldberg. Efficient algorithms for writing unlabeled graphs. *J. Algorithms*, 13:128–143, 1992.

[37] G. L. W. Hart, L. J. Nelson, and R. W. Forcade. Generating derivative structures at a fixed concentration. *Comp. Mater. Sci.*, 59:101–107, 2012.

[38] W. Deer, R. Howie, and J. Zussman. *An introduction to the rock forming minerals*. John Wiley, New York, 1992.

[39] P. C. Rickwood. On recasting analyses of garnet into end-member molecules. *Contrib. Mineral. Petrol.*, 18:175–198, 1968.

[40] F. J. Molster, L. B. F. M. Waters, N. R. Trams, H. van Winckel, L. Decin, J. T. van Loon, C. Jager, T. Henning, H. U. Kaufl, A. De Koter, and J. Bouwman. The composition and nature of the dust shell surrounding the binary AFGL 4106. *Astronomy and Astrophysics*, 350:163–180, 1999.