

Mathematical crystal chemistry II: Random search for ionic crystals and analysis on oxide crystals registered in ICSD

Ryotaro Koshoji ^{*,†}

National Institute of Advanced Industrial Science and Technology, Nagoya, Aichi 463-8560, Japan



(Received 31 March 2025; accepted 7 July 2025; published 21 July 2025)

Mathematical crystal chemistry views crystal structures as the optimal solutions of the mathematical optimization problem formalizing inorganic structural chemistry. This paper introduces the minimum and maximum atomic radii depending on the types of geometrical constraints, extending the concept of effective atomic sizes. These radii define permissible interatomic distances instead of interatomic forces, constraining feasible types and connections of coordination polyhedra. The definition derives the continuous optimization problem as dense packings of flexible atomic spheres. Additionally, discrete variables and constraint functions, which give a choice of creatable types of geometrical constraints depending on the spatial order of atoms, are implemented to formalize the feasible atomic environment, such as the composition of coordination polyhedra, resulting in acceleration of structure search and decrease of the number of local minima. The framework identifies unique optimal solutions corresponding to the structures of spinel, pyrochlore (α and β), pyroxene, quadruple perovskite, cuprate superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$, and iron-based superconductor LaFeAsO . Notably, up to 95% of oxide crystal structure types in the Inorganic Crystal Structure Database align with the optimal solutions preserving experimental structures despite the discretized feasible atomic radii. These findings highlight the role of mathematical optimization problems as a theoretical foundation for mathematical crystal chemistry, enabling efficient structure prediction.

DOI: [10.1103/6yl6-fr8b](https://doi.org/10.1103/6yl6-fr8b)

I. INTRODUCTION

The Inorganic Crystal Structure Database (ICSD) [1] contains around 250 000 experimental inorganic structures as of 2025, but they only covers a tiny portion of the possible inorganic compounds considering the infinite combinations of the chemical elements across the periodic table. This is due to the traditional trial-and-error approach to discover high-performance functional materials, being time-consuming and labor-intensive. The current state-of-the-art approaches to accelerate the discovery of unknown materials is the computational design of materials, but it remains challenging due to the sheer size of freedom, such as chemical compositions, unit cells, and the atomic arrangements.

The standard approaches for crystal structure prediction mainly combine density functional theory calculations with structure searching algorithms, such as random search [2–4], genetic algorithms [5–7], and particle-swarm optimization [8,9]. These methods lead to a number of major discoveries, such as the high-pressure stabilized superconductor LaH_{10} [10–16]. However, they suffer from their limited applicability

to complex structures with large number of atoms of several kinds in the unit cells, because they are based on the direct sampling over the entire energy surface; the large computational cost of *ab initio* simulations causes a problem of insufficient sampling for a large system.

In recent years, machine learning methods, such as graph neural networks representing a structure of compounds with nodes and edges [17–24], have been introduced to provide a faster way to predict properties without using density functional methods. The graph neural network approaches can be applied to crystal structure prediction with combining symmetry constraints [25], graph theory techniques to generate initial structures [26], or the structure search algorithm, such as particle-swarm optimization [27,28]. Diffusion-based generative model is also presented for tackling inverse design tasks under desired property constraints [29]. Biased symmetry-adapted artificial intelligence algorithms successfully predicted complex crystal structures, such as the garnet structure [30], because sampling with space group significantly simplifies the problem of crystal structure prediction by generating more promising initial structures than random sampling. The integer programming approach also found the garnet structure as the global optimum with the space group constraint to narrow down the combinations of the discretized atom sites in the fixed cubic unit cell, where the total energy is calculated by classical two-body potentials so that the structural optimization problem is converted into the integer programming formalism [31].

The previous study on mathematical crystal chemistry [32] proposed a mathematical optimization problem formalizing the empirical rules of inorganic structural chemistry

^{*}Present address: Institute for Solid State Physics, The University of Tokyo, Kashiwa 277-8581, Japan.

[†]Contact author: cosaji@issp.u-tokyo.ac.jp

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

[33–35]. The structural aspects, such as the sizes of atoms, the coordination polyhedra, and connection of polyhedra, are formalized as the objective or constraint functions. For example, a chemical bond is defined to be the inequality constraints on the distances between atoms, and the objective function is given by the volume of the unit cell to reproduce the packing of atomic spheres. The mathematical model successfully made a wide variety of crystal structures, such as spinel and α -pyrochlore structures. Besides, the number of optimal solutions was found to be much smaller than the number of local minima in the total energy by *ab initio* simulations.

This study provides a more direct mathematical formalization of inorganic structural chemistry based on the previous study [32] in order that the concepts of mathematical crystal chemistry should be clarified as summarized in Fig. 1. The previous study [32] explicitly constrains feasible connections of coordination polyhedra by maximum number of common bridging anions, but they should be constrained by minimum distances between cations, because inorganic structural chemistry considers that the connection of polyhedra results from the electrostatic stability as highly charged cations prefer the vertex sharing to edge or face sharing of coordination polyhedra to lengthen the distances between cations. Note that mathematical crystal chemistry defines interatomic forces as geometrical constraints on the interatomic distances. Besides, this study explicitly defines the minimum and maximum atomic radii depending on the type of geometrical constraints as the extension of the effective sizes of atoms depending on the types of bonding forces. Furthermore, the rule to assign one of the types of geometrical constraints to every pair of atoms is mathematically formulated by introducing the creatability function $\Lambda_{ijT}^{(G)}$ similar to activation functions in neural networks, resulting in unifying the paired optimization problem derived in the previous study [32]. The unified formalization is also shown to have the capability of discovering a wide variety of ionic crystal structures as optimal solutions from randomly generated structures. Finally, oxide crystals registered in ICSD are analyzed whether they can be the optimal solutions preserving the experimental structures by assigning suitable coordination features to atoms. As a result, 78%, 12%, and 4% of oxide crystals have isotypic, homeotypic, or isopointal optimal structures, respectively, despite the discretized feasible atomic radii. The successful results shows that mathematical optimization problem can be the theoretical foundation of mathematical crystal chemistry to explore the crystal structure prototypes with small computation.

This paper is organized as follows: In Sec. II, the theoretical foundation of mathematical crystal chemistry is discussed. In Sec. III, the computational aspect is described. In Sec. IV, the application results are shown. In Sec. V, the comprehensive summary is provided.

II. MATHEMATICAL CRYSTAL CHEMISTRY

Interatomic forces stabilize a crystal structure by keeping all the interatomic distances within the stable ranges. Accordingly, mathematical crystal chemistry defines an interatomic force as the geometrical constraint G that constrains the

feasible interatomic distance as

$$d_{ijT}^{(G)} \leq x_\sigma \leq D_{ijT}^{(G)}, \quad (1)$$

where $d_{ijT}^{(G)}$ and $D_{ijT}^{(G)}$ are the minimum and maximum distances between atom i and atom j in lattice T , respectively, and x_σ is given by

$$x_\sigma = |\mathbf{x}_j + \mathbf{T} - \mathbf{x}_i|, \quad (2)$$

with $\mathbf{x}_i$ being the cartesian coordinate of atom i in the original cell, $\mathbf{T}$ being the lattice vector, and σ being the abbreviated index of ijT . In the following discussion, the atom j in lattice T is called atom jT . Since randomly generated structures are too difficult to satisfy all the constraints formed between every pair of atoms, a geometrical constraint is relaxed to be

$$(1 - \varepsilon)d_\sigma^{(G)} \leq x_\sigma \leq (1 + \varepsilon)D_\sigma^{(G)}, \quad (3)$$

where ε is an error rate. The geometrical constraints are classified according to the types of interatomic forces based on inorganic structural chemistry, and one of the types of geometrical constraints is assigned to each pair of atoms.

Figures 2(a)–2(c) show the ionic bonds I, anionic constraints A, and cationic constraints C formed in the quadruple perovskite structure, respectively. The three types of the geometrical constraints are formed between pairs of atoms depending on the formal charges, and they show the different structural aspects of ionic crystals. In addition to the three kinds of constraints, this study defines nonionic bond N_I and metallic constraint M as the types of geometrical constraints for random structure search. The former excludes anions from coordination polyhedra of cations if they do not create ionic bonds, by setting the minimum distances of nonionic bond more than the maximum bonding distance of ionic bonds. The latter is the repulsive constraints formed between cations if they do not share any ionic bonded anion. Note that minimum interatomic distances of cationic constraints can be larger than that of metallic constraints to restrict the permissive linking of coordination polyhedra. Finally, Fig. 2(d) shows another aspect that ionic crystals are packings of ionic spheres, in fact, oxide ions in crystals with small cations tend to constitute the densest packings of hard spheres [36].

Since inorganic structural chemistry estimates a chemical bonding distance as the sum of atomic radii, mathematical crystal chemistry defines the minimum and maximum interatomic distances as

$$d_{ijT}^{(G)} = r_i^{(G)} + r_j^{(G)}, \quad (4)$$

$$D_{ijT}^{(G)} = R_i^{(G)} + R_j^{(G)}, \quad (5)$$

where $r^{(G)}$ and $R^{(G)}$ are the minimum and maximum atomic radii depending on the type of the geometrical constraint G , respectively. Since the interatomic forces corresponding to anionic, cationic, nonionic, metallic constraints are repulsive, the maximum interatomic distances of the four kinds should be ∞ . Therefore, the four kinds of the maximum atomic radii should be defined as

$$R^{(A)} = R^{(C)} = R^{(N_I)} = R^{(M)} = \infty. \quad (6)$$

The other six kinds of atomic radii are the factors to determine the feasible atomic environments, such as the maximum coordination number.

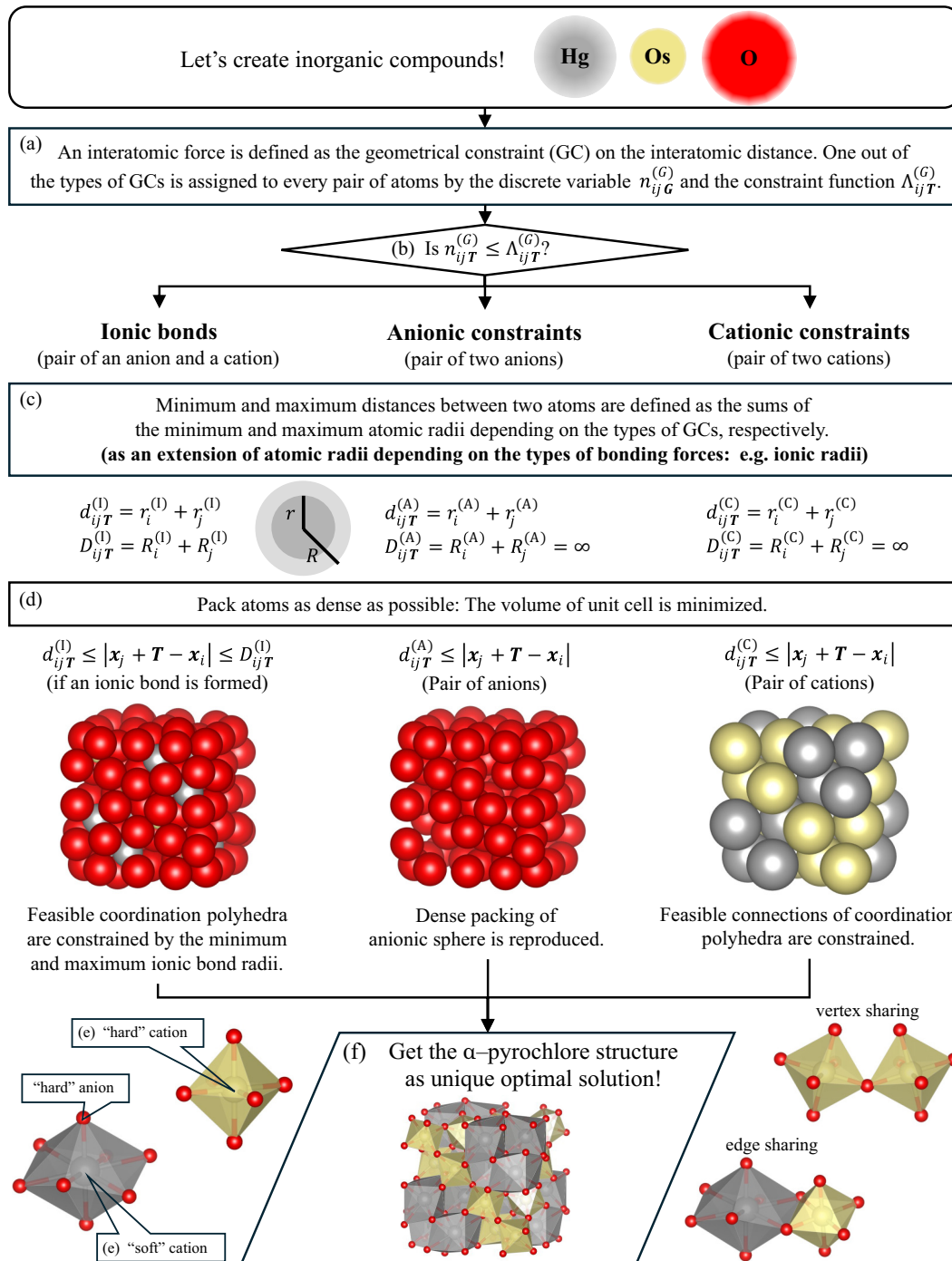


FIG. 1. Concept of mathematical crystal chemistry. (a) The geometrical constraint is defined in Eq. (1). Since one of the types of geometrical constraints is assigned to every pair of atoms, the sum of $n_{ijT}^{(G)}$ with respect to G must be one as given in Eq. (7). The creatability function $\Lambda_{ijT}^{(G)}$ is defined in Eq. (12). (b) The rule to assign geometrical constraints is mathematically formalized by inequality constraints as the sixth constraint in the mathematical optimization problem given in Eq. (17). (c) The minimum and maximum interatomic distances of a geometrical constraint are defined in Eqs. (4) and (5), respectively. (d) The volume of unit cell is the objective function of the mathematical optimization problem given in Eq. (17). (e) Flexible atomic sphere is discussed in Sec. II and Appendix A. (f) A wide variety of crystal structures identified as unique optimal solutions are discussed in Sec. IV.

The minimum cationic repulsion radius $r^{(C)}$ constrains the feasible connection of coordination polyhedra; large $r^{(C)}$ lengthens the distance between cations resulting in the infeasibility of the edge or face sharing of coordination polyhedra. This infeasibility corresponds to large repulsive force between

highly charged cations; inorganic structural chemistry considers that highly charged cations prefer the vertex sharing to the edge or face sharing of coordination polyhedra for reducing electrostatic repulsion by lengthening the interatomic distance. Figure 3 shows the vertex, edge, and face sharings of

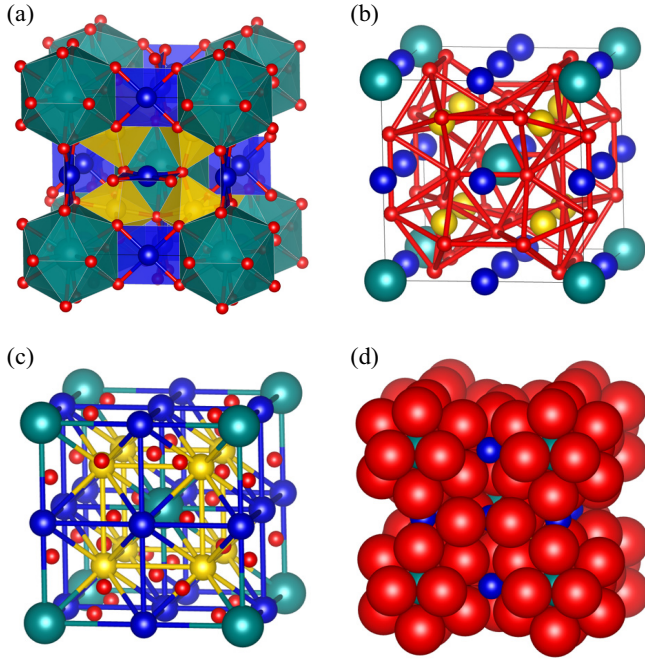


FIG. 2. Different structural aspects of ionic crystals depending on the type of the geometrical constraints that correspond to the types of interatomic forces. Cylinders connecting two atoms correspond to geometrical constraints. In this study, all the figures of crystal structures are drawn by VESTA [37]. (a) An ionic bond can be formed between a pair of an anion and a cation unless their interatomic distance will be infeasible when it is formed. (b) Anionic constraints are formed between every pair of anions. (c) A cationic constraint is formed between a pair of metal atoms if they have common bridging anions. (d) Packing of ionic spheres. If the sizes of cations are small enough, then they are placed in the tetrahedral or octahedral site of the densest packings of oxide ions.

coordination polyhedra in the quadruple perovskite structure, in which iron, copper, and calcium atoms make the octahedron, square, and regular icosahedron, respectively. This structure can be reproduced as the unique optimal solution when all the $r^{(C)}$ of iron, copper, and calcium atoms are set to be $C^{(Qd)}$, where the discretized feasible atomic radii in this study are defined in Appendix A. While the edge and face sharing of two octahedra is infeasible due to the large $r^{(C)}$, the edge sharing of square and regular icosahedron and the face sharing of octahedron and regular icosahedron are feasible, because the ionic bond radius of calcium is large.

Since many crystal structures, such as ionic crystals and intermetallic compounds, have the aspect of the packing of atomic spheres, the objective function of structural optimization problem is defined as the volume of unit cell Ω . Accordingly, the dense packings of flexible atomic spheres, which are the optimal solutions of the mathematical optimization problem, are regarded as crystal structure prototypes.

Inorganic structural chemistry can estimate the atomic environments of each atom from the given chemical composition, for example, cations will be surrounded by anions if the number of anions is large enough. Then, mathematical crystal chemistry introduces the discrete variables to constrain the feasible atomic environments. Let $n_{\sigma}^{(G)} \in \{0, 1\}$ is the number

of created geometrical constraints G assigned to the atoms i and jT . Since the pair is assigned to one of the types of geometrical constraints, the sum of $n_{\sigma}^{(G)} \in \{0, 1\}$ with respect to G must be one:

$$\sum_G n_{\sigma}^{(G)} = 1. \quad (7)$$

The rule of assignment is mathematically represented by the equality or inequality constraints on $n_{\sigma}^{(G)} \in \{0, 1\}$; let $\Lambda_{\sigma}^{(G)}(\Phi, \varepsilon)$ is the creatability function depending on the spatial order of atoms Φ defined by

$$\Phi = X \otimes n, \quad (8)$$

where n is the state vector of geometrical constraints and X is the structure matrix defined by

$$X = (x_1, \dots, x_n, \phi_1, \dots, \phi_n, t_1, t_2, t_3), \quad (9)$$

with ϕ_i being the coordination feature vector including atomic number and several kinds of atomic radii, and t_i being the lattice vector. Like an activation function in neural networks, this function returns one or more if the geometrical constraint G is creatable. The inequality constraints using $\Lambda_{\sigma}^{(G)}(\Phi, \varepsilon)$ given in the sixth constraint of Eq. (17) gives a choice of creatable types of geometrical constraints, and the suitable one is chosen for the feasibility and/or optimality of the structure. This study gives a simple definition of $\Lambda_{\sigma}^{(G)}(\Phi, \varepsilon)$ as follows: First, let $\chi_i^{(\varepsilon)}(X)$ be the function of

$$\chi_i^{(\varepsilon)}(X) = \begin{cases} 1 & (N_i^{(A)} \in \varepsilon) \\ 0 & (N_i^{(A)} \notin \varepsilon) \end{cases}, \quad (10)$$

where $N_i^{(A)}$ is the atomic number of atom i , ε is the subset of atomic numbers. In this study, the four kinds of subsets ε are defined: Metal atoms $\mathcal{M}$, anions $\mathcal{A}$, pnictogens $\mathcal{P}$, and oxygen $\mathcal{O}$. Second, let $\Xi_{\sigma}^{(G)}(X, \varepsilon)$ be the function defined by

$$\Xi_{\sigma}^{(G)}(X, \varepsilon) = \begin{cases} 1 & (\text{satisfy Eq. (3)}) \\ 0 & (\text{otherwise}) \end{cases}. \quad (11)$$

Accordingly, this study defines the creatability function $\Lambda_{\sigma}^{(G)}(\Phi, \varepsilon) \in \{0, 1\}$ as

$$\begin{aligned} \Lambda_{\sigma}^{(I)}(\Phi, \varepsilon) &= [\chi_i^{(\mathcal{M})} \chi_j^{(\mathcal{A})} + \chi_i^{(\mathcal{A})} \chi_j^{(\mathcal{M})}] \Xi_{\sigma}^{(I)}(X, \varepsilon), \\ \Lambda_{\sigma}^{(N_i)}(\Phi, \varepsilon) &= \chi_i^{(\mathcal{M})} \chi_j^{(\mathcal{A})} + \chi_i^{(\mathcal{A})} \chi_j^{(\mathcal{M})}, \\ \Lambda_{\sigma}^{(A)}(\Phi, \varepsilon) &= \chi_i^{(\mathcal{A})} \chi_j^{(\mathcal{A})}, \\ \Lambda_{\sigma}^{(C)}(\Phi, \varepsilon) &= \begin{cases} \chi_i^{(\mathcal{M})} \chi_j^{(\mathcal{M})} & (n_{\sigma}^{(\text{common})} > 0) \\ 0 & (n_{\sigma}^{(\text{common})} = 0) \end{cases}, \\ \Lambda_{\sigma}^{(N_M)}(\Phi, \varepsilon) &= \begin{cases} \chi_i^{(\mathcal{M})} \chi_j^{(\mathcal{M})} & (n_{\sigma}^{(\text{common})} = 0) \\ 0 & (n_{\sigma}^{(\text{common})} > 0) \end{cases}, \end{aligned} \quad (12)$$

where $n_{\sigma}^{(\text{common})}$ is the number of common bridging anions between metals i and jT that can be formalized as

$$n_{\sigma}^{(\text{common})} = \chi_i^{(\mathcal{M})} \chi_j^{(\mathcal{M})} \sum_{k, T'} \chi_k^{(\mathcal{A})} n_{ikT'}^{(I)} n_{jk(T'-T)}^{(I)}. \quad (13)$$

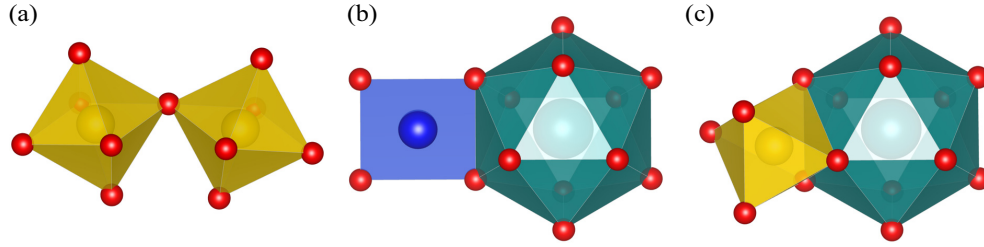


FIG. 3. Connection of polyhedra in the quadruple perovskite structure of $\text{CaCu}_3\text{Fe}_4\text{O}_{12}$. Calcium, copper, iron, and oxygen atoms are shown as dark-cyan, blue, gold, and red balls, respectively. All the minimum distances between cations are $2C^{(\text{Qd})}$, but the feasible connection of two polyhedra depends on the sizes of the two polyhedra. (a) Vertex sharing of two octahedra. The edge or face sharings are infeasible due to the large minimum distances between central cations. (b) Edge sharing of square and regular icosahedron. The edge sharing is feasible due to the large size of regular icosahedron. (c) Face sharing of octahedron and regular icosahedron. The face sharing is feasible due to the large size of regular icosahedron.

Note that $n_\sigma^{(\text{common})}$ corresponds to the connection of coordination polyhedra as

$$\begin{aligned} n_\sigma^{(\text{common})} &= 1 && \text{(Vertex sharing of polyhedra)} \\ n_\sigma^{(\text{common})} &= 2 && \text{(Edge sharing of polyhedra)} \\ n_\sigma^{(\text{common})} &\geq 3 && \text{(Face sharing of polyhedra).} \end{aligned} \quad (14)$$

The creatability function in this study returns one if the geometrical constraint is creatable but zero otherwise. Both the ionic and nonionic bonds can be formed between an anion and a cation, but the former has the additional condition that the interatomic distance has to be feasible if it is formed. Note that if $R^{(\text{I})} < r^{(\text{Ni})}$, the two kinds of the constraints satisfy

$$\Lambda_\sigma^{(\text{I})}(\Phi, 0) \cdot \Lambda_\sigma^{(\text{Ni})}(\Phi, 0) = 0. \quad (15)$$

In this study, $r^{(\text{Ni})}$ is defined with the coefficient $\gamma^{(\text{Ni})}$ as

$$r^{(\text{Ni})} = \gamma^{(\text{Ni})} R^{(\text{I})}. \quad (16)$$

The coefficient is set to be 1.2 for random structure search but 1.0 for the analysis of oxide crystals registered in ICSD. However, while anionic constraints can always be formed between two anions, cationic constraints cannot be formed unless common bridging anions exist and metallic constraints vice versa.

The mathematical optimization problem searching structural prototypes of ionic compounds is formalized as follows: First, the structure minimizes Ω . Second, all the interatomic distances must be feasible. Third to sixth, one of the types of geometrical constraints is assigned to every pair of atoms, and the minimum and maximum distances are the sums of the minimum and maximum atomic radii corresponding to the assigned types of geometrical constraints, respectively. Seventh constraint gives a choice of the creatable types of geometrical constraints for every pair of atoms. Accordingly,

the optimization problem for ionic crystals is given by

$$\begin{aligned} &\text{minimize} \quad \Omega, \\ &\text{subject to} \quad (1 - \varepsilon)d_\sigma \leq x_\sigma \leq (1 + \varepsilon)D_\sigma, \\ &\quad d_\sigma = \max_G [n_\sigma^{(\text{G})}(r_i^{(\text{G})} + r_j^{(\text{G})})], \\ &\quad D_\sigma = \min_G [n_\sigma^{(\text{G})}(R_i^{(\text{G})} + R_j^{(\text{G})})], \\ &\quad \sum_G n_\sigma^{(\text{G})} = 1, \\ &\quad n_\sigma^{(\text{G})} \in \{0, 1\}, \\ &\quad n_\sigma^{(\text{G})} \leq \Lambda_\sigma^{(\text{G})}(\Phi, \varepsilon). \end{aligned} \quad (17)$$

The optimization problem consists of the continuous optimization problem with respect to the atomic arrangement $\{\mathbf{x}_i\}$ in the variable crystal lattice $\{\mathbf{t}_i\}$ and the discrete optimization problem with respect to $n_\sigma^{(\text{G})}$. The discrete variables not only considerably reduces the number of optimal solutions but also accelerate the discovery of crystal structure prototypes. Note that if $\Lambda_\sigma^{(\text{I})}(\Phi, \varepsilon) = 1$ with $\varepsilon = 0$, then the ionic bond must be formed if $R^{(\text{I})} < r^{(\text{Ni})}$. In that case, the geometrical constraints are deterministically assigned to all the pairs of atoms.

In this study, the coordination numbers of every cation are fixed to accelerate the discovery of optimal structures by limiting the feasible coordination polyhedra, while anions have no constraint on the coordination numbers. Let $N^{(\text{I})}$ be the vector of the fixed coordination numbers depending on the subsets of atomic numbers; for example, if a cation is surrounded by four oxide ions and four pnictogen ions, then the coordination vector is given by $(N_{\text{O}}^{(\text{I})}, N_{\text{P}}^{(\text{I})}) = (4, 4)$. Since the creatability functions $\Lambda_\sigma^{(\text{I})}(\Phi, \varepsilon)$ given in Eq. (12) do not use the information on the coordination numbers, this study marks up the additional constraints in Eq. (17) given by

$$\sum_{j, \mathbf{T}} \chi_j^{(\varepsilon)} n_\sigma^{(\text{I})} = N_{i\varepsilon}^{(\text{I})}. \quad (18)$$

This equation constrains the coordination numbers depending on the subsets of atomic numbers. If an atom does not satisfy the constraints, then the structure is regarded as an infeasible solution.

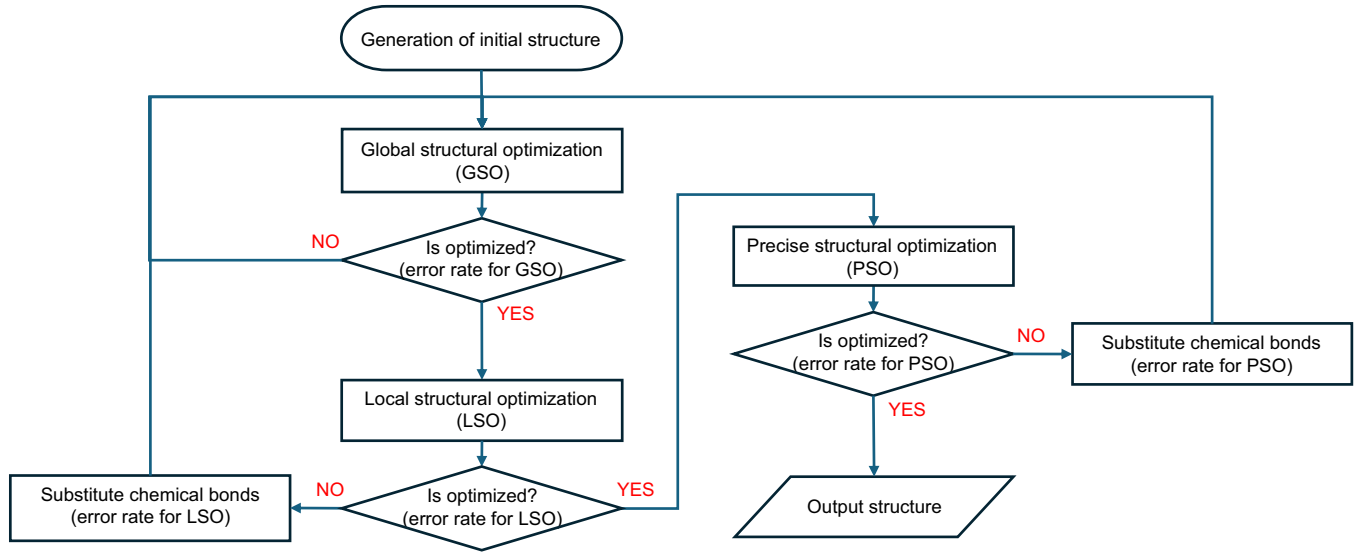


FIG. 4. Flow chart of the algorithm to find the optimal solutions of the problem given in Eq. (17) from random sampling. In the global optimization, $\mathbf{X}$ and $\mathbf{n}$ are alternately optimized, while in the local and precise optimization, only $\mathbf{X}$ is optimized to identify whether all the geometrical constraints can be feasible.

III. METHODS

The flowchart to solve the problem of Eq. (17) with Eq. (18) from random sampling is shown in Fig. 4. How to generate initial structures is discussed in Appendix B. The algorithm is composed of the three parts of the global, local, and precise optimization, which have larger error rate ε from left to right. Dynamic error rate is necessary as discussed in Appendix C. If a structure is optimal under an error rate, then the structure is judged whether it can be an optimal solution under smaller error rate by optimizing with smaller displacement sizes of atoms and crystal lattice. The global optimization is aimed at finding different network of ionic bonds by large deformation of the structure, while the local and precise optimization is aimed at judging whether all the geometrical constraints can be feasible. Therefore, $\mathbf{X}$ and $\mathbf{n}$ are alternately optimized in the global optimization to escape from infeasible region with large displacement size, while the optimization on $\mathbf{n}$ is only applied at the beginning in the local and precise optimization. Note that the optimization of $\mathbf{X}$ with fixed $\mathbf{n}$ corresponds to the structural optimization whose algorithm is summarized in Appendix C, while the optimization of $\mathbf{n}$ with fixed $\mathbf{X}$ corresponds to the replacement of the geometrical constraints.

In many cases, it is difficult for a structure to escape from the infeasible region by only the alternative optimization of $\mathbf{X}$ and $\mathbf{n}$. Therefore, if a structure is regarded as an infeasible structure after the local or precise optimization, then the ionic bonds are randomly substituted as follows: First, one cation owing infeasible ionic bonds is randomly chosen. Note that if a cation does not satisfy the implicit constraints on the feasible connection of coordination polyhedra, then some ionic bonds are stretched after the structural optimization, because the repulsive force is dominant compared to the attractive force and the pressure as discussed in Appendix C. Second, the cation randomly choose one of owing nonionic bonds. Third, all the owing ionic bonds are converted into nonionic bonds.

Finally, the chosen nonionic bond is converted into ionic bond. This operation enables the infeasible cation to move into a neighboring site of the packing of oxide ions.

How to replace the types of geometrical constraints is simple: First, every cation creates as many creatable ionic bonds as possible under the constraint on the coordination numbers. Second, if a pair of an anion and a cation does not have an ionic bond, then a nonionic bond is formed between the pair. Third, cationic constraints are formed if the two cations have common bridging anions. Fourth, if a pair of two cations does not have a cationic constraint, then a metallic constraint is formed between the pair. Fifth, anionic constraints are formed between every pair of anions. This operation can be mathematically formalized as

$$\begin{aligned}
 &\text{maximize} \quad \sum_i \sum_{j \in T} \chi_i^{(\mathcal{M})} n_{\sigma}^{(I)}, \\
 &\quad \sum_G n_{\sigma}^{(G)} = 1, \\
 &\quad n_{\sigma}^{(G)} \in \{0, 1\}, \\
 &\quad n_{\sigma}^{(G)} \leq \Lambda_{\sigma}^{(G)}(\mathbf{X}, \varepsilon), \\
 &\quad \sum_{j \in T} \chi_j^{(\mathcal{E})} n_{\sigma}^{(I)} \leq N_{i\mathcal{E}}^{(I)}.
 \end{aligned} \tag{19}$$

Note that since an ionic bond is not creatable when the interatomic distance is larger, a cation does not necessarily satisfy the equality constraints on the coordination numbers. Besides, since the optimization problem of Eq. (19) does not include the constraint on which neighboring anions around a central cation are chosen to create ionic bonds, how to choose anions for creating ionic bonds is implementation-defined; the algorithm used in this study makes every cation create ionic bonds with closest anions to the cation as the sum of the ionic bonding distances is minimized.

TABLE I. Chemical compositions to which the mathematical optimization problem is applied. The coordination feature, the number of randomly generated structures, and the discovered optimal solutions are also listed. A coordination feature is written as $(N^{(A)}, \mathcal{E}, R^{(I)}, r^{(I)}, R^{(R)}, N_I)$ with $R \in \{A, C\}$, where the atomic number $N^{(A)}$ is expressed by element symbol for a straightforward manner. Oxygen has the coordination feature given by $(O, \mathcal{O}, C^{(O)}, C^{(O)}, C^{(O)}, \mathcal{M}_*)$, where oxygen can create ionic bonds with metal atoms freely as expressed by $N_I = (\mathcal{M}_*)$.

Chemical composition	Coordination features ϕ_i	Number of generated structures	Discovered optimal solutions
$\text{Si}_4\text{Fe}_8\text{O}_{16}$	$(\text{Fe}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(O)}, \mathcal{O}_6)$ $(\text{Si}, \mathcal{M}, C^{(Q)}, C^{(Q)}, C^{(Pd)}, \mathcal{O}_4)$	5×10^6	spinel structure
$\text{Si}_4\text{Fe}_8\text{O}_{16}$	$(\text{Fe}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(O)}, \mathcal{O}_6)$ $(\text{Si}, \mathcal{M}, C^{(Q)}, C^{(Q)}, C^{(O)}, \mathcal{O}_4)$	1×10^7	spinel structure $\beta\text{-Mg}_2\text{SiO}_4$ structure other 207 structures
$\text{Si}_4\text{Fe}_8\text{O}_{16}$	$(\text{Fe}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(O)}, \mathcal{O}_6)$ $(\text{Si}, \mathcal{M}, C^{(Q)}, C^{(Q)}, C^{(E)}, \mathcal{O}_4)$	1×10^6	spinel structure $\beta\text{-Mg}_2\text{SiO}_4$ structure Mg_2SiO_4 structure other 5590 structures
$\text{Si}_4\text{Fe}_4\text{O}_{12}$	$(\text{Fe}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(O)}, \mathcal{O}_6)$ $(\text{Si}, \mathcal{M}, C^{(Q)}, C^{(Q)}, C^{(O)}, \mathcal{O}_4)$	5×10^6	$\text{CaMgSi}_2\text{O}_6$ structure Ga_2O_3 structure other 602 structures
$\text{Si}_2\text{Fe}_{14}\text{O}_{20}$	$(\text{Fe}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(O)}, \mathcal{O}_6)$ $(\text{Si}, \mathcal{M}, C^{(Q)}, C^{(Q)}, C^{(O)}, \mathcal{O}_4)$	1×10^6	$\text{SiFe}_7\text{O}_{10}$ structure other 66 structures
$\text{Hg}_4\text{Os}_4\text{O}_{14}$	$(\text{Hg}, \mathcal{M}, C^{(Wd)}, C^{(S)}, C^{(Hd)}, \mathcal{O}_8)$ $(\text{Os}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Hd)}, \mathcal{O}_6)$	5×10^6	α -pyrochlore structure
$\text{Ca}_4\text{Os}_4\text{O}_{14}$	$(\text{Ca}, \mathcal{M}, C^{(Td)}, C^{(S)}, C^{(Hd)}, \mathcal{O}_8)$ $(\text{Os}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Hd)}, \mathcal{O}_6)$	5×10^6	α -pyrochlore structure weberite structure other 6 structures
$\text{Rb}_2\text{Os}_4\text{O}_{12}$	$(\text{Rb}, \mathcal{M}, C^{(Qv)}, C^{(Sd)}, C^{(V)}, \mathcal{O}_{18})$ $(\text{Os}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Hd)}, \mathcal{O}_6)$	5×10^6	β -pyrochlore structure
$\text{Ca}_2\text{Pt}_3\text{O}_8$	$(\text{Ca}, \mathcal{M}, C^{(E)}, C^{(E)}, C^{(Pd)}, \mathcal{O}_6)$ $(\text{Pt}, \mathcal{M}, C^{(S)}, C^{(H)}, C^{(Td)}, \mathcal{O}_6)$	1×10^6	$\text{Na}_2\text{Mn}_3\text{O}_8$ structure other 575 structures
$\text{Ca}_4\text{Ir}_4\text{O}_{12}$	$(\text{Ca}, \mathcal{M}, C^{(N)}, C^{(E)}, C^{(Td)}, \mathcal{O}_8)$ $(\text{Ir}, \mathcal{M}, C^{(H)}, C^{(Q)}, C^{(Td)}, \mathcal{O}_6)$	5×10^6	pyroxene structure
$\text{CaCu}_3\text{Fe}_4\text{O}_{12}$	$(\text{Ca}, \mathcal{M}, C^{(Wd)}, C^{(Wd)}, C^{(Qd)}, \mathcal{O}_{12})$ $(\text{Cu}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Qd)}, \mathcal{O}_4)$ $(\text{Fe}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Qd)}, \mathcal{O}_6)$	5×10^6	quadruple perovskite structure
$\text{YBa}_2\text{Cu}^{[4I]}\text{Cu}_2^{[5Y]}\text{O}_7$	$(\text{Y}, \mathcal{M}, C^{(N)}, C^{(E)}, C^{(Qd)}, \mathcal{O}_8)$ $(\text{Ba}, \mathcal{M}, C^{(Td)}, C^{(O)}, C^{(Qd)}, \mathcal{O}_{10})$ $(\text{Cu}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Pd)}, \mathcal{O}_4)$ $(\text{Cu}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Qd)}, \mathcal{O}_5)$	1×10^6	$\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ structure
$\text{Y}_2\text{Ba}_4\text{Cu}_2^{[4I]}\text{Cu}_4^{[5Y]}\text{O}_{14}$	$(\text{Y}, \mathcal{M}, C^{(N)}, C^{(E)}, C^{(Qd)}, \mathcal{O}_8)$ $(\text{Ba}, \mathcal{M}, C^{(Td)}, C^{(O)}, C^{(Qd)}, \mathcal{O}_{10})$ $(\text{Cu}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Pd)}, \mathcal{O}_4)$ $(\text{Cu}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(Qd)}, \mathcal{O}_5)$	5×10^6	$\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ structure other 5 structure
$\text{La}_4\text{Cu}_2\text{O}_8$	$(\text{Cu}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(O)}, \mathcal{O}_6)$ $(\text{La}, \mathcal{M}, C^{(O)}, C^{(O)}, C^{(Sd)}, \mathcal{O}_9)$	5×10^5	K_2NiF_4 structure Distorted K_2NiF_4 structure
$\text{La}_2\text{Fe}_2\text{As}_2\text{O}_2$	$(\text{La}, \mathcal{M}, C^{(O)}, C^{(O)}, C^{(Nd)}, \mathcal{O}_4\mathcal{P}_4)$ $(\text{Fe}, \mathcal{M}, C^{(H)}, C^{(H)}, C^{(O)}, \mathcal{P}_4)$ $(\text{As}, \mathcal{P}, C^{(Ed)}, C^{(Ed)}, C^{(Ed)}, \mathcal{M}_*)$	1×10^6	ZrCuSiAs structure

IV. RESULT AND DISCUSSION

A. Crystal structure reproduction

The mathematical optimization problem is applied to several chemical compositions listed in Table I. The structure types are identified by the structure fingerprints defined in Appendix D. Note that largely distorted structures are regarded as different from the symmetric structure. Many coordination features ϕ_i are almost the same as those automatically assigned in the analysis discussed in Sec. IV B. Since cations

in a crystal structure have similar cationic repulsion radii, we can guess that cations tend to be evenly distributed. This study also indicate that the number of optimal solutions is much smaller than the number of local minima in the total energy by *ab initio* simulations [32].

Previous study explicitly constrains the number of common bridging anions as

$$n_{\sigma}^{(\text{common})} \leq N_{\sigma}^{(\text{common})}, \quad (20)$$

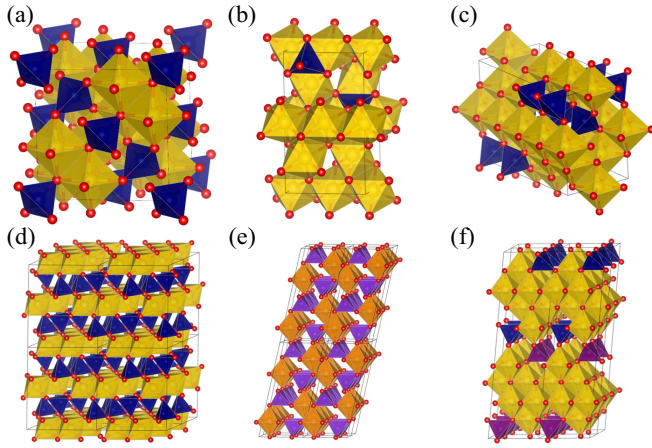


FIG. 5. Silicon atoms are shown as dark-blue balls. A silicon atom makes a tetrahedron, while an iron atom makes an octahedron. (a) Spinel structure of $\text{Si}_4\text{Fe}_8\text{O}_{16}$. (b) Mg_2SiO_4 structure of $\text{Si}_4\text{Fe}_8\text{O}_{16}$. (c) $\beta\text{-Mg}_2\text{SiO}_4$ structure of $\text{Si}_4\text{Fe}_8\text{O}_{16}$. (d) $\text{CaMgSi}_2\text{O}_6$ structure of $\text{Si}_4\text{Fe}_4\text{O}_{12}$. (e) Ga_2O_3 structure of $\text{Ga}_4^{[4r]}\text{Ga}_4^{[6o]}\text{O}_{12}$. The blue-violet and dark-orange balls correspond to gallium atoms whose coordination number is four and six, respectively. The former makes a tetrahedron, while the latter makes an octahedron. (f) $\text{SiFe}_7\text{O}_{10}$ structure of $\text{Si}_2\text{Fe}_{14}\text{O}_{20}$. The purple balls correspond to iron atoms whose coordination number is four.

where $N_{\sigma}^{(\text{common})}$ is the maximum number of common bridging anions. However, this study only constrains the distances between the cations resulting in the implicit constraints on the connection of coordination polyhedra, which conceptualizes more directly the reason why highly charged cations tend to prefer the vertex sharing to the edge or face sharing, since mathematical crystal chemistry defines interatomic forces as the constraints on interatomic distances. This alleviation also enables us to automatically set the $N_{\sigma}^{(\text{common})}$ according to the $r^{(C)}$ of the chemical elements. For example, as shown in Fig. 3, two octahedra in $\text{CaCu}_3\text{Fe}_4\text{O}_{12}$ cannot share their edges or faces due to the large $r^{(C)}$, but an octahedron and a regular icosahedron can share their faces due to the large ionic bond radius of Ca. Additionally, larger $r^{(C)}$ can decrease the number of optimal solutions. For example, the spinel structure of Fe_2SiO_4 shown in Fig. 5(a) can be a unique optimal solution when $r^{(C)}$ of silicon atoms is set to be $C^{(\text{Pd})}$. However, Fe_2SiO_4 can also constitute the Mg_2SiO_4 structure shown in Fig. 5(b). This structure, in which a tetrahedron and an octahedron share their edges, can be the optimal solution when $r^{(C)}$ of silicon atoms is set to be $C^{(\text{E})}$. Besides, $\text{Fe}_{2+x}\text{Si}_{1-x}\text{O}_4$ can constitute the $\beta\text{-Mg}_2\text{SiO}_4$ structure shown in Fig. 5(c). This structure can be the optimal solution when $r^{(C)}$ of silicon atoms is set to be $C^{(\text{O})}$. In terms of the different compositions, the $\text{CaMgSi}_2\text{O}_6$, Ga_2O_3 , and $\text{SiFe}_7\text{O}_{10}$ structures shown in Figs. 5(d)–5(f), respectively, can be found as the optimal solutions when $r^{(C)}$ of silicon atoms is also set to be $C^{(\text{O})}$. This results indicate that too large $r^{(C)}$ makes it difficult to reproduce the crystal polymorphism.

The α -pyrochlore structure shown in Fig. 6(a) can be the unique optimal solution when $R^{(\text{I})}$ of mercury atoms, whose coordination number is eight, is set to be $C^{(\text{Wd})}$. However, the weberite structure shown in Fig. 6(b) can be one of the

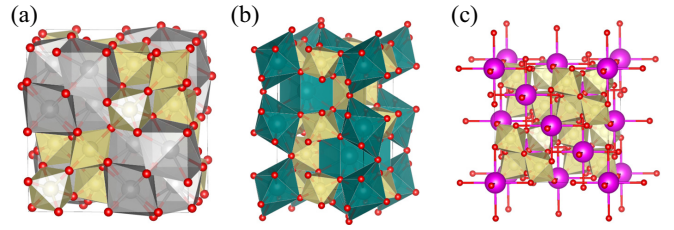


FIG. 6. (a) α -pyrochlore structure of $\text{Hg}_4\text{Os}_4\text{O}_{14}$. Mercury and osmium atoms are shown as dark-gray and khaki balls, respectively. A mercury atom makes a distorted hexagonal bipyramid, while an osmium atom makes an octahedron. (b) Weberite structure of $\text{Ca}_4\text{Os}_4\text{O}_{14}$. A calcium atom makes a hexagonal bipyramid, while an osmium atom makes an octahedron. (c) β -pyrochlore structure of $\text{Rb}_2\text{Os}_4\text{O}_{12}$. Rubidium atoms are shown as magenta balls, respectively. A rubidium atom makes a large octahedron, while an osmium atom makes a closed-packed octahedron.

optimal solutions with the α -pyrochlore structure when $R^{(\text{I})}$ of calcium atoms, whose coordination number is also eight, is set to be $C^{(\text{Td})}$. This result clearly supports the empirical rules of inorganic structural chemistry that the sizes of atoms are one of the main factors to determine the crystal structures. However, β -pyrochlore structure shown in Fig. 6(c) can also be unique optimal solution when $R^{(\text{I})}$ of rubidium atoms, whose coordination number is eighteen, is set to be $C^{(\text{Qv})}$. However, although the six ionic bonds of the rubidium is about 0.8 times smaller than the other 12 ionic bonds, the structure cannot be reproduced if the coordination number of rubidium is set to be six with decreasing $R^{(\text{I})}$; in that case, a rubidium atom makes a largely distorted octahedron resulting in contact with another rubidium atom. It might be better for alkali metals to alleviate the constraints on the coordination numbers into the constraints that the cation has only to be surrounded by anions.

While a calcium ion in the weberite structure constitute the hexagonal bipyramid, the element can constitute a wide variety of coordination polyhedra. Figure 7(a) shows the crystal structure of $\text{Ca}_2\text{Pt}_3\text{O}_8$, and a calcium atom constitutes the triangular prism. Figure 7(b) shows the pyroxene structure of $\text{Ca}_4\text{Ir}_4\text{O}_{12}$, and a calcium atom constitutes the distorted square antiprism. Figure 7(c) shows the quadruple perovskite

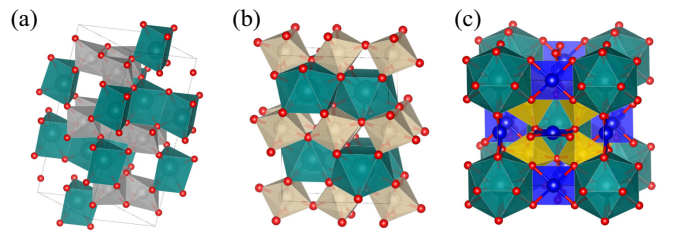


FIG. 7. (a) $\text{Na}_2\text{Mn}_3\text{O}_8$ structure of $\text{Ca}_2\text{Pt}_3\text{O}_8$. Platinum atoms are shown as silver balls. A platinum atom makes an octahedron, while a calcium atom makes a trigonal prism. (b) Pyroxene structure of $\text{Ca}_4\text{Ir}_4\text{O}_{12}$. Iridium atoms are shown as wheat balls. An iridium atom makes an octahedron, while a calcium atom makes a distorted square antiprism. (c) Quadruple perovskite structure of $\text{CaCu}_3\text{Fe}_4\text{O}_{12}$. A copper atom makes a square, an iron atom makes an octahedron, and a calcium atom makes a regular icosahedron.

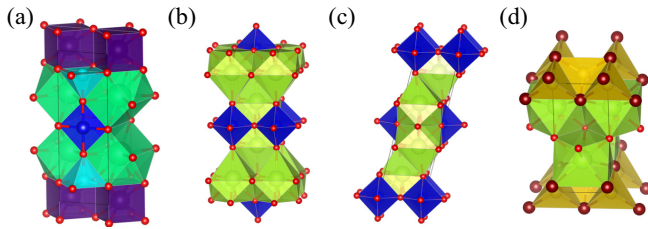


FIG. 8. (a) $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ structure of $\text{YBa}_2\text{Cu}_3\text{O}_7$. Yttrium and barium atoms are shown as indigo and spring-green balls, respectively. The former makes the cubic while the latter is surrounded by ten oxide ions. The aqua balls correspond to copper atoms which makes the square pyramid, while the blue balls correspond to copper atoms which makes the square. (b) K_2NiF_4 structure of $\text{La}_4\text{Cu}_2\text{O}_8$. Lanthanum atoms are shown as green-yellow balls. A lanthanum atom makes a capped square antiprism, while a copper atom makes an octahedron. (c) Distorted K_2NiF_4 structure with $C2/m$ symmetry. Lanthanum atoms are shown as green-yellow balls. A lanthanum atom makes a distorted capped square antiprism, while a copper atom makes an octahedron. (d) ZrCuSiAs structure of $\text{La}_2\text{Fe}_2\text{As}_2\text{O}_2$. Arsenic atoms are shown as dark-red balls. A lanthanum atom makes a square antiprism, while an iron atom makes a tetrahedron.

structure of $\text{CaCu}_3\text{Fe}_4\text{O}_{12}$, and a calcium atom constitutes the regular icosahedron. These results indicate that the mathematical crystal chemistry can reproduce a wide variety of coordination polyhedra resulting in making many kinds of crystal structures.

Previous algorithm was hard to reproduce the square coordinations around Cu atoms, but a copper atom in the quadruple perovikite structure makes the square coordination owing to large $r^{(C)}$. This study also finds the crystal structure of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ shown in Fig. 8(a). Previous study found the K_2NiF_4 structure shown in Fig. 8(b) as the unique optimal solution, but this study additionally finds the distorted K_2NiF_4 structure shown in Fig. 8(c) due to small $r^{(C)}$. Finally, the crystal structure of LaFeAsO shown in Fig. 8(d) can also be found as unique optimal solution. Since this structure contains two kinds of anions, we can expect that mathematical crystal chemistry has the capability of crystal structure prediction of not only oxides but also mixed-anion compounds [38].

Despite the successful reproduction of crystal structures, computational cost of the algorithm is found to be larger than the previous one [32]. Throughout the optimization process, assignment of geometrical constraints is updated depending on the spatial order of atoms many times, however, in most cases, some of assigned geometrical constraints cannot be feasible. One of the main reasons of the infeasibility comes from the infeasible linking of polyhedra, but the algorithm in this study cannot recognize them before the local structural optimization which is aimed at judging whether all the interatomic distances can be feasible or not. Note that the local structural optimization begins just after all the cations satisfy the constraints on the compositions of the coordination polyhedra. Therefore, the inequality constraint of Eq. (20), which can be determined in advance based on the atomic radii, should be included in Eq. (19) to avoid the wasteful computations as done in the previous study [32].

Mathematical optimization problem given in Eq. (17) has the capability to include additional constraints, such as crystal systems or space groups obtained from x-ray diffraction. Initial structure generations under such constraints would reduce the number of discrete optimizations that are aimed at finding the feasible assignments of geometrical constraint. Therefore, the additively constrained optimization problem is expected to be advantageous not only for the discovery of complex crystal structures but also the determination of crystal structures based on experimental data as one of the optimal solutions.

Mathematical crystal chemistry is not easy to reproduce the “accurate” crystal structures of materials; for example, the packings of flexible atomic spheres seem to be incapable of modeling the Jahn-Teller effect. However, *ab initio* simulations are easy to optimize a designed crystal structure prototype into the “accurate” crystal structure with evaluating the standard enthalpy of formation. The mathematical optimization model in this study enables the exhaustive search for crystal structure prototypes from the huge coordination space, leading to the discovery of unknown crystal structures. Therefore, we can expect that crystal structure predictions based on mathematical crystal chemistry leads to the computational discovery of unknown materials by integrating *ab initio* simulations.

B. Analysis on oxide crystals registered in ICSD

Metal oxides exhibit a wide variety of physical properties, such as high-temperature superconductivity and magnetism. Since ICSD contains a lot of crystal structures of metal oxides, this study analyzes whether they can be the optimal solutions by assigning suitable coordination features to every atom. First, metal oxides satisfying the five following conditions are collected from the ICSD database:

- (1) The number of metal atoms is not more than that of oxide ions
- (2) No anions except for oxide ions
- (3) No hydrogen atoms
- (4) No partial occupation of atom sites
- (5) The number of atoms per unit cell is not more than hundred

Second, the different types of them are collected by referencing structure fingerprints. As a result, 6475 structure types are discovered.

Since structural reoptimization after the assignment of coordination features transforms the experimental structures, the degree of structural similarity between two structures are defined as listed in Table II based on the previous study on inorganic structure types [39,40]. Note that even if two structures are homeotypic, they do not necessarily have the same structure type since in some cases a reoptimized structure is largely deformed. However, if the best degree of structural similarity is isopointal, then we can expect that the experimental and reoptimized structures have the same structure type, because this case occurs when an experimental structure is too distorted due to, for example, the Jahn-Teller effect; since the coordination number is underestimated due to the stretched ionic bonds, the experimental and reoptimized structures cannot be isotypic due to the increased coordination numbers

TABLE II. Definitions of the degree of structural similarity between two structures. In addition to the space group type and the Wyckoff sequence, which are utilized in the previous studies for identifying inorganic structural types [39,40], the geometrical constraints are also utilized to identify the coordination numbers and the connections of coordination polyhedra. The structure fingerprint (SF) and the simple structure fingerprint (SSF) are defined in Appendix D.

Coincidence	Definition
Isotypic	Optimal solution has the same SF.
Homeotypic	Optimal solution has the different space group type or site symmetries, but the same SSF.
Isopointal	Optimal solution has the same space group type and site symmetries, but different SSF.
Optimal	Optimal solution has a different space group type and SSF.
Infeasible	An infeasible solution.

of the reoptimized structure. Finally, if two structures are optimal, then in many cases the optimized structure is largely deformed.

The flowchart to search the best degree of structural similarity between experimental and reoptimized structures is shown in Fig. 9. First, the promising combinations of coordination features are estimated by the algorithm discussed in Appendix E. Note that the number of possible combinations of coordination features is too huge due to 25 kinds of discretized feasible atomic radii defined in Appendix A. Second, a promising combination of coordination features is analyzed whether the experimental structure can be an optimal solution after reoptimization. If the reoptimized structure is an optimal solution, then the experimental structure is additionally reoptimized 20 times after randomly fluctuating atomic position. The best degree of structural similarity is determined as the worst one in the 20 steps.

The result is shown in Table III. A great majority of the 6475 kinds of structure type have the optimal solutions preserving the experimental structures regardless of the sizes of the random fluctuations of the atomic positions. When the maximum size is set to be zero, 78%, 12%, and 4% of them

have the isotypic, homeotypic, or isopointal optimal structures, respectively, despite the discrete feasible radii listed in Table IV. The successful results supports assumption that appropriate discrete radii are enough to model the inorganic crystal structures; in other words, the feasibility of crystal structure is not sensitive to the minimum and maximum interatomic distances of the geometrical constraints. We can expect that not only oxide crystals but also the crystal structures of chalcogenides, halides, and mixed-anion compounds [38] can also be the optimal solutions of the mathematical optimization problem formalizing inorganic structural chemistry, since they have the structural similarity with oxide crystals.

V. CONCLUSION

Mathematical crystal chemistry formalizes the empirical rules of inorganic structural chemistry as the mathematical optimization problem. The structural aspects of crystal structures such as linked polyhedra and packing of flexible atomic spheres result from the objective and constraint functions in the optimization problem. Based on the previous study that defined interatomic forces as the geometrical constraints

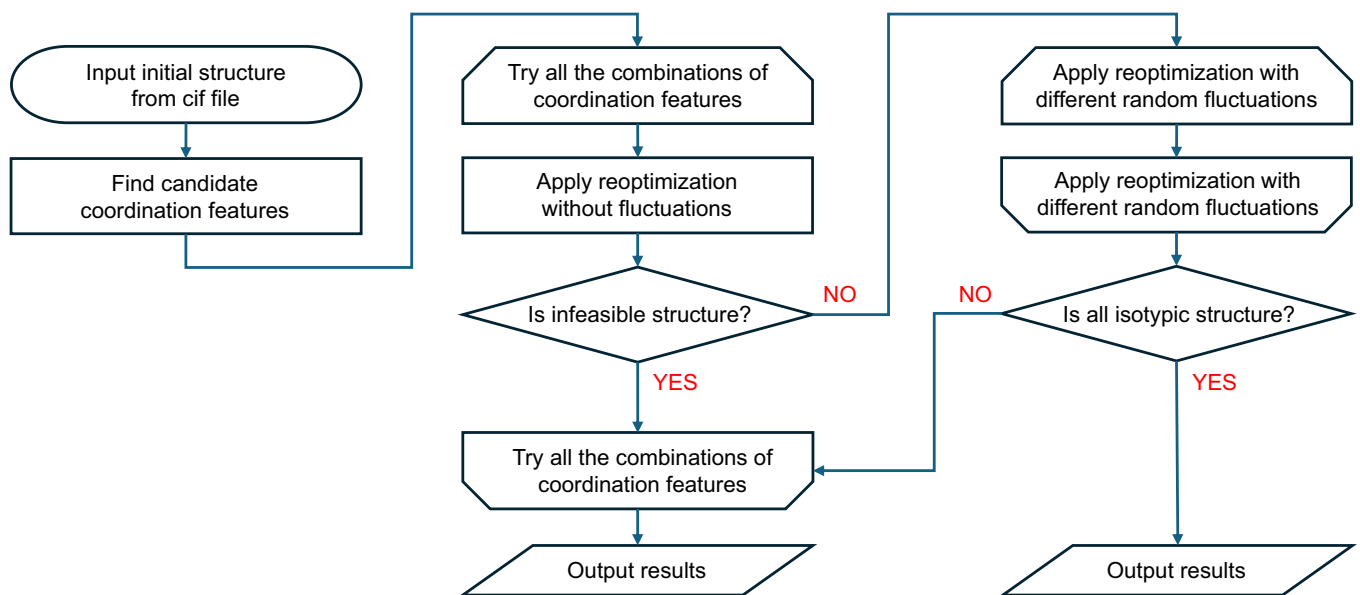


FIG. 9. Flow chart of the scheme to find the best degree of structural similarity between the experimental and reoptimized structures. All the promising combination of coordination features are analyzed whether the reoptimized structure can be an optimal solution with structural similarity until the isotypic optimal solution is found.

TABLE III. Numbers of structure types whose best structural similarities are isotypic, homeotypic, isopointal, optimal, infeasible, or exceptional, are listed. Each structure type of metal oxides is analyzed whether it can be the optimal solution of Eq. (17) preserving the experimental structure by assigning the suitable coordination feature to each atom, and the best degree of structural similarity between the experimental and reoptimized structure is determined. Each structure type is reoptimized after random fluctuation on the positions of atoms, where the maximum displacement size of each atom $\Delta y_i^{(\max)}$ is set to be 0 or $0.01R_i^{(Z)}$. Note that if an exceptional error such as a negative atomic radius arises during the analysis, then the structure is regarded as an exceptional structure.

Degree of structural similarity	Number of structure types ($\Delta y_i^{(\max)} = 0$)	Number of structure types ($\Delta y_i^{(\max)} = 0.01R_i^{(Z)}$)
Isotypic	5077	4826
Homeotypic	778	885
Isopointal	280	309
Optimal	11	20
Infeasible	141	302
Exceptional	188	133

keeping interatomic distances in stable ranges [32], this study provides a more direct extension of inorganic structural chemistry to refine the theoretical foundation of mathematical

TABLE IV. Discretized feasible atomic radii in this study. The values are the minimum radii so that the central cation can be in contact with all the surrounding n oxygen ions. The symbols of the feasible radii are determined by the number prefixes derived from Latin, Greek, or English to avoid same symbols.

Symbol	Radius	M.C.N
O	$C^{(O)} = 1.400000$	
Z	$C^{(Z)} = 0.000000$	
T	$C^{(T)} = 0.216581$	3
Q	$C^{(Q)} = 0.314643$	4
P	$C^{(P)} = 0.579899$	5
H	$C^{(H)} = 0.579899$	6
S	$C^{(S)} = 0.827755$	7
E	$C^{(E)} = 0.903460$	8
N	$C^{(N)} = 1.024871$	9
D	$C^{(D)} = 1.165450$	10
Ud	$C^{(Ud)} = 1.262958$	11
Wd	$C^{(Wd)} = 1.262958$	12
Td	$C^{(Td)} = 1.527604$	13
Qd	$C^{(Qd)} = 1.598300$	14
Pd	$C^{(Pd)} = 1.701956$	15
Hd	$C^{(Hd)} = 1.779744$	16
Sd	$C^{(Sd)} = 1.846584$	17
Ed	$C^{(Ed)} = 1.940422$	18
Nd	$C^{(Nd)} = 2.062955$	19
V	$C^{(V)} = 2.080892$	20
Uv	$C^{(Uv)} = 2.21177$	21
Wv	$C^{(Wv)} = 2.27852$	22
Tv	$C^{(Tv)} = 2.36085$	23
Qv	$C^{(Qv)} = 2.36240$	24
Pv	$C^{(Pv)} = 2.53936$	25

crystal chemistry. First, the minimum and maximum values of the feasible interatomic distance are explicitly defined by the sums of minimum and maximum atomic radii depending on the types of geometrical constraints, respectively, where the geometrical constraints are defined according to the types of interatomic forces based on inorganic structural chemistry. These radii highlights the aspect that crystal structures are packings of flexible atomic spheres according to multiple types of the atomic radii as shown in Fig. 1. The definition is an extension of inorganic structural chemistry which estimates a bonding distance by the sum of the effective sizes of atoms depending on the types of bonding forces. Second, in place of the explicit constraints on the maximum number of common bridging anions, minimum feasible distances between cations implicitly constrain the connection of coordination polyhedra, resulting in the consistency that the constraints on the interatomic distances shape the spatial order of the atoms. Note that inorganic structural chemistry considers that highly charged cations prefer the vertex sharing to edge or face sharing of coordination polyhedra to reduce the electrostatic repulsion by lengthening the distance between cations. Third, discrete variables and functions are defined to constrain the feasible atomic environment of each atom, for example, the assigned type of geometrical constraints between every pair of atoms is expressed by the discrete variable, and the rule of assignment is formalized by the inequality constraints on the discrete variables with introducing the creatability function. This study gives the simple definition of the creatability function, which returns one if the geometrical constraint is creatable but zero otherwise as activation functions in neural networks. Optimizing the assignment of the types of geometrical constraints depending on the spatial order of the atoms guides a randomly generated structure in the infeasible region toward one of the optimal solutions by breaking the barriers between local minima of the continuous optimization problem that is the dense packing problem of flexible atomic spheres.

The framework makes it possible to unify the paired mathematical optimization problem derived in the previous study [32], resulting in the identification of a wide variety of crystal structures including the spinel, Mg_2SiO_4 , β - Mg_2SiO_4 , Ga_2O_3 , $CaMgSi_2O_6$, pyrochlore (α and β), weberite, $Na_2Mg_3O_8$, pyroxene, quadruple perovskite, $YBa_2Cu_3O_{7-x}$, K_2NiF_4 , and $ZrCuSiAs$ structures. Owing to the effectiveness of the strict constraints on the minimum distances between cations, a majority of them can be found as unique optimal solutions, and besides, the square coordination of Cu atom is successfully reproduced. The result shows that the constraints on distances between cations are as dominant as the ionic bonds and the dense packings of anions when modeling ionic crystal structures. The formalization has some problems, such as increased computational cost due to the implicit of the constraints on the connection of coordination polyhedra, which can eliminate unpromising networks of ionic bonds. However, they can explicitly be included in the maximization problem of the number of ionic bonds, which is solved every several steps of the global structural optimization, so that random structure search is accelerated. Furthermore, oxide crystals registered in ICSD are analyzed whether they can be the optimal solutions. As a result, this study finds that up to 95% of oxide crystal structure types align with the optimal

solutions preserving the experimental structures, and at least 83% of them have the same symmetry and atomic environments despite the discretized feasible atomic radii.

The successful results strongly imply that the mathematical optimization problem deserves to be the theoretical foundation of mathematical crystal chemistry enabling systematic discovery of crystal structures of not only oxides but also, for example, chalcogenides and mixed-anion compounds with small computations. Crystal structure prediction can be regarded as the problem of how to simultaneously maximize the packing fractions according to the multiple types of the atomic radii, but the creatability functions express another aspect that crystal structures are linked polyhedra by creating ionic bonds between suitable pairs of an anion and a cation. This study can be regarded as the precursor of mathematical crystal chemistry which is the study of how to formalize the empirical rules of inorganic structural chemistry as the objective or constraint functions in mathematical optimization problem.

ACKNOWLEDGMENTS

This work is supported by The Kazuchika Okura Memorial Foundation. This computation was conducted using FUJITSU Server PRIMERGY GX2570 (Wisteria/BDEC-01) at the Information Technology Center, The University of Tokyo, and the facilities (supercomputer Ohtaka) of the Supercomputer Center, the Institute for Solid State Physics, the University of Tokyo. The author thanks Prof. T. Ozaki, T. Yajima, D. Hirai, J. Yamaura, and Y. Okamoto for kindly discussing on mathematical crystal chemistry. Finally, the author also thanks Profs. S. Iwata and Y. Kanno for discussing mathematical aspects.

DATA AVAILABILITY

The algorithm is implemented in the open-source program package [41]. The author declares that the data supporting the findings of this study are contained within this paper, and all results reported in this work can be reproduced by this code.

APPENDIX A: DISCRETIZED FEASIBLE ATOMIC RADII

Inorganic structural chemistry considers that all anions surrounding the central cation must be in contact with the central cation for electrostatic stability. In fact, larger cations in ionic crystals tend to be surrounded by more oxide ions. Accordingly, the minimum central radius that can be surrounded by n oxide ion is defined as the optimal solution of the optimization problem given by

$$\begin{aligned} &\text{minimize } C_0, \\ &\text{subject to } |\mathbf{x}_j - \mathbf{x}_i| = C_i + C_j, \end{aligned} \quad (\text{A1})$$

where C_0 and $C_k = C^{(O)}$ ($k \neq 0$) are the radii of the central cation and oxide ions, respectively. The optimal solutions of Eq. (A1) are listed in Table IV and the examples of the corresponding coordination polyhedra are shown in Figs. 10(a)–10(c). The central cations of the three kinds of polyhedra are the “hard” spheres whose minimum and maximum ionic bond radii are the same, but the diversity of coordination polyhedra are enhanced by “soft” spheres whose

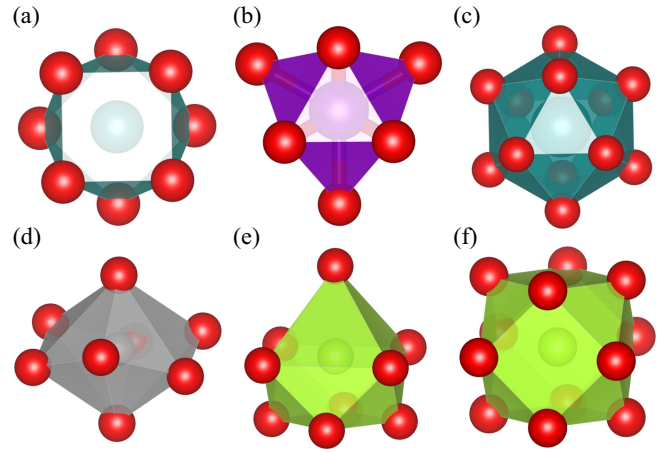


FIG. 10. A wide variety of coordination polyhedra can be generated by changing $r^{(I)}$, $R^{(I)}$, and/or the coordination number. (a) Square antiprism: The central cation with $r^{(I)} = R^{(I)} = C^{(E)}$ is surrounded by 8 oxide ions. (b) Triply capped trigonal prism: The central cation with $r^{(I)} = R^{(I)} = C^{(N)}$ is surrounded by 9 oxide ions. (c) Regular icosahedron: The central cation with $r^{(I)} = R^{(I)} = C^{(Wd)}$ is surrounded by 12 oxide ions. (d) Distorted hexagonal bipyramid: The central cation with $r^{(I)} = C^{(S)}$ and $R^{(I)} = C^{(Wd)}$ is surrounded by 8 oxide ions. This coordination polyhedra can be seen in the α -pyrochlore structure shown in Fig. 6(a). (e) Capped square antiprism: The central cation with $r^{(I)} = R^{(I)} = C^{(O)}$ is surrounded by 9 oxide ions. (f) Cuboctahedron: The central cation with $r^{(I)} = R^{(I)} = C^{(O)}$ is surrounded by 12 oxide ions.

minimum ionic bond radii are less than the maximum ionic bond radii. For example, the distorted hexagonal bipyramid shown in Fig. 10(d) can be generated by an atom whose minimum and maximum ionic bond radii are $C^{(S)}$ and $C^{(Wd)}$, respectively. Besides, even if the coordination numbers are the same, two hard spheres having the different radii can generate different coordination polyhedra. For example, the hard cation whose ionic bond radii is $C^{(Wd)}$ generates the regular icosahedron shown in Fig. 10(b), while the cation whose ionic bond radii is $C^{(O)}$ generates the cuboctahedron shown in Fig. 10(f). However, even if the sizes of the two “hard” spheres are the same, they can generate different coordination polyhedra if the coordination numbers are different. For example, both the capped square antiprism shown in Fig. 10(e) and cuboctahedron shown in Fig. 10(f) can be generated by “hard” spheres whose ionic bond radii are $C^{(O)}$. In summary, only the three simple parameters of the minimum ionic bond radius, maximum ionic bond radius, and the coordination number enable us to generate a wide variety of coordination polyhedra. In this study, all the atomic radii utilized in this study are discretized so that they have to be equal to one of the feasible radii.

APPENDIX B: 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

$$1 \leq b, c \leq 3, \quad \frac{1}{3}\pi \leq \alpha, \beta, \gamma \leq \frac{2}{3}\pi, \quad (\text{B1})$$

TABLE V. Empirically determined geometrical optimization parameters.

Parameter	Value for global optimization	Value for local optimization	Value for precise optimization
S	100	2000	4000
Initial $\Delta x_s^{(\max)}$	0.60	0.3	0.1
Final $\Delta x_s^{(\max)}$	0.59999	0.05	0.005
ε	1.0	0.2	0.05

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

$$\frac{1}{|\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)|} \left[\sum_{i=1}^M \frac{4}{3} \pi (R_i^{(l)})^3 \right] = 0.7. \quad (\text{B2})$$

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

APPENDIX C: STRUCTURAL OPTIMIZATION ALGORITHM

The structural optimization algorithm devised in the previous study [32] is summarized to present the paper in a self-contained manner. The inequality constraints are approximated by the hard-spherical potential as

$$d_\sigma \leq x_\sigma \Rightarrow U_{\min}(x_\sigma) \equiv \max[0, k_\downarrow(d_{ij} - x_\sigma)], \quad (\text{C1})$$

$$x_\sigma \leq D_\sigma \Rightarrow U_{\max}(x_\sigma) \equiv \max[0, k_\uparrow(x_\sigma - D_{ij})], \quad (\text{C2})$$

where $k_\downarrow$ and $k_\uparrow$ are a common constant for repulsive and attractive forces, respectively. Accordingly, the structural optimization problem is given by

$$\text{minimize } H(\mathbf{X}), \quad (\text{C3})$$

where the objective function $H(\mathbf{X})$ is defined as

$$H(\mathbf{X}) \equiv \sum_{\sigma} [U_{\min}(x_\sigma) + U_{\max}(x_\sigma)] + P\Omega, \quad (\text{C4})$$

with P being the pressure. The minimization problem is solved by iterative-balance methods [32,42–44] as follows: Let $\Delta \mathbf{x}_i$ and $\Delta \mathbf{t}_i$ be the displacements of $\mathbf{x}_i$ and $\mathbf{t}_i$ calculated by

$$\Delta \mathbf{x}_i = -\xi_i \frac{\partial H(\mathbf{X})}{\partial \mathbf{x}_i}, \quad \Delta \mathbf{t}_i = -\zeta_i \frac{\partial H(\mathbf{X})}{\partial \mathbf{t}_i}, \quad (\text{C5})$$

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

$$|\Delta \mathbf{x}_i| \leq \Delta x_s^{(\max)}, \quad |\Delta \mathbf{t}_i| \leq \gamma \Delta x_s^{(\max)}, \quad (\text{C6})$$

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

$$\Delta x_s^{(\max)} = \Delta x_0^{(\max)} \left(\frac{\Delta x_S^{(\max)}}{\Delta x_0^{(\max)}} \right)^{\frac{s}{S}}, \quad (\text{C7})$$

where S is the maximum number of optimization steps, and γ is defined as

$$\gamma = 0.02L \cdot \Delta x_s^{(\max)}, \quad (\text{C8})$$

with L being the number of atoms per unit cell. As $\Delta x_s^{(\max)}$ is gradually reduced, an interatomic distance is oscillating

around the feasible region of value. Therefore, even if the assignment of geometrical constraints is correct, an interatomic distance may be regarded as infeasible when $\Delta x_s^{(\max)}$ is larger than the feasible error εd_σ . Accordingly, a dynamic error rate is necessary to recognize infeasible assignments of geometrical constraints with small computation.

In this study, the force constants are set to be

$$P = 1.0, \quad k_\uparrow = 30.0, \quad k_\downarrow = -100.0. \quad (\text{C9})$$

In the global optimization, $\mathbf{n}$ is optimized every 20 structural optimization steps, and the unit cell is refined per 1000 steps by using the software SPGLIB [45]. In the local optimization, if a structure cannot be an optimal solution after the first 2000 steps, then the structure is optimized again after initializing $\Delta x_s^{(\max)}$, because in some cases, not all the interatomic distances converge into the feasible regions. The other parameters used in the structural optimization are listed in Table V. Note that the error rate ε is set to be zero throughout the structural optimization.

The cutoff distances of the ionic bonds are set to be $2D_\sigma^{(l)}$ to reduce the computational cost. Similarly, the cutoff radii of the nonionic bonds, anionic constraints, cationic constraints, and metallic constraints are set to be $2d_\sigma^{(G)}$.

APPENDIX D: STRUCTURE FINGERPRINT

The previous study uses several criterias, such as the space group types, Wyckoff sequence, c/a ratio, and β ranges, to identify the inorganic structure types [39,40]. In many cases, the space group types and Wyckoff sequence are enough to distinguish two crystal structures. Based on the insight into the inorganic crystal structures, the structure fingerprint (SF) is defined by not only the symmetry of the crystal but also the geometrically constrained state. SF is specified by the space group number and the list of the atomic environments (AEs), where AE is defined by the subset the atom belongs, the site symmetry symbol, the number of ionic bonds, and the connection of polyhedra which can be recognized from the geometrical constraints. Additionally, the simplified structure fingerprint (SSF) is defined as the list of simplified atomic environment which is defined as the AE without the site symmetry symbol. In this study, the space groups are determined by using SPGLIB [45].

APPENDIX E: RULE TO ESTIMATE THE PROMISING COORDINATION FEATURES FOR ANALYSIS OF OXIDE CRYSTALS

In the analysis on the oxide crystals registered in ICSD, the promising ranges of $R^{(l)}$, $r^{(l)}$, and $r^{(C)}$ of each atom site are assumed following the flowchart shown in Fig. 11. The

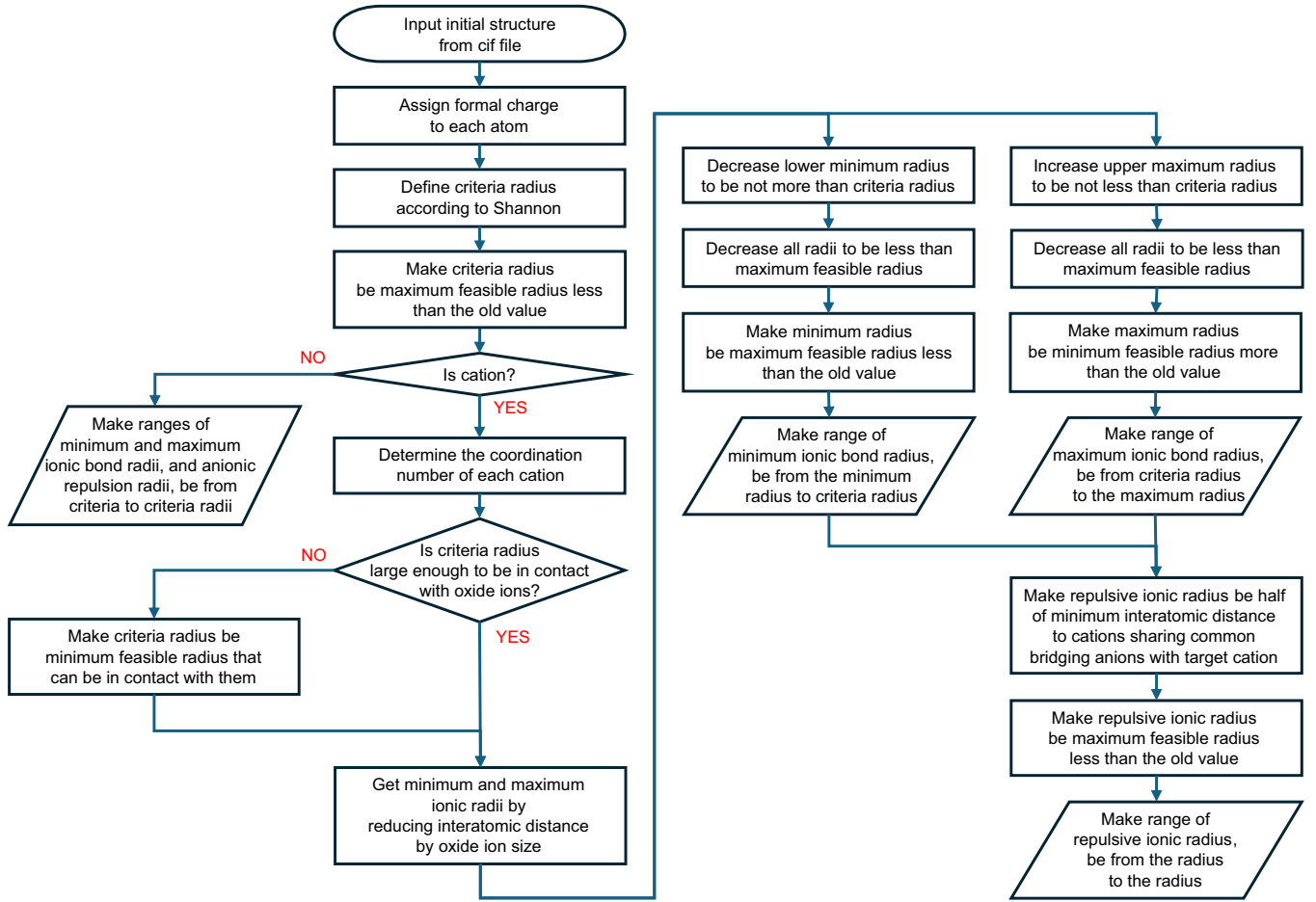


FIG. 11. Flow chart of the scheme to estimate the promising combinations of coordination features. The promising ranges of $r^{(l)}$, $R^{(l)}$, and $r^{(c)}$ of each atom site are assumed based on the ionic radii by Shannon [47].

formal charge of each atom is determined so that the sum of ionization potentials are minimized:

$$\begin{aligned} & \text{minimize} \quad \sum_{i=1}^M \sum_{n=0}^{q_i} I_i^{(n)} \chi_i^{(\mathcal{M})}, \\ & \text{subject to} \quad q_i^{(\min)} \leq q_i \leq q_i^{(\max)}, \\ & \quad |q_i - q_j| \chi_i^{(\mathcal{A})} \chi_j^{(\mathcal{A})} \leq 1, \end{aligned} \quad (\text{E1})$$

where $I_i^{(n)}$ is the experimentally determined n th ionization potential [46], q_i is the formal charge of atom i , and $q_i^{(\min)}$ and $q_i^{(\max)}$ are the minimum and maximum formal charges of atom i , respectively. Note that the second constraint makes the anions be evenly reduced.

If the number of the combinations of the coordination features derived from the algorithm shown in Fig. 11 is more than 100 000, the range of $r^{(l)}$ is limited to the criteria radius defined in Fig. 11. If it does not work, then the range of $R^{(l)}$ is also limited to the criteria radius defined in Fig. 11. However, if the number of the combinations is not more than 100 000, then the ranges of $R^{(l)}$ and $r^{(c)}$ are extended as follows: First, the range of $R^{(l)}$ is extended so that the maximum coordination number can be two larger unless $C^{(\text{Pv})} < R^{(l)}$. Second, the range of $r^{(c)}$ is extended so that it contains one additional feasible radius which is lowest in the feasible radii more than the criteria radius. However, these extensions of the ranges are stopped if the number of the combinations of the coordination features is more than 100 000. Finally, $r^{(M)}$ of each cation is set to be $C^{(O)}$.

- [1] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, The inorganic crystal structure data base, *J. Chem. Inf. Comput. Sci.* **23**, 66 (1983).
 [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] *Computational Materials Discovery*, edited by A. R. Oganov, G. Saleh, and A. G. Kvashnin (The Royal Society of Chemistry, London, UK, 2019).
- [6] A. R. Oganov and C. W. Glass, Crystal structure prediction using *ab initio* evolutionary techniques: Principles and applications, *J. Chem. Phys.* **124**, 244704 (2006).
- [7] X. Liu, H. Niu, and A. R. Oganov, Copex: co-evolutionary crystal structure prediction algorithm for complex systems, *npj Comput. Mater.* **7**, 199 (2021).
- [8] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Crystal structure prediction via particle-swarm optimization, *Phys. Rev. B* **82**, 094116 (2010).
- [9] 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).
- [10] F. Peng, Y. Sun, C. J. Pickard, R. J. Needs, Q. Wu, and Y. Ma, Hydrogen clathrate structures in rare earth hydrides at high pressures: Possible route to room-temperature superconductivity, *Phys. Rev. Lett.* **119**, 107001 (2017).
- [11] H. Liu, I. I. Naumov, R. Hoffmann, N. W. Ashcroft, and R. J. Hemley, Potential high-T_c superconducting lanthanum and yttrium hydrides at high pressure, *Proc. Natl. Acad. Sci. USA* **114**, 6990 (2017).
- [12] Z. M. Geballe, H. Liu, A. K. Mishra, M. Ahart, M. Somayazulu, Y. Meng, M. Baldini, and R. J. Hemley, Synthesis and stability of lanthanum superhydrides, *Angew. Chem. Int. Ed.* **57**, 688 (2018).
- [13] A. P. Drozdov, P. P. Kong, V. S. Minkov, S. P. Besedin, M. A. Kuzovnikov, S. Mozaffari, L. Balicas, F. F. Balakirev, D. E. Graf, V. B. Prakapenka, E. Greenberg, D. A. Knyazev, M. Tkacz, and M. I. Erements, Superconductivity at 250 K in lanthanum hydride under high pressures, *Nature (London)* **569**, 528 (2019).
- [14] M. Somayazulu, M. Ahart, A. K. Mishra, Z. M. Geballe, M. Baldini, Y. Meng, V. V. Struzhkin, and R. J. Hemley, Evidence for superconductivity above 260 K in lanthanum superhydride at megabar pressures, *Phys. Rev. Lett.* **122**, 027001 (2019).
- [15] V. Struzhkin, B. Li, C. Ji, X.-J. Chen, V. Prakapenka, E. Greenberg, I. Troyan, A. Gavriluk, and H.-k. Mao, Superconductivity in La and Y hydrides: Remaining questions to experiment and theory, *Matter Radiat. Extremes* **5**, 028201 (2020).
- [16] I. A. Troyan, D. V. Semenov, A. G. Kvashnin, A. V. Sadakov, O. A. Sobolevskiy, V. M. Pudalov, A. G. Ivanova, V. B. Prakapenka, E. Greenberg, A. G. Gavriluk, I. S. Lyubutin, V. V. Struzhkin, A. Bergara, I. Errea, R. Bianco, M. Calandra, F. Mauri, L. Monacelli, R. Akashi, and A. R. Oganov, Anomalous high-temperature superconductivity in YH₆, *Adv. Mater.* **33**, 2006832 (2021).
- [17] 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).
- [18] 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).
- [19] 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).
- [20] 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).
- [21] J. Cheng, C. Zhang, and L. Dong, A geometric-information-enhanced crystal graph network for predicting properties of materials, *Commun. Mater.* **2**, 92 (2021).
- [22] V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, Benchmarking graph neural networks for materials chemistry, *npj Comput. Mater.* **7**, 84 (2021).
- [23] K. Choudhary and B. DeCost, Atomistic line graph neural network for improved materials property predictions, *npj Comput. Mater.* **7**, 185 (2021).
- [24] 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).
- [25] 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).
- [26] 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).
- [27] G. Cheng, X.-G. Gong, and W.-J. Yin, Crystal structure prediction by combining graph network and optimization algorithm, *Nat. Commun.* **13**, 1492 (2022).
- [28] 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).
- [29] 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, H. Yang *et al.*, A generative model for inorganic materials design, *Nature (London)* **639**, 624 (2025).
- [30] 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).
- [31] 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).
- [32] R. Koshiji and T. Ozaki, Mathematical crystal chemistry: A mathematical programming approach to crystal structures of inorganic compounds, *Phys. Rev. Mater.* **8**, 113801 (2024).
- [33] U. Müller, *Inorganic Structural Chemistry*, 2nd ed. (John Wiley & Sons, Ltd, New York, NY, 2007).
- [34] A. F. Wells, *Structural Inorganic Chemistry*, 5th ed. (Oxford University Press, Oxford, UK, 1984).
- [35] L. Pauling, The principles determining the structure of complex ionic crystals, *J. Am. Chem. Soc.* **51**, 1010 (1929).
- [36] T. C. Hales, A proof of the Kepler conjecture, *Ann. Math.* **162**, 1065 (2005).
- [37] K. Momma and F. Izumi, VESTA3 for three-dimensional visualization of crystal, volumetric and morphology data, *J. Appl. Crystallogr.* **44**, 1272 (2011).
- [38] H. Kageyama, H. Ogino, T. Zhu, and T. Hasegawa, *Mixed-anion Compounds* (Royal Society of Chemistry, London, UK, 2024).
- [39] R. Allmann and R. Hinek, The introduction of structure types into the inorganic crystal structure database ICSD, *Acta Crystallogr. Sect. A* **63**, 412 (2007).

- [40] J. Lima-de Faria, E. Hellner, F. Liebau, E. Makovicky, and E. Parthé, Nomenclature of inorganic structure types: Report of the international union of crystallography commission on crystallographic nomenclature subcommittee on the nomenclature of inorganic structure types, *Acta Crystallogr. Sect. A* **46**, 1 (1990).
- [41] The software to produce structural prototypes of ionic compounds and analyze crystal structures is available at <https://aist.repo.nii.ac.jp/records/2003344>.
- [42] R. Koshiji, M. Kawamura, M. Fukuda, and T. Ozaki, Diverse densest binary sphere packings and phase diagram, *Phys. Rev. E* **103**, 023307 (2021).
- [43] R. Koshiji and T. Ozaki, Densest ternary sphere packings, *Phys. Rev. E* **104**, 024101 (2021).
- [44] R. Koshiji and T. Ozaki, Diverse densest ternary sphere packings, *J. Phys. Commun.* **6**, 075002 (2022).
- [45] K. S. Atsushi Togo and I. Tanaka, Spglib: A software library for crystal symmetry search, *Sci. Technol. Adv. Mater.: Methods* **4**, 2384822 (2024).
- [46] S. S. Batsanov and A. S. Batsanov, *Introduction to Structural Chemistry* (Springer, Berlin, 2012).
- [47] R. D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, *Acta Crystallogr. Sect. A* **32**, 751 (1976).