

Crystal structure prediction by generalized disjunctive programming

Ryotaro Koshoji ^{*}

Institute for Solid State Physics, The University of Tokyo, Kashiwa 277-8581, Japan



(Received 9 June 2026; accepted 29 July 2026; published 31 August 2026)

Inorganic structural chemistry leads naturally to a theory for crystal structure prediction formalized by a generalized disjunctive programming, which is formulated using continuous and Boolean variables to involve the algebraic equations, disjunctions, and logic propositions. Since the feasibilities of continuous variables change drastically depending on Boolean variables and vice versa, iterative optimization of continuous and Boolean variables efficiently transforms a randomly generated initial structure into an optimal solution. Boolean variables are introduced to elucidate the “combinatorial backbone” of the original continuous optimization problem, which corresponds to the graphs describing crystal structures with clear chemical meanings. Since many subgraphs are discovered in numerous graphs generated through structural optimizations, on-the-fly learning of the feasibilities of them accelerates the discovery of the optimal solutions. This theory enables designing a wide variety of crystal structures with small computations based on only the atomic radii and feasible coordination numbers of each atom.

DOI: [10.1103/6vcv-5zwf](https://doi.org/10.1103/6vcv-5zwf)

I. INTRODUCTION

Crystal structure prediction is the optimization problem of the Gibbs free energy, which is the function of continuous variables including spatial arrangement of atoms through the many-body wavefunction. This is a very difficult optimization problem like a long-standing scandal in physics [1] due to countless local optima, but many techniques using metaheuristics such as random search methods [2–4], particle swarm optimization [5–8], or evolutionary algorithms [9–16], have made it now possible to predict a crystal structure given only its composition for systems that are not too complicated. They are designed to focus on the set of reasonable possibilities for arrangements of atoms, which is much less than the set of all possible spatial orders of atoms.

Essentially, deterministic algorithms to solve optimization problems must visit every local optimum, but conventional nonlinear optimization techniques for real variables, which rely on such concepts as derivatives and gradients, have been mainly concerned with finding local optima around initial states. While there are methods to replace a too difficult original problem with a more easily solvable relaxed problem [17], no deterministic polynomial-time algorithm to solve nondeterministic polynomial (NP)-hard problems has been found, as if the algorithm is consistent with $P \neq NP$ conjecture. A common mechanism behind NP-hardness of nonconvex continuous optimization is that the continuous feasible region or objective contains a “combinatorial backbone,” such

as vertices of a polytope, or sign patterns, etc. The global optimization over the continuous space contains the solution of such an NP-hard discrete problem, and therefore becomes hard [18]. In contrast to the formal theory, the branch and bound algorithm for integer programmings, which is capable of rapidly eliminating large parts of the optimization domain from consideration, is devised to find the global optimum more efficiently than random samplings [19]. Then, the next question is whether we can introduce integer variables into crystal structure prediction so as to not only clarify the mathematical structure of crystal structure prediction such as NP-hardness but also suggest efficient algorithms for handling it.

Almost all the eigenvalue problems for many-body wavefunctions are unsolvable. The density functional theory, which replaces them with different auxiliary systems that can be solved more easily, makes large-scale crystal structure searches possible, but the exact form of the exchange-correlation energy is unknown. That is, nobody knows the explicit mapping of spatial orders of atoms to the objective function for crystal structure prediction. However, experimental synthesizers of inorganic compounds know the empirical rules to elucidate the crystal structures, which are organized as the science named inorganic structural chemistry [20–23]. It deals with both the continuous and discrete aspects of crystal structures; the former is the packing of atomic spheres, and the latter is the coordination polyhedra and the linking of them.

It was not until 2000s that Kepler conjecture was proved positively despite the simpleness of the problem [24]. The densest packings of spheres of only two or three sizes are still unknown; however, many methods have predicted them by computer [25–34]. They were applied to design unknown crystal structures such as quaternary hydrides under high pressure [35], but the predictive power from them seems to be limited [32–35].

^{*}Contact author: cosaji@issp.u-tokyo.ac.jp

A lot of program packages have been developed to represent the discrete aspects of the structures of chemical compounds by graphs [36–42]. Many machine-learning techniques also use the graph neural networks to provide a prediction of not only material properties [43–50] but also crystal structures [51–58]. However, it is still a challenge to inversely design solid-state materials with graphs because the reconstruction of the original crystal structure from the graph is difficult. Infinite periodic three-dimensional structures with translational, rotational, and permutational invariances cannot be represented as undirected finite graphs unlike molecular inverse design [55].

II. THEORY

This article shows that the next logical step would be to combine the continuous and discrete aspects to develop an efficient solution method for crystal structure prediction. It is not very difficult to do: Inorganic structural chemistry leads naturally to a theory using continuous and discrete variables, which not only elucidates the mathematical structure of crystal structure prediction but also provides a high-speed algorithm to design prototypes of crystal structures.

First, we define the feasible atomic environment of each atom as

$$(r^{(G)}, R^{(G)}, c^{(G)}, C^{(G)}) \quad (\forall G \in \mathcal{G}), \quad (1)$$

where $r^{(G)}, R^{(G)} \in \mathbb{R}$ are the minimum and maximum radii, respectively, $c^{(G)}, C^{(G)} \in \mathbb{N}$ are the minimum and maximum coordination numbers, respectively, and $G \in \mathcal{G}$ corresponds to a type of interatomic force such as ionic bond, covalent bond, and electrostatic repulsion. If G does not correspond to a chemical bond, it is quite natural to set $R^{(G)} = \infty$, $c^{(G)} = 0$, and $C^{(G)} = \infty$. Think about atom $j\mathbf{T}$ as the atom j in lattice $\mathbf{T} \in \mathbb{Z}^3$. To constrain the feasible interatomic distance between atoms i and $j\mathbf{T}$, we must select one type of interatomic distance constraint G between them. At this point, we need to introduce the Boolean variables $Y_{ij\mathbf{T}}^{(G)} \in \{\text{True}, \text{False}\}$, where $Y_{ij\mathbf{T}}^{(G)}$ is True if and only if G is selected. We can now write the

constraints on interatomic distance:

$$\bigvee_{G \in \mathcal{G}} \left[d_{ij}^{(G)} \leq |\mathbf{x}_j + A\mathbf{T} - \mathbf{x}_i| \leq D_{ij}^{(G)} \right], \quad (2)$$

where $d_{ij}^{(G)}$ and $D_{ij}^{(G)}$ fix the feasible range of interatomic distance, $A \equiv (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ is the lattice vectors of the unit cell, and $\mathbf{x}_i$ is the position vector of atom i . Each term of disjunction, which is associated with a Boolean variable and the inequality constraint on interatomic distance, is linked together by or operator ($\vee$). When a disjunctive term is selected ($Y_{ij\mathbf{T}}^{(G)}$ is True), the corresponding inequalities are enforced. Otherwise, the constraints are ignored. Since every pair of atoms should select one type of interatomic forces, we impose the constraints that exactly one of the Boolean variables can be True:

$$\bigvee_{G \in \mathcal{G}} Y_{ij\mathbf{T}}^{(G)}, \quad (3)$$

where $\bigvee$ is the exclusive or (xor) operator. Now, recall that each atom has the feasible coordination numbers with respect to G . We arrive at the constraint on the number of active Boolean variables, which is represented by cardinality clauses defined by the following predicates: Exactly($m, Y_s, \forall s \in \mathcal{S}$) enforces that exactly m of the Boolean variables Y_s are True, AtLeast($m, Y_s, \forall s \in \mathcal{S}$) enforces that at least m of Y_s are True, and AtMost($m, Y_s, \forall s \in \mathcal{S}$) enforces that at most m are True.

We are now ready to construct the theory for crystal structure prediction. Starting from inorganic structural chemistry, we replace the original optimization problem of crystal structure prediction, which is too difficult to solve, by the different auxiliary problem that relies on the intersections of disjunctions of algebraic constraints (equality and inequality constraints on continuous variables) controlled by Boolean variables. Such a model is known as generalized disjunctive programming (GDP) [59–65], which can include logic propositions expressed by logic operators such as AND ($\wedge$), OR ($\vee$), XOR ($\bigvee$), negation ($\neg$), implication ($\Rightarrow$), and equivalence ($\iff$). We are now ready to write down the GDP to design prototypes of crystal structures. Here, it is

$$\begin{aligned} & \underset{A, \mathbf{x}_i, Y_{ij\mathbf{T}}^{(G)}}{\text{minimize}} && |\det A|, \\ & \text{subject to} && \text{AtLeast}(c_i^{(G)}, Y_{ij\mathbf{T}}^{(G)}, \forall j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3) && (\forall i \in \mathcal{I}, \forall G \in \mathcal{G}), \\ & && \text{AtMost}(C_i^{(G)}, Y_{ij\mathbf{T}}^{(G)}, \forall j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3) && (\forall i \in \mathcal{I}, \forall G \in \mathcal{G}), \\ & && \bigvee_{G \in \mathcal{G}} Y_{ij\mathbf{T}}^{(G)} && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3), \\ & && \bigvee_{G \in \mathcal{G}} \left[d_{ij\mathbf{T}}^{(G)} \leq |\mathbf{x}_j + A\mathbf{T} - \mathbf{x}_i| \leq D_{ij\mathbf{T}}^{(G)} \right] && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3), \\ & && \neg \Lambda_{ij}^{(G)} \Rightarrow \neg Y_{ij\mathbf{T}}^{(G)} && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3, \forall G \in \mathcal{G}), \\ & && \Xi_{ij}^{(G)} \Rightarrow Y_{ij\mathbf{T}}^{(G)} && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3, \forall G \in \mathcal{G}), \\ & && d_{ij\mathbf{T}}^{(G)} \equiv r_i^{(G)} + r_j^{(G)}, \quad D_{ij\mathbf{T}}^{(G)} \equiv R_i^{(G)} + R_j^{(G)}, \quad \mathbf{0} \leq A^{-1}\mathbf{x}_i \leq \mathbf{1}, \\ & && A \in \mathbb{R}^{3 \times 3}, \quad \mathbf{x}_i \in \mathbb{R}^3, \quad Y_{ij\mathbf{T}}^{(G)} \in \{\text{True}, \text{False}\}. \end{aligned} \quad (4)$$

The optimal solutions (prototypes of crystal structures) of the auxiliary GDP must include all the meaningful optimal solutions of the original optimization problem. Note that since we are mainly interested in questions about whether there are feasible solutions satisfying all the constraints, the auxiliary GDP is essentially a feasibility problem whose objective function is identically zero. However, the numerical performance indicates that the volume of unit cell is convenient as the objective function to design crystal structures from random sampling. Since the GDP model has the concept that it is an extension of inorganic structural chemistry that emphasizes the importance of the closest packings of spheres, it is reasonable to choose the objective function as it is. Besides, since inorganic structural chemistry compiles the tables of atomic radii depending on the types of interatomic force to estimate interatomic distances by the sum of atomic radii [21–23,66–69], the GDP model constrains the minimum and maximum interatomic distances by the sum of minimum and maximum atomic radii, respectively, to find crystal structures as feasible solutions:

$$d_{ij}^{(G)} = r_i^{(G)} + r_j^{(G)}, \quad D_{ij}^{(G)} = R_i^{(G)} + R_j^{(G)}. \quad (5)$$

Finally, the Boolean constants $\Lambda_{ij}^{(G)}$ and $\Xi_{ij}^{(G)}$ in the logic propositions constrain the possible choices of G between atoms i and j . For example, if both atoms have positive or negative oxidation-state charges, an ionic bond is not creatable. Then, we can constrain the rule to create ionic bonds as

$$\Lambda_{ij}^{(I)} = \begin{cases} \text{True}, & q_i q_j < 0, \\ \text{False}, & \text{otherwise}, \end{cases} \quad (6)$$

where I corresponds to ionic bonds and q_i is the oxidation state of atom i .

Solid-state chemistry defines chemical bond to briefly describe the mechanism that stabilizes crystal structures. While there is no standard derivation of chemical bonds from quantum mechanics, the software LOBSTER [70–72] calculates the indicator using a simple formula to quantify chemical bonding properties consistent with chemical intuition. The theory explains that a covalent bond among atoms is formed by interfering wavefunctions consisting of their overlapping orbitals, while an ionic bond is formed by electrostatic interactions between charged ions [70–72]. While covalency and ionicity are often referred to as opposing bonding concepts, all the chemical bonds in chemical compounds mix both the two concepts. Therefore, it seems unreasonable for the GDP model to explicitly distinguish the types of interatomic forces as different types of interatomic distance constraints, but the constraint given in Eq. (1) can include the mixtures of the concepts. In fact, since Alkali metals form ionic bonds with weak covalency, both the ionic radius and coordination numbers are very flexible. On the other hand, since transition elements form ionic bonds with strong covalency, the ionic radius is nearly fixed and the stable coordination numbers are limited.

In general, GDPs can be reformulated as mixed-integer nonlinear programmings (MINLPs) [59–63,73]. To do so, bi-

nary variables $y_{ijT}^{(G)} \in \{0, 1\}$ are introduced to have one-to-one correspondences with the Boolean variables: $Y_{ijT}^{(G)} = \text{True}$ is equivalent to $y_{ijT}^{(G)} = 1$, while $Y_{ijT}^{(G)} = \text{False}$ is equivalent to $y_{ijT}^{(G)} = 0$. It is straightforward to convert logic propositions into algebraic forms using binary variables. The reformulation of Eq. (4) as an MINLP is given in Appendix A. Compared to general MINLPs, the remarkable feature of the auxiliary GDP given in Eq. (4) is that all the integer variables, which are equivalent to Boolean variables, are only used for indicating whether certain constraints on real variables are required to be satisfied or not.

III. METHODS

There are two major solution methods for GDP, namely, GDP branch and bound method and logic-based outer approximation [63]. The former directly branches on the disjunctions to generate nodes selecting enforced inequalities. This procedure is repeated until the tree search is exhausted; if a node is infeasible, then it is pruned. The advantage of this method is that it requires the evaluation of fewer nodes than the conventional branch and bound methods. On the other hand, the latter iteratively solves the subproblem and the master problem. The subproblem fixes all the Boolean variables to optimize the real variables. If the optimal solution of the subproblem is discovered, the collection of the optimized real variables and the fixed Boolean variables is the optimal solution of the original GDP. If not, the master problem is solved to find a different activation of Boolean variables from the past, where all the differentiable functions of real variables are linearized by their gradients to simultaneously optimize both the real and Boolean variables.

Now, we will begin to develop an efficient solution method for the auxiliary GDP given in Eq. (4) based on the two conventional solution methods. As the logic-based outer approximation, we create two subproblems, which are iteratively solved. One is the optimization problem of the real variables A and $\mathbf{x}_i$ for the fixed Boolean variables $Y_{ijT}^{(G)}$:

$$\begin{aligned} & \underset{A, \mathbf{x}_i}{\text{minimize}} && |\det A|, \\ & \text{subject to} && \bigvee_{G \in \mathcal{G}} \left[d_{ijT}^{(G)} \leq x_{ijT} \leq D_{ijT}^{(G)} \right], \\ & && d_{ijT}^{(G)} \equiv r_i^{(G)} + r_j^{(G)}, \\ & && D_{ijT}^{(G)} \equiv R_i^{(G)} + R_j^{(G)}, \\ & && x_{ijT} \equiv |\mathbf{x}_j + A\mathbf{T} - \mathbf{x}_i|, \\ & && \mathbf{0} \leq A^{-1} \mathbf{x}_i \leq \mathbf{1}, \\ & && A \in \mathbb{R}^{3 \times 3}, \quad \mathbf{x}_i \in \mathbb{R}^3. \end{aligned} \quad (7)$$

The other is the optimization problem of Boolean variables $Y_{ijT}^{(G)}$ for the fixed real variables A and $\mathbf{x}_i$, which is identical to

the integer programming:

$$\begin{aligned}
& \underset{b_i^{(G)}, Y_{ijT}^{(G)}}{\text{maximize}} && \sum_{i \in \mathcal{I}} \sum_{G \in \mathcal{B}} b_i^{(G)}, \\
& \text{subject to} && \text{ML}(\mathbf{Y}) = \text{True}, \\
& && \text{Exactly}(b_i^{(G)}, Y_{ijT}^{(G)}, \forall j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3) && (\forall i \in \mathcal{I}, \forall G \in \mathcal{G}), \\
& && \text{AtMost}(C_i^{(G)}, Y_{ijT}^{(G)}, \forall j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3) && (\forall i \in \mathcal{I}, \forall G \in \mathcal{G}), \\
& && \bigvee_{G \in \mathcal{G}} Y_{ijT}^{(G)} && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3), \\
& && \bigvee_{G \in \mathcal{G}} \left[|\mathbf{x}_j + A\mathbf{T} - \mathbf{x}_i| \leq (1 + \varepsilon)(R_i^{(G)} + R_j^{(G)}) \right] && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3), \\
& && \neg \Lambda_{ij}^{(G)} \Rightarrow \neg Y_{ijT}^{(G)} && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3, \forall G \in \mathcal{G}), \\
& && \Xi_{ij}^{(G)} \Rightarrow Y_{ijT}^{(G)} && (\forall i, j \in \mathcal{I}, \forall \mathbf{T} \in \mathbb{Z}^3, \forall G \in \mathcal{G}), \\
& && b_i^{(G)} \in \mathbb{N}, \quad Y_{ijT}^{(G)} \in \{\text{True}, \text{False}\}.
\end{aligned} \tag{8}$$

The set $\mathcal{B}$ is the subset of G corresponding to chemical bonds, and $0 \leq \varepsilon$ is an error. It is important to note that this subproblem has a similarity with the master problem of the logic-based outer approximation in the sense that Boolean variables $Y_{ijT}^{(G)}$ are optimized in preparation for minimizing the objective function of original GDP. If each atom forms as many chemical bonds with neighboring atoms as possible, the volume of the unit cell will be minimized after optimization of real variables A and $\mathbf{x}_i$. Then, we arrive at the objective function of the subproblem being the number of chemical bonds. It is also important to note that we remove both the lower bounds of feasible coordination numbers and interatomic distances, because adjacent atoms in a disordered structure can be either very close or very far apart. This relaxation is reasonable since this subproblem is aimed at providing an opportunity to make a different network of chemical bonds so that all the interatomic distances can be feasible. Finally, it is better to memorize the feasibilities of subsets of Boolean variables $Y_{ijT}^{(G)}$ as logic constraints:

$$\text{ML}(\mathbf{Y}) = \text{True}, \tag{9}$$

where $\mathbf{Y}$ is all the set of Boolean variables $Y_{ijT}^{(G)}$.

It is inevitable to solve the two subproblems iteratively many times when an initial structure is randomly generated. Since the feasible conditions of real variables change drastically depending on the Boolean variables and vice versa, well-controlled annealing causes a domino effect to largely transform an initial structure into an optimal solution. Such an annealing can be realized by applying the steepest-descent method to the subproblem given in Eq. (7): First, we introduce the linear penalty functions for the constraints on interatomic distances

$$\begin{aligned}
d_\sigma^{(G)} \leq x_\sigma &\Rightarrow U_{\min}^{(G)}(x_\sigma) \equiv \max[0, k_\downarrow(d_\sigma^{(G)} - x_\sigma)], \\
x_\sigma \leq D_\sigma^{(G)} &\Rightarrow U_{\max}^{(G)}(x_\sigma) \equiv \max[0, k_\uparrow(x_\sigma - D_\sigma^{(G)})],
\end{aligned}$$

with $k_\downarrow$ and $k_\uparrow$ being the constants, and σ being the abbreviated index of ijT . Accordingly, we have the relaxed

optimization problem for real variables with $0 < P$:

$$\text{minimize} \quad \sum_{\sigma, G} [U_{\min}^{(G)}(x_\sigma) + U_{\max}^{(G)}(x_\sigma)] + P|\det A|.$$

The displacements of real variables are calculated from the gradients of the relaxed objective function. This method is shown to be effective not only for optimizing packings of spheres accurately with small computation [32–34,74,75] but also for annealing structures without destroying feasible atomic environments too much by adjusting the maximum size of the displacements [74,75]. This annealing is aimed at creating an opportunity to find different spatial arrangements of atoms so that all the logic constraints on Boolean variables can be feasible. If it succeeds, the real variables are optimized by gradually decreasing the maximum size of the displacement to verify whether or not all the interatomic distances can be feasible.

The set of Boolean variables $\mathbf{Y}$ represents the graphs describing crystal structures with clear chemical meanings; for example, the subgraphs corresponding to ionic bonds can represent not only coordination polyhedra but also linkings of them, and similarly, the central cations of linked polyhedra are joined by the edge corresponding to electrostatic repulsions. Recall that each edge enforces the associated constraint on the interatomic distance. As detailed in the previous study [75], two polyhedra can generally be linked by sharing a common vertex, a common edge, or a common face, but not all of them can be feasible if the central atoms must be far apart. Similarly, even if the coordination number is feasible, the composition of the coordination polyhedron designated by the subgraph is not necessarily feasible.

A considerable amount of subgraphs, which correspond to coordination polyhedra or linking of them, are discovered in numerous graphs generated by the solution method. Since it is very futile to optimize real variables with an infeasible graph, it is advisable to add constraints on Boolean variables so that infeasible subgraphs are efficiently removed as if the GDP branch and bound method prunes infeasible nodes from the tree search. While the cutting-plane method for integer programming iteratively refines the feasible condition by

adding linear inequalities [76,77], the GDP model can easily make such constraints by memorizing infeasible subgraphs. Accordingly, the subproblem given in Eq. (8) includes the on-the-fly logic constraint given in Eq. (9), which consists of only the memories of feasible and infeasible subgraphs. Note that the feasibility of a subgraph can be verified by optimizing the corresponding substructure as discussed in Appendix B 4.

IV. RESULTS AND DISCUSSION

The software MARICI (*Mathematical Architecture for Realizing Inorganic Crystalline materials using mixed-Integer nonlinear programming*) [78] implements the solution method discussed above. The algorithm is detailed in Appendix B. MARICI can efficiently design prototypes of crystal structures based on only the atomic radii and feasible coordination numbers of each atom. For example, MARICI can design pyroxene and the four superstructures of the perovskite type in about tens of seconds, and besides, not only simple ternary oxides such as the normal spinel but also simple quaternary oxides such as $\text{BaDyFe}_4\text{O}_7$ in about one second by only using ten cores of Intel® Core™ Ultra 9 185H (2.30 GHz), 32 GB. It may be important to note that numerical performance indicates that the memories of the feasibilities of subgraphs increases the possibility of discovering optimal solutions by at least 10 times. In addition to the computational efficiency, MARICI can find a wide variety of crystal structures including not only ionic compounds (α -pyrochlore, quadruple perovskite, $\text{YBa}_2\text{Cu}_3\text{O}_7$, $\text{NaBe}_4\text{SbO}_7$, jarosite without hydrogen, etc.) but also Zintl phases (MgB_2 , ThSi_2 , AuBe_5 , etc.) whose crystal structures are composed of both the covalent and ionic bonds. Furthermore, as detailed in Appendix C 1, the number of optimal solutions of the auxiliary GDP is much smaller than that of the original optimization problem, which means that the GDP model successfully extracts only the meaningful optimal solutions of the original optimization problem. Finally, as detailed in Appendix C 2, an exhaustive search for quaternary oxides indicates that the GDP model can design many unknown prototypes of crystal structures whose existence in nature would not be surprising. Not only when designed prototypes have high symmetries but also when they belong to the space groups with a smaller occurrence in nature, there are many cases where they are similar to experimentally known crystal structures.

The GDP model cannot identify the synthesizable prototypes of crystal structures since it does not explicitly solve the minimization problem of Gibbs free energy. However, owing to the high-speed discoveries of prototypes, the model is useful to scan the coordination space as a preprocessor of *ab initio* simulation to evaluate formation enthalpies. While the model has no idea how to assign chemical elements to each atom site, the table of ionic radii depending on the oxidation state and the coordination number [21–23,66–69] may be useful to limit the possible choices. In general, ionic radius of each atom depends on the oxidation state and the coordination number as if inorganic structural chemistry interprets that a larger atom forms more chemical bonds so that it is in contact with as many neighboring atoms as possible [20–23].

Since MARICI generates initial structures randomly and after the structures are largely transformed to find an optimal solution, each “pseudoatom” in optimal structures tends to form as many chemical bonds as possible, as if a local packing fraction of atomic spheres is maximized. Therefore, it is difficult for MARICI to deal with materials for which one chemical element adopts two different coordination polyhedra. However, while inverse spinel such as Fe_3O_4 is difficult to find, normal spinel such as Co_3O_4 is very easy to find, because different “pseudoatoms” can be defined depending on the oxidation states of chemical elements. Here, it is important to note that once a prototype of a crystal structure is designed, only the remained problem is how to assign chemical elements to each atom site. Besides, in general, if two atoms are the same chemical element with the same oxidation state, they tend to realize the same atomic environment: It is widely known that smaller number of Wyckoffs are preferred. Therefore, the difficulty of finding crystal structures for which one element adopts two different coordination polyhedra is not a critical problem. Finally, it is also important to note that the suitable ionic radii for four or six coordination are largely different. Assume that closest packing of oxide ions where the ionic radius of oxide ion is 1.4 Å. In that case, the radii of cations placed in tetrahedral and octahedral sites are about 0.32 and 0.6 Å at maximum, respectively. Therefore, if the atomic radius of a “pseudoatom” is about 0.6 Å, tetrahedral coordination is not stable (the coordination polyhedra is often largely distorted).

V. CONCLUSION

This article clarifies that the typical form of GDP can naturally formulate the empirical rules of inorganic structural chemistry, so that the crystal structures of not only ionic compounds but also Zintl phases can be discovered as local optima with small computations. The solution method for the GDP consists of two subproblems so that continuous and integer variables are iteratively optimized. Since the optimization of real variables with an infeasible set of integer variables is very futile, infeasible subsets of integer variables are memorized as logical constraints on integer variables, as if the cutting-plane method refines the feasible condition by adding linear inequalities. The efficiency of the solution method indicates that introducing integer variables accelerates visiting every local optimum by extracting the “combinatorial backbone” of an original continuous optimization problem, which enables discovering only the meaningful optimal solutions of the original problem. While many machine-learning techniques have achieved great success [79,80], we can expect that symmetry-oriented crystal structure prediction [81–83] with the GDP model will discover the majority of known and unknown prototypes of crystal structures based on only the atomic radii and feasible coordination numbers of each atom. This preprocessing before *ab initio* simulations will broaden the applicability of computational discovery of chemical compounds owing to the efficiency of mathematical crystal chemistry.

ACKNOWLEDGMENTS

This work used computational resources of supercomputer Fugaku provided by the RIKEN Center for Computational Science (Projects No. hp250425 and No. hp260027). This

work also used Fujitsu Server PRIMERGY CX2550 M7 (Miyabi-C) at the Information Technology Center, The University of Tokyo, and the facilities (supercomputers Ohtaka and Kugui) of the Supercomputer Center, The Institute for Solid State Physics, The University of Tokyo. This work was supported by JST ASPIRE (Grant No. JP-MJAP2314), JSPS KAKENHI (Grant No. 26H02231), The Kazuchika Okura Memorial Foundation, and Nippon Sheet Glass Foundation for Materials Science and Engineering. The author is indebted to many researchers including Prof. Richard Dronskowski, Prof. Yoshihiro Kanno, Prof. Taisuke Ozaki, Prof. Jun-ichi Yamaura, and Prof. Yoshihiko Okamoto. The author has to thank, in particular, Dr. Jun Takahashi who not only discussed NP-hardness but also gave me suggestions on the draft.

DATA AVAILABILITY

The software MARICI (*Mathematical Architecture for Realizing Inorganic Crystalline materials using mixed-Integer nonlinear programming*) is available in Ref. [78]. MARICI ver. 2026.0 is distributed under the MIT License. This software can efficiently not only design prototypes of crystal structures but also analyze a large dataset of crystal structures. Since a graphical user interface is distributed for Windows users, everyone can easily begin to predict crystal structures. It is thus straightforward to confirm that the GDP model can find a wide variety of crystal structures of not only ionic compounds but also Zintl phases with remarkable computational speed. The data of designed prototypes of crystal structures discussed in Appendixes is also available in Ref. [85].

APPENDIX A: REFORMULATION AS MIXED-INTEGERS NONLINEAR PROGRAMMING

A GDP [59–65] can easily be reformulated as an MINLP [59,61–63]. To do this, binary variables $y_{ijT}^{(G)} \in \{0, 1\}$, $\xi_{ijT}^{(G)} \in \{0, 1\}$, and $\lambda_{ijT}^{(G)} \in \{0, 1\}$ are introduced to have one-to-one

correspondences with the Boolean variables $Y_{ijT}^{(G)}$, $\Xi_{ijT}^{(G)}$, and $\Lambda_{ijT}^{(G)}$:

$$\begin{aligned} Y_{ijT}^{(G)} = \text{True} &\iff y_{ijT}^{(G)} = 1, \\ Y_{ijT}^{(G)} = \text{False} &\iff y_{ijT}^{(G)} = 0, \\ \Xi_{ijT}^{(G)} = \text{True} &\iff \xi_{ijT}^{(G)} = 1, \\ \Xi_{ijT}^{(G)} = \text{False} &\iff \xi_{ijT}^{(G)} = 0, \\ \Lambda_{ijT}^{(G)} = \text{True} &\iff \lambda_{ijT}^{(G)} = 1, \\ \Lambda_{ijT}^{(G)} = \text{False} &\iff \lambda_{ijT}^{(G)} = 0. \end{aligned}$$

It is straightforward to convert logic propositions into algebraic forms using binary variables, for example, the logic constraint

$$\bigvee_{G \in \mathcal{G}} Y_{ijT}^{(G)} \quad (\text{A1})$$

can be converted into

$$\sum_{G \in \mathcal{G}} y_{ijT}^{(G)} = 1. \quad (\text{A2})$$

Similarly, there are several ways such as big- M or hull reformulation [63] to reformulate the disjunctions:

$$\bigvee_{G \in \mathcal{G}} \left[d_{ijT}^{(G)} \leq |x_j + AT - x_i| \leq D_{ijT}^{(G)} \right]. \quad (\text{A3})$$

Among them, the big- M approach converts the disjunctions into

$$y_{ijT}^{(G)} d_{ijT}^{(G)} \leq |x_j + AT - x_i|, \quad (\text{A4})$$

$$|x_j + AT - x_i| \leq D_{ijT}^{(G)} + M(1 - y_{ijT}^{(G)}), \quad (\text{A5})$$

where M is a huge constant. Accordingly, one of the reformulations as an MINLP from the auxiliary GDP given in Eq. (4) is as follows:

$$\begin{aligned} &\underset{A, x_i, y_{ijT}^{(G)}}{\text{minimize}} && |\det A|, \\ &\text{subject to} && c_i^{(G)} \leq \sum_{jT} y_{ijT}^{(G)} \leq C_i^{(G)} && (\forall i \in \mathcal{I}, \forall G \in \mathcal{G}), \\ & && \sum_{G \in \mathcal{G}} y_{ijT}^{(G)} = 1 && (\forall i, j \in \mathcal{I}, \forall T \in \mathbb{Z}^3), \\ & && y_{ijT}^{(G)} d_{ijT}^{(G)} \leq |x_j + AT - x_i| \leq D_{ijT}^{(G)} + M(1 - y_{ijT}^{(G)}) && (\forall i, j \in \mathcal{I}, \forall T \in \mathbb{Z}^3), \\ & && \xi_{ij}^{(G)} \leq y_{ijT}^{(G)} \leq \lambda_{ij}^{(G)} && (\forall i, j \in \mathcal{I}, \forall T \in \mathbb{Z}^3, \forall G \in \mathcal{G}), \\ & && d_{ijT}^{(G)} \equiv r_i^{(G)} + r_j^{(G)}, \quad D_{ijT}^{(G)} \equiv R_i^{(G)} + R_j^{(G)}, \quad \mathbf{0} \leq A^{-1}x_i \leq \mathbf{1}, \\ & && A \in \mathbb{R}^{3 \times 3}, \quad x_i \in \mathbb{R}^3, \quad y_{ijT}^{(G)} \in \{0, 1\}. \end{aligned} \quad (\text{A6})$$

APPENDIX B: ALGORITHM FOR SOLUTION METHOD

This Appendix details the algorithm executing the solution method for the auxiliary GDP given in Eq. (4). The

algorithm is implemented in the software MARICI [78], which is distributed under the MIT License.

The algorithm discussed in this Appendix is almost identical to that discussed in the previous studies [74,75]. However,

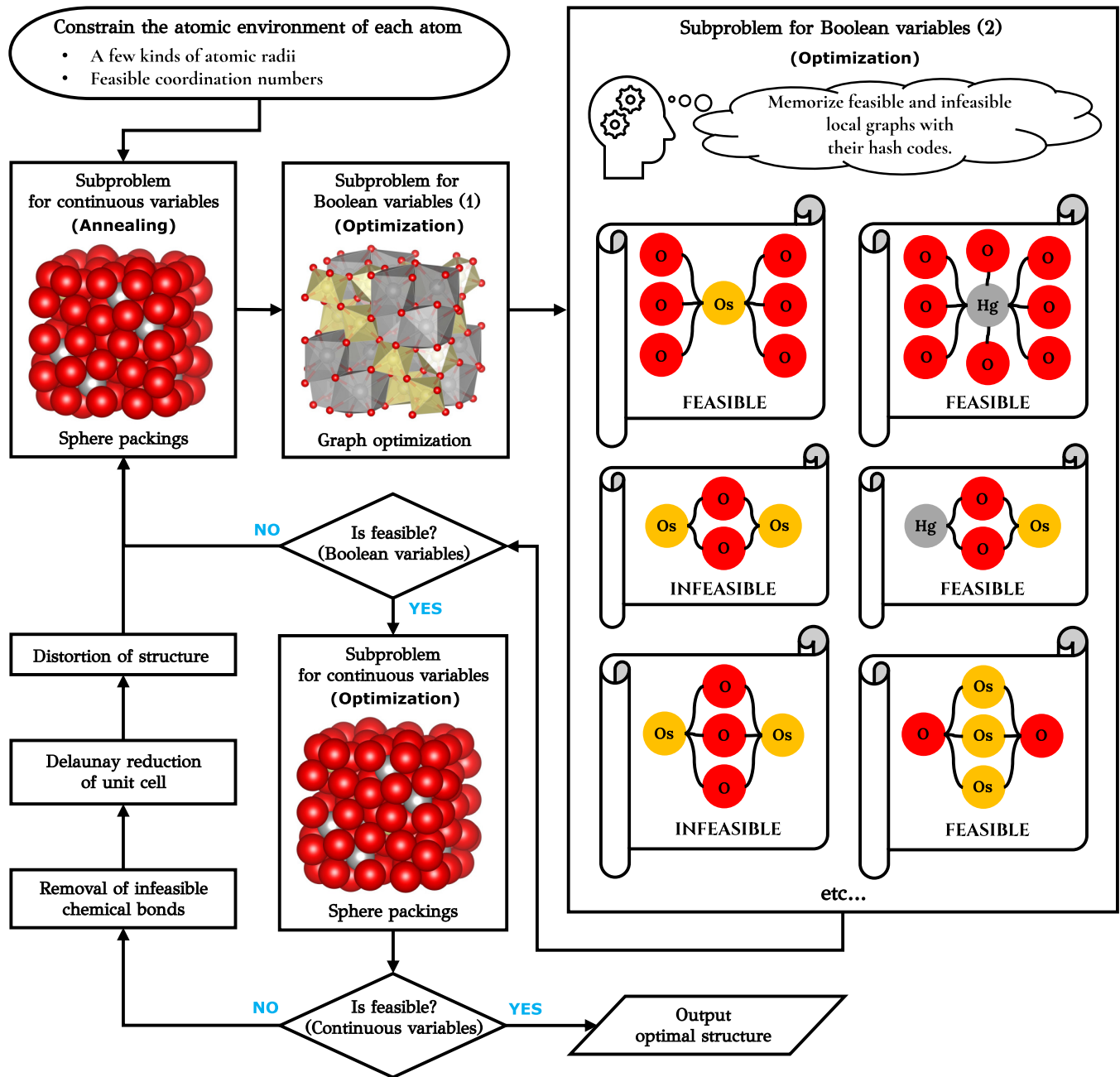


FIG. 1. The flowchart of the algorithm executing the solution method for the auxiliary GDP given in Eq. (4). Crystal structures are drawn by VESTA [84].

since the source code for MARICI is developed from scratch to resolve the technical problems, the computational efficiency is significantly improved. Besides, it is found that the on-the-fly constraint consisting of the memories of the feasible and infeasible subgraphs also improves the efficiency for finding the optimal solutions. The flowchart of the algorithm is shown in Fig. 1. Two subproblems are iteratively solved to optimize the continuous and Boolean variables with fixing the other.

1. Input data

When we choose the chemical compositions for crystal structure prediction, we must assign an oxidation state to each

atom (we call a chemical element with an oxidation state as “pseudoatom”). This is because oxidation states are used for judging whether or not ionic bonds are creatable between atoms as

$$\Lambda_{ij}^{(I)} = \begin{cases} \text{True}, & q_i q_j < 0, \\ \text{False}, & \text{otherwise.} \end{cases} \quad (\text{B1})$$

The software MARICI supports the five kinds of G : The constraints I , C , E , I_e , and C_e correspond to ionic bonds, covalent bonds, electrostatic repulsions, ionic exclusive repulsions,

and covalent exclusive repulsions, respectively. The remained “creatabilities” $\Lambda_{ij}^{(G)}$ are defined as follows:

$$\Lambda_{ij}^{(C)} = \begin{cases} \text{False}, & q_i q_j < 0, \\ \text{True}, & \text{otherwise}, \end{cases} \quad (\text{B2})$$

$$\Lambda_{ij}^{(E)} = \begin{cases} \text{True}, & q_i q_j > 0, \\ \text{False}, & \text{otherwise}, \end{cases} \quad (\text{B3})$$

$$\Lambda_{ij}^{(I_e)} = \begin{cases} \text{True}, & q_i q_j < 0, \\ \text{False}, & \text{otherwise}, \end{cases} \quad (\text{B4})$$

$$\Lambda_{ij}^{(C_e)} = \text{True}. \quad (\text{B5})$$

Recall that the feasible atomic environments of each atom are determined by only the atomic radii and feasible coordination numbers as given in Eq. (1). Since MARICI sets the exclusive radii as

$$\begin{aligned} r^{(I_e)} &= \chi R^{(I)}, \\ r^{(C_e)} &= \chi R^{(C)}, \\ R^{(I_e)} &= R^{(C_e)} = \infty, \end{aligned} \quad (\text{B6})$$

with $\chi = 1.3$, the indispensable parameters to determine the feasible atomic environments of each pseudoatom are given by

$$(r^{(C)}, R^{(C)}, c^{(C)}, C^{(C)}, r^{(I)}, R^{(I)}, c^{(I)}, C^{(I)}, R^{(E)}).$$

If you search only ionic crystal structures, the necessary parameters are as follows:

$$(r^{(I)}, R^{(I)}, c^{(I)}, C^{(I)}, R^{(E)}).$$

2. Initial structure generation

An initial structure is randomly generated as follows: Let $(a, b, c, \alpha, \beta, \gamma)$ be the lattice parameters, and they are initialized randomly under the condition of

$$1 \leq b, c \leq 3, \quad \frac{1}{3}\pi \leq \alpha, \beta, \gamma \leq \frac{2}{3}\pi, \quad (\text{B7})$$

where a is set to be 1. Next, lattice vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ are scaled to satisfy the condition of

$$\frac{1}{|\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)|} \left[\sum_{i=1}^M \frac{4}{3} \pi (R_i^{(I)})^3 \right] = 0.7. \quad (\text{B8})$$

Finally, all the atoms are randomly distributed in the unit cell.

3. Subproblem for continuous variables

The solution method consists of the two subproblems. One of them is the optimization problem for continuous variables, which is given in Eq. (7). The GDP can be reformulated as an

MINLP:

$$\begin{aligned} &\text{minimize}_{A, \mathbf{x}_i} |\det A|, \\ &\text{subject to} \quad y_{ijT}^{(G)} d_{ijT}^{(G)} \leq x_{ijT}, \\ &\quad x_{ijT} \leq D_{ijT}^{(G)} + M(1 - y_{ijT}^{(G)}), \\ &\quad d_{ijT}^{(G)} \equiv r_i^{(G)} + r_j^{(G)}, \\ &\quad D_{ijT}^{(G)} \equiv R_i^{(G)} + R_j^{(G)}, \\ &\quad x_{ijT} \equiv |\mathbf{x}_j + A\mathbf{T} - \mathbf{x}_i|, \\ &\quad \mathbf{0} \leq A^{-1}\mathbf{x}_i \leq \mathbf{1}, \\ &\quad A \in \mathbb{R}^{3 \times 3}, \quad \mathbf{x}_i \in \mathbb{R}^3. \end{aligned} \quad (\text{B9})$$

The solution method for this problem is identical to that discussed in the previous study [74,75], but it is summarized to present this article in a self-contained manner. The inequality constraints are approximated by the hard-spherical penalty potential as

$$\begin{aligned} d_\sigma^{(G)} \leq x_\sigma &\Rightarrow U_{\min}^{(G)}(x_\sigma) \equiv \max[0, k_\downarrow (d_\sigma^{(G)} - x_\sigma)], \\ x_\sigma \leq D_\sigma^{(G)} &\Rightarrow U_{\max}^{(G)}(x_\sigma) \equiv \max[0, k_\uparrow (x_\sigma - D_\sigma^{(G)})], \end{aligned}$$

where $k_\downarrow$ and $k_\uparrow$ are common constants for repulsive and attractive forces, respectively, and σ is the abbreviated index of ijT . Accordingly, the objective function of the subproblem is approximated as

$$\text{minimize} \quad H(\mathbf{X}, A), \quad (\text{B10})$$

where $H(\mathbf{X}, A)$ is defined as

$$H(\mathbf{X}, A) \equiv \sum_{\sigma, G} [U_{\min}^{(G)}(x_\sigma) + U_{\max}^{(G)}(x_\sigma)] + P|\det A|,$$

with $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, N being the number of atoms per unit cell, and P being the pressure. The continuous variables are optimized by steepest-descent method as follows: Let $\Delta \mathbf{x}_i$ and $\Delta \mathbf{a}_i$ be the displacements of $\mathbf{x}_i$ and $\mathbf{a}_i$. They are calculated by

$$\Delta \mathbf{x}_i = -\xi_i \frac{\partial H(\mathbf{X}, A)}{\partial \mathbf{x}_i}, \quad \Delta \mathbf{a}_i = -\zeta_i \frac{\partial H(\mathbf{X}, A)}{\partial \mathbf{a}_i}, \quad (\text{B11})$$

where the constants ξ_i and ζ_i are scaled to satisfy the condition

$$|\Delta \mathbf{x}_i| \leq \Delta x_s^{(\max)}, \quad |\Delta \mathbf{a}_i| \leq \gamma \Delta x_s^{(\max)}, \quad (\text{B12})$$

with $\Delta x_s^{(\max)}$ being the maximum displacement of atoms in sth optimization step and γ being a constant. $\Delta x_s^{(\max)}$ is calculated as

$$\Delta x_s^{(\max)} = \Delta x_0^{(\max)} \left(\frac{\Delta x_s^{(\max)}}{\Delta x_0^{(\max)}} \right)^{\frac{s}{S}}, \quad (\text{B13})$$

where S is the maximum steps of the steepest-descent method. In this study, γ is given by

$$\gamma = 0.02, \quad (\text{B14})$$

and the force constants are set to be

$$P = 1.0, \quad k_\uparrow = 30.0, \quad k_\downarrow = -100.0. \quad (\text{B15})$$

When the continuous variables are annealed by this subproblem, the parameters are given by

$$S = 25, \quad \Delta x_0^{(\max)} = \Delta x_S^{(\max)} = 0.5. \quad (\text{B16})$$

It is important to note that in the annealing process, the tracers of every pair of atoms are updated per 100 steps, the unit cell is refined per 200 steps by using the software SPGLIB [87], and a structure is randomly distorted largely per 2500 steps.

When the continuous variables are optimized by this subproblem, we make two steps composed of local and precise structural optimization. In the former, the parameters are given by

$$S = 2000, \quad \Delta x_0^{(\max)} = 0.3, \quad \Delta x_S^{(\max)} = 0.05, \quad (\text{B17})$$

and in the latter, they are given by

$$S = 4000, \quad \Delta x_0^{(\max)} = 0.1, \quad \Delta x_S^{(\max)} = 0.005. \quad (\text{B18})$$

Note that in the local optimization, if a structure is failed to reach an optimal solution after the first 2000 steps, the structure is optimized again after initializing $\Delta x_S^{(\max)}$, because in some cases, all the interatomic distances cannot converge into the feasible values at once.

When MARICI appraises the feasibility of the set of continuous variables, MARICI allows the small error of the feasibilities of interatomic distances as

$$(1 - \varepsilon)d_{ijT}^{(G)} \leq |\mathbf{x}_j + \mathbf{AT} - \mathbf{x}_i| \leq (1 + \varepsilon)D_{ijT}^{(G)}, \quad (\text{B19})$$

where $0 \leq \varepsilon$ is a small value. When MARICI appraises the inequality constraints after annealing, local optimization, and precise optimization, ε is set to be

$$\varepsilon = 1.0, \quad \varepsilon = 0.1, \quad \varepsilon = 0.05, \quad (\text{B20})$$

respectively.

After annealing, if all the logic constraints on Boolean variables are feasible, MARICI checks the feasibilities of all the interatomic distances with $\varepsilon = 1.0$. If they are optimal, MARICI starts a local optimization. After that, if all the interatomic distances are optimal with $\varepsilon = 0.1$, MARICI begins a precise optimization. After that, if all the interatomic distances are optimal with $\varepsilon = 0.05$, MARICI regards the structure as a local optimum of the auxiliary GDP given in Eq. (4).

To reduce the computational cost, the criteria to track a pair of atoms is given as follows:

$$0 < q_i q_j \Rightarrow \tau(R_i^{(I)} + R_j^{(I)}), \quad (\text{B21})$$

$$0 = q_i q_j \Rightarrow \tau(R_i^{(C)} + R_j^{(C)}), \quad (\text{B22})$$

$$0 > q_i q_j \Rightarrow \tau(r_i^{(E)} + r_j^{(E)}), \quad (\text{B23})$$

where τ is set to be 4.0.

4. Subproblem for Boolean variables

The other subproblem is the optimization problem for Boolean variables given in Eq. (8). The GDP can also be reformulated as an MINLP:

$$\begin{aligned} &\underset{y_{ijT}^{(G)}}{\text{maximize}} && \sum_i \sum_{jT} \sum_{G \in \mathcal{B}} y_{ijT}^{(G)}, && (\text{B24}) \\ &\text{subject to} && \text{ML}(\mathbf{y}) = \text{True}, && (\text{B25}) \end{aligned}$$

$$\sum_{jT} y_{ijT}^{(G)} \leq C_i^{(G)} \quad (\forall i, j \in \mathcal{I}, \forall T \in \mathbb{Z}^3), \quad (\text{B26})$$

$$\sum_{G \in \mathcal{G}} y_{ijT}^{(G)} = 1 \quad (\forall i, j \in \mathcal{I}, \forall T \in \mathbb{Z}^3), \quad (\text{B27})$$

$$|\mathbf{x}_j + \mathbf{AT} - \mathbf{x}_i| \leq (1 + \varepsilon)(R_i^{(G)} + R_j^{(G)}) + M(1 - y_{ijT}^{(G)}) \quad (\forall i, j \in \mathcal{I}, \forall T \in \mathbb{Z}^3), \quad (\text{B28})$$

$$\xi_{ij}^{(G)} \leq y_{ijT}^{(G)} \leq \lambda_{ij}^{(G)} \quad (\forall i, j \in \mathcal{I}, \forall T \in \mathbb{Z}^3, \forall G \in \mathcal{G}), \quad (\text{B29})$$

$$y_{ijT}^{(G)} \in \{0, 1\}. \quad (\text{B30})$$

Equation (B25) is the logic constraint consisting of the memories of feasible and infeasible subgraphs, where $\mathbf{y}$ is all the set of binary variables $y_{ijT}^{(G)}$. Since this integer programming is essentially the logic-based optimization under the logic constraints, it is easy to solve as follows: First, if a constraint G would be enforced between a pair of atoms, the interatomic distance must be feasible with respect to G as written in Eq. (B28). It is important to note that the maximum atomic radius of each atom is finite only if G corresponds to a chemical bond. Therefore, the inequality constraints of Eq. (B28) implicitly limit the creatable chemical bonds. Second, to maximize the total number of chemical bonds, each

atom should create as many creatable chemical bonds with neighboring atoms as possible. However, each atom must follow the constraint on the maximum coordination numbers given in Eq. (B26) and the other logic constraints given in Eqs. (B27) and (B29). If the number of creatable chemical bonds is more than the maximum coordination number, MARICI selects shorter chemical bonds as many as possible. Finally, inappropriate chemical bonds are removed to satisfy the constraint of Eq. (B25).

Two polyhedra can generally be linked by sharing a common vertex, a common edge, or a common face, but not all of them are feasible if the central atoms must be far

TABLE I. The definitions of pseudoatoms.

Symbol	Feasible oxidation states	Feasible atomic environment ($r^{(C)}$, $R^{(C)}$, $c^{(C)}$, $C^{(C)}$, $r^{(I)}$, $R^{(I)}$, $c^{(I)}$, $C^{(I)}$, $R^{(E)}$)
O	2−	(0, 0, 0, 0, 1.40, 1.40, 0, ∞, 1.40)
Aa	1−	(1.1, 1.3, 2, 2, 1.80, 1.80, 0, ∞, 1.80)
Ab	2−	(1.1, 1.3, 3, 3, 1.80, 1.80, 0, ∞, 1.80)
Ea	1+, 2+, 3+, 4+, 5+, 6+	(0, 0, 0, 0, 0.30, 0.35, 4, 4, 1.40)
Eb	1+, 2+, 3+, 4+, 5+, 6+	(0, 0, 0, 0, 0.30, 0.35, 4, 4, 1.60)
Ec	2+	(0, 0, 0, 0, 0.60, 0.60, 4, 4, 1.40)
Ed	2+	(0, 0, 0, 0, 0.60, 0.60, 5, 5, 1.40)
Ee	2+	(0, 0, 0, 0, 0.60, 0.60, 6, 6, 0.90)
Ef	1+, 2+, 3+, 4+, 5+, 6+	(0, 0, 0, 0, 0.60, 0.60, 6, 6, 1.40)
Eg	1+, 2+, 3+, 4+, 5+, 6+	(0, 0, 0, 0, 0.60, 0.60, 6, 6, 1.60)
Eh	4+	(0, 0, 0, 0, 0.60, 0.70, 6, 6, 1.40)
Ei	3+	(0, 0, 0, 0, 0.80, 1.00, 8, 8, 1.60)
Ej	1+	(0, 0, 0, 0, 0.80, 1.30, 8, 8, 1.80)
Ek	2+	(0, 0, 0, 0, 1.20, 1.40, 10, 10, 1.60)
El	1+, 2+, 3+	(0, 0, 0, 0, 1.40, 1.40, 12, 12, 1.80)
Em	1+, 2+, 3+	(0, 0, 0, 0, 1.40, 1.60, 12, 12, 1.80)
En	2+	(0, 0, 0, 0, 1.40, 1.60, 12, 7, 1.80)
Eo	2+	(0, 0, 0, 0, 1.40, 1.80, 12, 12, 1.80)

apart. Similarly, even if the coordination number is feasible, a coordination polyhedron consisting of a different chemical composition is not necessarily feasible. Since Eq. (B25), which is not easy to formulate, remembers not only all the infeasible coordination polyhedra of each atom but also all the infeasible linking of coordination polyhedra, we can easily find them all in a structure. An infeasible coordination polyhedra or linking of them must be made feasible by randomly removing chemical bonds that cause the infeasibility.

Technically, MARICI memorizes all the feasible and infeasible subgraphs corresponding to coordination polyhedra of each atom or linking of coordination polyhedra. MARICI uses the hash table to remember them all: The key is a subgraph, and the value is its feasibility. To define the equality of subgraphs, atoms are sorted with respect to such as atomic numbers and oxidation states when creating a subgraph. The hash code can be generated from the collections of which atoms are linked with each atom. Each time MARICI finds a subgraph that is not memorized, MARICI extracts the substructure corresponding to the subgraph to judge whether it is feasible or not. Accordingly, the on-the-fly constraint given in Eq. (B25) is regularly updated through crystal structure prediction. The structural optimization for a substructure is executed by the same method discussed in Appendix B 3. The parameters for local optimization are given by

$$S = 2000, \quad \Delta x_0^{(\max)} = 0.2, \quad \Delta x_S^{(\max)} = 0.1, \quad (\text{B31})$$

while those for precise optimization are given by

$$S = 10\,000, \quad \Delta x_0^{(\max)} = 0.1, \quad \Delta x_S^{(\max)} = 0.001. \quad (\text{B32})$$

The feasible error ε is set to be 0.01. Note that a random fluctuation is applied to the substructure before structural optimization, and the operation is repeated 5–10 times until a

feasible substructure is discovered. If it fails, the substructure is regarded as infeasible.

APPENDIX C: CRYSTAL STRUCTURE PREDICTION

1. Prediction of experimentally known crystal structures

Table I lists the pseudoatoms defined by the feasible oxidation states, the feasible coordination numbers, and the atomic radii. Note that both the feasible coordination numbers and atomic radii corresponding to covalent bonds are meaningless when we search ionic crystal structures. Table II lists the results of the example prediction for the experimentally known crystal structures, which shows that mathematical crystal chemistry can find a wide variety of crystal structures. It is important to emphasize that MARICI can design the crystal structures of Zintl phases consisting of both the covalent and ionic bonds. Besides, it is also important to note that the number of local optima is much smaller than that of the original continuous optimization problem for the Gibbs free energy: In some cases, an experimentally known crystal structure is obtained as a unique optimal solution. These results indicate that the optimal solutions of the GDP model represent only the meaningful optimal solutions, which look promising as stable crystal structures, of the original optimization problem.

There are two optimal solutions for $\text{EcEd}_2\text{EiEk}_2\text{O}_7$, but both of them correspond to the $\text{YBa}_2\text{Cu}_3\text{O}_7$ structure; one of the optimal solutions is a distorted $\text{YBa}_2\text{Cu}_3\text{O}_7$ structure. While Ec and Ed correspond to the coppers whose coordination numbers are four and five, respectively, the coordination polyhedra are very similar; Ec creates a square and Ed creates a square pyramid. Then, $\text{YBa}_2\text{Cu}_3\text{O}_7$ structure should be discovered as the optimal solution for $\text{Ec}_3\text{EiEk}_2\text{O}_7$, but it is impossible for MARICI. This is due to the uniform definition of the ionic exclusive radii $r^{(L)}$ as given in Eq. (B6): If the

TABLE II. The number of randomly generated initial structures depending on the chemical compositions are listed. The number of discovered optimal solutions and the discovered experimentally known crystal structures are also listed.

Composition	Number of initial structures	Number of discovered solutions	Experimental crystal structures
Ea ₂ Ef ₄ O ₈	1×10^3	13	Spinel
Ea ₄ Ef ₈ O ₁₆	1×10^5	340	Spinel structure β -Mg ₂ SiO ₄
Ee ₄ El ₄ O ₁₂	5×10^3	4	Perovskite BaNiO ₃ BaMnO ₃ Cs ₂ NaCrO ₆
Eg ₄ Ej ₄ O ₁₄	3×10^4	1	α -pyrochlore
Eh ₄ Ei ₄ O ₁₂	1×10^5	2	Pyroxene
Ec ₃ Ef ₄ ElO ₁₂	1×10^5	31	Quadruple perovskite
EaE ₄ ElO ₇	1×10^5	1	YBaFe ₄ O ₇
Ea ₂ Ef ₈ El ₂ O ₁₄	3×10^5	2	YBaFe ₄ O ₇ NaBe ₄ SbO ₇
Eb ₂ Eg ₃ ElO ₁₄	2×10^5	22	Jarosite without hydrogen
EcEd ₂ EiEk ₂ O ₇	1×10^5	2	YBa ₂ Cu ₃ O ₇
En ₆ Aa ₆	5×10^3	5	CaSi
EO ₄ Ab ₈	2×10^4	15	MgB ₂ ThSi ₂ AuBe ₅

exclusive radii could be arranged depending on each pseudoatom, YBa₂Cu₃O₇ structure would be discovered as the optimal solution for Ec₃EiEk₂O₇.

MARICI achieves high computational efficiency for designing prototypes of crystal structures; for example, not only simple ternary oxides such as normal spinel but also simple quaternary oxides such as YBaFe₄O₇ are discovered in about one second by just using ten cores of Intel® Core™ Ultra 9 185H (2.30 GHz), 32 GB. Besides, all the optimal solution of Ee₄El₄O₁₂ corresponding to the experimentally known crystal structures, which are Perovskite, BaNiO₃, BaMnO₃, and Cs₂NaCrO₆ structures, are discovered in tens of seconds. Furthermore, α -pyrochlore and pyroxene are discovered in less than 30–60 s, where the flexible ionic radius of Ej in the structure of α -pyrochlore demands larger computational cost. The quadruple perovskite and YBa₂Cu₃O₇ structures are difficult to find due to the square coordinations of Ec: The ionic radius of Ec is large enough to form octahedra as Ef [74,75], but copper can create only four ionic bonds. It may be important to note that the number of optimal solutions for discovering the quadruple perovskite is much larger than that in the previous study [75] due to larger ionic radius of El. Finally, it seems that crystal structure prediction of Zintl phases generally demands larger computational costs than that of ionic compounds.

2. Exhaustive search for crystal structures

This work shortly reports the result of exhaustive search for crystal structures. More detailed results will be given in future works. We focus on the two chemical systems: One is Ea_kEf_l^qEl_m^rO_n²⁻ and the other is Eb_kEg_l^qEm_m^rO_n²⁻, where the

feasible chemical compositions are constrained as follows:

$$\begin{aligned}
 1 &\leq p, q \leq 6, \\
 1 &\leq r \leq 3, \\
 pk + ql + rm - 2n &\in \{0, 1\}, \\
 1 &\leq k + l + m + n \leq 25, \\
 \frac{1}{7} &\leq \frac{l}{k}, \frac{m}{k}, \frac{m}{l} \leq 7.
 \end{aligned} \tag{C1}$$

Accordingly, we obtain 3055 kinds of chemical compositions per each chemical system. The numbers of randomly generated initial structures are listed in Table III. If we remove the optimal solutions whose space group numbers are between 1 and 9, the numbers of the discovered optimal solutions of Ea_kEf_l^qEl_m^rO_n²⁻ and Eb_kEg_l^qEm_m^rO_n²⁻ are 4227 and 689, respectively.

Among the 230 kinds of space groups, some of them are not commonly realized by experimentally known crystal structures, while some of the others such as $Fm\bar{3}m$ (No. 225) are very popular. On the other hand, through this exhaustive

TABLE III. The number of randomly generated initial structures depending on the number of atoms per unit cell.

Number of atoms per unit cell	Number of initial structure
1–10	1×10^4
11–15	5×10^4
16–20	1×10^5
21–25	2×10^5

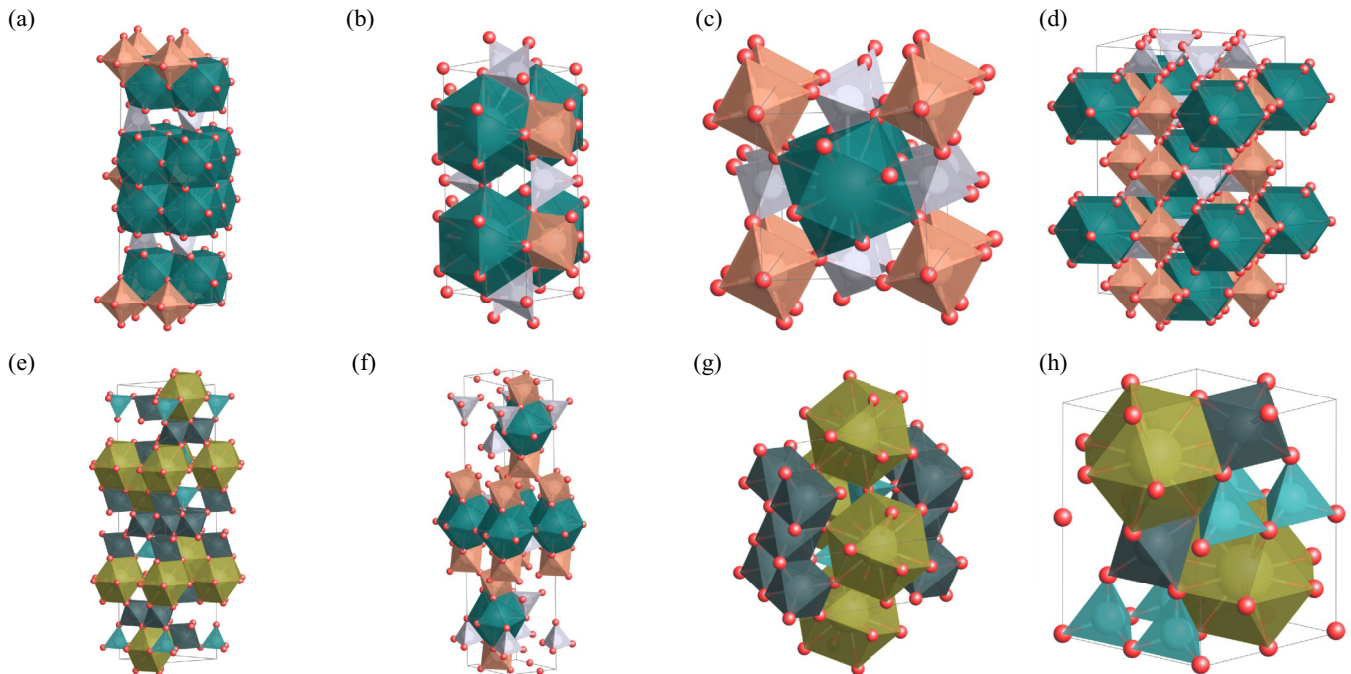


FIG. 2. The prototypes of crystal structures designed by MARICI. Crystal structures are drawn by OpenMX Viewer [86]. (a) The optimal solution of $\text{EbEgEm}_2\text{O}_8$. It belongs to the space group of $Ccc2$ (No. 37). (b) The optimal solution of $\text{Eb}_2\text{EgEmO}_7$. It belongs to the space group of $Pccm$ (No. 49). (c) The optimal solution of $\text{Eb}_2\text{EgEmO}_9$. It belongs to the space group of $P4$ (No. 75). (d) The optimal solution of $\text{Eb}_4\text{Eg}_2\text{EmO}_{18}$. It belongs to the space group of $I4$ (No. 79). (e) The optimal solution of $\text{Ea}_2\text{Ef}_5\text{El}_2\text{O}_{14}$. It belongs to the space group of $R\bar{3}m$ (No. 166). It corresponds to the crystal structure of $\text{Ba}_2\text{Cr}_7\text{O}_{14}$. (f) The optimal solution of $\text{Eb}_2\text{Eg}_2\text{EmO}_{11}$. It belongs to the space group of $R32$ (No. 155). It does not correspond to any experimentally known crystal structures, but it is similar to the crystal structure of $\text{BaNb}_2\text{V}_2\text{O}_{11}$. (g) The optimal solution of $\text{EaEf}_2\text{El}_3\text{O}_{11}$. It belongs to the space group of $P\bar{3}m1$ (No. 164). It does not correspond to any experimentally known crystal structures, but it is similar to the crystal structure of $\text{BaV}_2\text{Zn}_3\text{O}_{10}\text{H}_2$. (h) The optimal solution of $\text{Ea}_3\text{EfElO}_7$. It belongs to the space group of $P6_3mc$ (No. 186). It does not correspond to any experimentally known crystal structures, but it is similar to the crystal structure of $\text{NaBe}_4\text{SbO}_7$, which is the optimal solution of $\text{Ea}_2\text{Ef}_8\text{El}_2\text{O}_{14}$.

search for crystal structures, MARICI finds many prototypes of crystal structures belonging the space groups with a smaller occurrence in nature: Figures 2(a)–2(d) show the optimal solutions of $\text{EbEgEm}_2\text{O}_8$, $\text{Eb}_2\text{EgEmO}_7$, $\text{Eb}_2\text{EgEmO}_9$, and $\text{Eb}_4\text{Eg}_2\text{EmO}_{18}$, respectively, which belong to the space groups of $Ccc2$ (No. 37), $Pccm$ (No. 49), $P4$ (No. 75), and $I4$ (No. 79), respectively. They do not correspond to any experimentally known crystal structures.

MARICI successfully designs the structures corresponding to the experimentally known complex crystal structures such as $\text{Ba}_2\text{Cr}_7\text{O}_{14}$, which is the optimal solution of $\text{Ea}_2\text{Ef}_5\text{El}_2\text{O}_{14}$ shown in Fig. 2(e). Besides, MARICI also designs many unknown structures whose existence in nature would not be surprising; for example, Figs. 2(f)–2(h) show the optimal

solutions of $\text{Eb}_2\text{Eg}_2\text{EmO}_{11}$, $\text{EaEf}_2\text{El}_3\text{O}_{11}$, and $\text{Ea}_3\text{EfElO}_7$, respectively, which belong to the space groups of $R32$ (No. 155), $P\bar{3}m1$ (No. 164), and $P6_3mc$ (No. 186), respectively. They do not correspond to any experimentally known crystal structures, but they are similar to the crystal structures of $\text{BaNb}_2\text{V}_2\text{O}_{11}$, $\text{BaV}_2\text{Zn}_3\text{O}_{10}\text{H}_2$, and $\text{NaBe}_4\text{SbO}_7$, respectively.

These results show promise for the computational discovery of unknown inorganic compounds based on mathematical crystal chemistry. While the theory cannot assume the suitable assignments of chemical elements to each atom site, the algorithm is useful to scan the coordination space as a pre-processor before evaluating formation enthalpies by *ab initio* simulations.

- [1] J. Maddox, Crystals from first principles, *Nature (London)* **335**, 201 (1988).
- [2] C. J. Pickard and R. J. Needs, High-pressure phases of silane, *Phys. Rev. Lett.* **97**, 045504 (2006).
- [3] C. J. Pickard and R. J. Needs, *Ab initio* random structure searching, *J. Phys.: Condens. Matter* **23**, 053201 (2011).
- [4] C. J. Pickard, Beyond theory-driven discovery: Introducing hot random search and datum-derived structures, *Faraday Discuss.* **256**, 61 (2025).
- [5] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Crystal structure prediction via particle-swarm optimization, *Phys. Rev. B* **82**, 094116 (2010).
- [6] H. Wang, Y. Wang, J. Lv, Q. Li, L. Zhang, and Y. Ma, Calypso structure prediction method and its wide application, *Comput. Mater. Sci.* **112**, 406 (2016).
- [7] X. Shao, J. Lv, P. Liu, S. Shao, P. Gao, H. Liu, Y. Wang, and Y. Ma, A symmetry-orientated divide-and-conquer method for crystal structure prediction, *J. Chem. Phys.* **156**, 014105 (2022).

- [8] X. Luo, Z. Wang, P. Gao, J. Lv, Y. Wang, C. Chen, and Y. Ma, Deep learning generative model for crystal structure prediction, *npj Comput. Mater.* **10**, 254 (2024).
- [9] *Computational Materials Discovery*, edited by A. R. Oganov, G. Saleh, and A. G. Kvashnin (The Royal Society of Chemistry, 2019).
- [10] A. R. Oganov and C. W. Glass, Crystal structure prediction using *ab initio* evolutionary techniques: Principles and applications, *J. Chem. Phys.* **124**, 244704 (2006).
- [11] X. Liu, H. Niu, and A. R. Oganov, Copex: Co-evolutionary crystal structure prediction algorithm for complex systems, *npj Comput. Mater.* **7**, 199 (2021).
- [12] I. A. Kruglov, A. V. Yanilkin, Y. Propad, A. B. Mazitov, P. Rachitskii, and A. R. Oganov, Crystal structure prediction at finite temperatures, *npj Comput. Mater.* **9**, 197 (2023).
- [13] Z. Falls, P. Avery, X. Wang, K. P. Hilleke, and E. Zurek, The XtalOpt evolutionary algorithm for crystal structure prediction, *J. Phys. Chem. C* **125**, 1601 (2021).
- [14] S. Hajinazar and E. Zurek, XTALOPT version 13: Multi-objective evolutionary search for novel functional materials, *Comput. Phys. Commun.* **304**, 109306 (2024).
- [15] S. Racioppi, A. O.-de-la Roza, S. Hajinazar, and E. Zurek, Powder X-ray diffraction assisted evolutionary algorithm for crystal structure prediction, *Digital Discovery* **4**, 73 (2025).
- [16] S. Hajinazar and E. Zurek, XtalOpt version 14: Variable-composition crystal structure search for functional materials through Pareto optimization, *Comput. Phys. Commun.* **320**, 109910 (2026).
- [17] K. Tomiyasu, R. Oishi-Tomiyasu, M. Matsuda, and K. Matsuhira, A new mathematical approach to finding global solutions of the magnetic structure determination problem, *Sci. Rep.* **8**, 16228 (2018).
- [18] L. E. Gibbons, D. W. Hearn, P. M. Pardalos, and M. V. Ramana, Continuous characterizations of the maximum clique problem, *Math. Oper. Res.* **22**, 754 (1997).
- [19] V. V. Gusev, D. Adamson, A. Deligkas, D. Antypov, C. M. Collins, P. Krysta, I. Potapov, G. R. Darling, M. S. Dyer, P. Spirakis, and M. J. Rosseinsky, Optimality guarantees for crystal structure prediction, *Nature (London)* **619**, 68 (2023).
- [20] L. Pauling, The principles determining the structure of complex ionic crystals, *J. Am. Chem. Soc.* **51**, 1010 (1929).
- [21] A. F. Wells, *Structural Inorganic Chemistry*, 5th ed. (Oxford University Press, 1984).
- [22] U. Müller, *Inorganic Structural Chemistry*, 2nd ed. (John Wiley & Sons, Ltd., 2007).
- [23] S. S. Batsanov and A. S. Batsanov, *Introduction to Structural Chemistry* (Springer, 2012).
- [24] T. C. Hales, A proof of the Kepler conjecture, *Ann. Math.* **162**, 1065 (2005).
- [25] L. Fillion and M. Dijkstra, Prediction of binary hard-sphere crystal structures, *Phys. Rev. E* **79**, 046714 (2009).
- [26] J. K. Kummerfeld, T. S. Hudson, and P. Harrowell, The densest packing of AB binary hard-sphere homogeneous compounds across all size ratios, *J. Phys. Chem. B* **112**, 10773 (2008).
- [27] P. I. O'Toole and T. S. Hudson, New high-density packings of similarly sized binary spheres, *J. Phys. Chem. C* **115**, 19037 (2011).
- [28] T. S. Hudson, Dense sphere packing in the NaZn_{13} structure type, *J. Phys. Chem. C* **114**, 14013 (2010).
- [29] T. S. Hudson and P. Harrowell, Structural searches using isopointal sets as generators: Densest packings for binary hard sphere mixtures, *J. Phys.: Condens. Matter* **23**, 194103 (2011).
- [30] A. B. Hopkins, Y. Jiao, F. H. Stillinger, and S. Torquato, Phase diagram and structural diversity of the densest binary sphere packings, *Phys. Rev. Lett.* **107**, 125501 (2011).
- [31] A. B. Hopkins, F. H. Stillinger, and S. Torquato, Densest binary sphere packings, *Phys. Rev. E* **85**, 021130 (2012).
- [32] R. Koshiji, M. Kawamura, M. Fukuda, and T. Ozaki, Diverse densest binary sphere packings and phase diagram, *Phys. Rev. E* **103**, 023307 (2021).
- [33] R. Koshiji and T. Ozaki, Densest ternary sphere packings, *Phys. Rev. E* **104**, 024101 (2021).
- [34] R. Koshiji and T. Ozaki, Diverse densest ternary sphere packings, *J. Phys. Commun.* **6**, 075002 (2022).
- [35] R. Koshiji, M. Fukuda, M. Kawamura, and T. Ozaki, Prediction of quaternary hydrides based on densest ternary sphere packings, *Phys. Rev. Mater.* **6**, 114802 (2022).
- [36] V. A. Blatov, A. P. Shevchenko, and D. M. Proserpio, Applied topological analysis of crystal structures with the program ToposPro, *Cryst. Growth Des.* **14**, 3576 (2014).
- [37] O. Delgado-Friedrichs, M. D. Foster, M. O'Keeffe, D. M. Proserpio, M. M. Treacy, and O. M. Yaghi, What do we know about three-periodic nets? *J. Solid State Chem.* **178**, 2533 (2005).
- [38] M. O'Keeffe, M. A. Peskov, S. J. Ramsden, and O. M. Yaghi, The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets, *Acc. Chem. Res.* **41**, 1782 (2008).
- [39] M. O'Keeffe and O. M. Yaghi, Deconstructing the crystal structures of metal-organic frameworks and related materials into their underlying nets, *Chem. Rev.* **112**, 675 (2012).
- [40] M. O'Keeffe, Nets, tiles, and metal-organic frameworks, *APL Mater.* **2**, 124106 (2014).
- [41] T. Yokoyama, K. Ichikawa, and H. Naito, Crystal structure generation based on polyhedra using dual periodic graphs, *Cryst. Growth Des.* **24**, 2168 (2024).
- [42] T. Yokoyama, K. Ichikawa, and H. Naito, From polyhedra to crystals: A graph-theoretic framework for crystal structure generation, *CrystEngComm* **28**, 2293 (2026).
- [43] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.* **120**, 145301 (2018).
- [44] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.* **31**, 3564 (2019).
- [45] C. W. Park and C. Wolverton, Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery, *Phys. Rev. Mater.* **4**, 063801 (2020).
- [46] Y. Jiang, D. Chen, X. Chen, T. Li, G.-W. Wei, and F. Pan, Topological representations of crystalline compounds for the machine-learning prediction of materials properties, *npj Comput. Mater.* **7**, 28 (2021).
- [47] V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, Benchmarking graph neural networks for materials chemistry, *npj Comput. Mater.* **7**, 84 (2021).
- [48] J. Cheng, C. Zhang, and L. Dong, A geometric-information-enhanced crystal graph network for predicting properties of materials, *Commun. Mater.* **2**, 92 (2021).

- [49] K. Choudhary and B. DeCost, Atomistic line graph neural network for improved materials property predictions, *npj Comput. Mater.* **7**, 185 (2021).
- [50] P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer, and P. Friederich, Graph neural networks for materials science and chemistry, *Commun. Mater.* **3**, 93 (2022).
- [51] T. Long, N. M. Fortunato, I. Opahle, Y. Zhang, I. Samathrakris, C. Shen, O. Gutfleisch, and H. Zhang, Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures, *npj Comput. Mater.* **7**, 66 (2021).
- [52] J. Schmidt, L. Pettersson, C. Verdozzi, S. Botti, and M. A. L. Marques, Crystal graph attention networks for the prediction of stable materials, *Sci. Adv.* **7**, eabi7948 (2021).
- [53] G. Cheng, X.-G. Gong, and W.-J. Yin, Crystal structure prediction by combining graph network and optimization algorithm, *Nat. Commun.* **13**, 1492 (2022).
- [54] C.-N. Li, H.-P. Liang, X. Zhang, Z. Lin, and S.-H. Wei, Graph deep learning accelerated efficient crystal structure search and feature extraction, *npj Comput. Mater.* **9**, 176 (2023).
- [55] H. Xiao, R. Li, X. Shi, Y. Chen, L. Zhu, X. Chen, and L. Wang, An invertible, invariant crystal representation for inverse design of solid-state materials using generative deep learning, *Nat. Commun.* **14**, 7027 (2023).
- [56] R. Pawar, A. Rout, S. Bhattacharjee, and P. Motamarri, Robust lightweight graph neural network framework for accelerating crystal structure prediction, *J. Chem. Inf. Model.* **65**, 6470 (2025).
- [57] T. Hong, J. Yang, and G. Cao, Crystal structure prediction based on diffusion model and graph network optimization, *J. Phys.: Mater.* **8**, 035011 (2025).
- [58] S. Yu, J. Wang, Y. Han, Q. Jia, Z. Liang, Z. Yang, Y. Pan, H. Gao, and J. Sun, A generalized method for refining and selecting random crystal structures using graph theory, *J. Chem. Phys.* **163**, 094106 (2025).
- [59] *Mixed Integer Nonlinear Programming*, edited by J. Lee and S. Leyffer (Springer, New York, NY, 2012).
- [60] I. E. Grossmann, *Advanced Optimization for Process Systems Engineering*, Cambridge Series in Chemical Engineering (Cambridge University Press, 2021).
- [61] P. M. Castro and I. E. Grossmann, Generalized disjunctive programming as a systematic modeling framework to derive scheduling formulations, *Ind. Eng. Chem. Res.* **51**, 5781 (2012).
- [62] I. E. Grossmann and F. Trespalacios, Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming, *AIChE J.* **59**, 3276 (2013).
- [63] F. Trespalacios and I. E. Grossmann, Review of mixed-integer nonlinear and generalized disjunctive programming methods, *Chem. Eng. Tech.* **86**, 991 (2014).
- [64] H. D. Perez and I. E. Grossmann, Extensions to generalized disjunctive programming: Hierarchical structures and first-order logic, *Optim. Eng.* **25**, 959 (2024).
- [65] D. Ovalle, A. Bhatia, C. D. Laird, and I. E. Grossmann, A logic-based decomposition for the global optimization of the multiperiod blending problem using symmetry-breaking cuts, *Ind. Eng. Chem. Res.* **65**, 3981 (2026).
- [66] R. D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, *Acta Crystallogr. A* **32**, 751 (1976).
- [67] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, Covalent radii revisited, *Dalton Trans.* **37**, 2832 (2008).
- [68] A. A. B. Baloch, S. M. Alqahtani, F. Mumtaz, A. H. Muqabel, S. N. Rashkeev, and F. H. Alharbi, Extending Shannon's ionic radii database using machine learning, *Phys. Rev. Mater.* **5**, 043804 (2021).
- [69] M. Alsaman, Y. A. Alghofaili, A. A. Baloch, H. Alsadah, A. A. Alsaui, S. M. Alqahtani, A. H. Muqabel, and F. H. Alharbi, Outliers in Shannon's effective ionic radii table and the table extension by machine learning, *Comput. Mater. Sci.* **228**, 112350 (2023).
- [70] R. Dronskowski, *Chemical Bonding* (De Gruyter, Berlin, Boston, 2023).
- [71] P. C. Müller, C. Ertural, J. Hempelmann, and R. Dronskowski, Crystal orbital bond index: Covalent bond orders in solids, *J. Phys. Chem. C* **125**, 7959 (2021).
- [72] P. C. Müller, L. S. Reitz, D. Hemker, and R. Dronskowski, Orbital-based bonding analysis in solids, *Chem. Sci.* **16**, 12212 (2025).
- [73] N. V. Sahinidis, M. Tawarmalani, and M. Yu, Design of alternative refrigerants via global optimization, *AIChE J.* **49**, 1761 (2003).
- [74] R. Koshiji and T. Ozaki, Mathematical crystal chemistry: A mathematical programming approach to crystal structures of inorganic compounds, *Phys. Rev. Mater.* **8**, 113801 (2024).
- [75] R. Koshiji, Mathematical crystal chemistry II: Random search for ionic crystals and analysis on oxide crystals registered in ICSD, *Phys. Rev. Res.* **7**, 033078 (2025).
- [76] T. Dewenter and A. K. Hartmann, Phase transition for cutting-plane approach to vertex-cover problem, *Phys. Rev. E* **86**, 041128 (2012).
- [77] G. Claussen and A. K. Hartmann, Cutting-plane algorithms and solution whitening for the vertex-cover problem, *Phys. Rev. E* **106**, 035305 (2022).
- [78] <https://marici.mathematical-crystal-chemistry.solutions/>.
- [79] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Scaling deep learning for materials discovery, *Nature (London)* **624**, 80 (2023).
- [80] C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, and H. Yang, *et al.*, A generative model for inorganic materials design, *Nature (London)* **639**, 624 (2025).
- [81] Y. Zhao, E. M. D. Siriwardane, Z. Wu, N. Fu, M. Al-Fahdi, M. Hu, and J. Hu, Physics guided deep learning for generative design of crystal materials with symmetry constraints, *npj Comput. Mater.* **9**, 38 (2023).
- [82] S. Fredericks, K. Parrish, D. Sayre, and Q. Zhu, PyXtal: A Python library for crystal structure generation and symmetry analysis, *Comput. Phys. Commun.* **261**, 107810 (2021).
- [83] Q. Zhang, A. Choudhury, and A. Chernatynskiy, A symmetry-oriented crystal structure prediction method for crystals with rigid bodies, *J. Phys.: Condens. Matter* **37**, 095901 (2025).

- [84] K. Momma and F. Izumi, *VESTA 3* for three-dimensional visualization of crystal, volumetric and morphology data, *J. Appl. Crystallogr.* **44**, 1272 (2011).
- [85] <https://marici.mathematical-crystal-chemistry.solutions/optimal-solutions/>.
- [86] Y.-T. Lee and T. Ozaki, OpenMX viewer: A web-based crystalline and molecular graphical user interface program, *J. Mol. Graphics Modell.* **89**, 192 (2019).
- [87] K. S. Atsushi Togo and I. Tanaka, Spglib: A software library for crystal symmetry search, *Sci. Technol. Adv. Mater.: Methods* **4**, 2384822 (2024).