Mathematical Crystal Chemistry

Ryotaro Koshoji*

National Institute of Advanced Industrial Science and Technology (AIST), Nagoya, 463-8560, Japan

Taisuke Ozaki[†]

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

(Dated: April 25, 2024)

Efficient heuristics have predicted many functional materials such as high-temperature superconducting hydrides, while inorganic structural chemistry explains why and how the crystal structures are stabilized. Here we develop the paired mathematical programming formalism for searching and systematizing the structural prototypes of crystals. The first is the minimization of the volume of the unit cell under the constraints of only the minimum and maximum distances between pairs of atoms. We show the capabilities of linear relaxations of inequality constraints to optimize structures by the steepest-descent method, which is computationally very efficient. The second is the discrete optimization to assign five kinds of geometrical constraints including chemical bonds for pairs of atoms. Under the constraints, the two object functions, formulated as mathematical programming, are alternately optimized to realize the given coordination numbers of atoms. This approach successfully generates a wide variety of crystal structures of oxides such as spinel, pyrochlore-α, and K_2NiF_4 structures.

Discovering advanced materials enables technological innovation in many areas such as batteries, photovoltaics, and catalysis. Since electronic structure depends on the spatial arrangements of atoms, predictions of unknown crystal structures have a major impact on the discovery of better materials. However, computational structure searching is still difficult despite the progress of computational power, since structure prediction is a large-scale nonconvex minimization problem of the total energy, which has a large number of local minima, concerning the atomic arrangement and crystal lattice.

Finding the global minimum essentially involves visiting every local minimum, but some heuristic approaches with more or less intelligent ways such as evolutionally algorithms [1–4] and particle swarm-intelligence approach [5–7] have predicted novel materials including binary hydrides that have been experimentally synthesized as high-temperature superconductors under high pressure [8–14]. The drawbacks of these methods are that there are no guarantees that the most stable structure are found, and similarly, they can miss synthesizable unique structures. Furthermore, they cannot explain why and how the predicted structures are stabilized. This is also true of the deep learning approach [15–18] and the random structure searching methods [19, 20], while the latter may lead to peculiar structures without a *priori* knowledge.

The integer programming approach has been developed to deterministically find the lowest energy structure with guarantees [21]. The global optimum of integer programming can be found deterministically by brute force with branch-and-cut algorithms that are capable of rapidly eliminating large parts of the optimization domain from consideration. To apply this universal method, all the atoms are placed at discrete positions within a fixed cubic unit cell, and the total energy is represented by the sum of two-body classical interaction potentials [21]. This approach has successfully found complex crystals such as garnet structure, but the fixed unit cells and approximated total energies may limit its applicability. Instead of using approximated total energies, geometrical approaches may be effective to search crystal structures, for example, Yokoyama *et al.* have developed the dual periodic graph approach to generate structures composed of space-filling polyhedra [22].

Important structural principles for ionic crystals have been summarized by Pauling in the five rules [23–25]. George *et al.* have shown the limited predictive power of the Pauling rules since a few oxides simultaneously satisfy them [26], but the rules offer the basic concepts for inorganic structural chemistry. For example, the crystal structures of oxides, chalcogenides, and mixed anion compounds are linked polyhedra with explicit chemical bonds, especially between anions and cations [24, 25]. Anions such as oxygen prefer to constitute the Barrow packings, which are the densest packings of spheres [27], and small cations are placed in the tetrahedral and octahedral sites, while large cations typically have a coordination number greater than 6, and in some cases, constitutes the Barrow packings with anions [24, 25]. Small and large cations generally behave as hard and soft spheres, respectively. Highly polar bonds do not favor edge-sharing polyhedra and especially face-sharing polyhedra because of the increased electrostatic repulsion between central atoms [24, 25].

The structure of amorphous semiconductors is well represented by continuous random network models which require a condition that each atom should satisfy fully its bonding needs [28–31]. Two atoms interact only if they are explicitly bonded, and the total energy is given

* r.koshoji@aist.go.jp
† t-ozaki@issp.u-tokyo.ac.jp

by the Keating potential [31, 32]. The network is reproduced periodically by bond transpositions accepted with the Metropolis acceptance probability, conserving four coordination. This approach can generate structurally and electronically good structures comparable with experiments from random initial structures.

In this study, we propose a paired mathematical programming approach that finds candidate structures of crystals satisfying a series of empirical rules. The first is the minimization problem of the volume of the unit cell within the constraints of minimum and maximum distances between pairs of atoms. The constraints represent linked polyhedra and simultaneously the packings of anionic spheres regardless of the kinds of chemical bonds. We approximate the inequality constraints as the linear potentials for the structural relaxation. The second is the maximization problem of the number of chemical bonds. The five kinds of geometrical constraints, which impose restrictions on the distances between every pair of atoms, are assigned by the network of chemical bonds. The two optimization problems are alternately solved to find a proper structure where every cation satisfies its bonding needs; the transpositions of geometrical constraints transform random initial structures into linked polyhedra as the *sillium* approach for amorphous silicons [28–31]. We define a feasible solution as a structure satisfying all the constraints in the two optimization problems and all the bonding needs of atoms, and the other cases are denoted as infeasible solution. An optimal solution in this study is defined to be the feasible solution whose volume of the unit cell is minimized. The paired mathematical programming we propose is an efficient method which tries to find the optimal solution for a model system with a given composition.

Let us start our discussion by introducing the first object function in the paired mathematical programming formalism. The geometrical constraints are composed of the minimum and maximum distances between pairs of atoms. Suppose $\boldsymbol{x}_i$ is the cartesian coordinate of atom i, and $x_{ij\boldsymbol{T}}$ is the distances between atoms i and $j\boldsymbol{T}$ given by

$$x_{ij\boldsymbol{T}} = |\boldsymbol{x}_j + \boldsymbol{T} - \boldsymbol{x}_i|, \qquad (1)$$

where the atom $j\boldsymbol{T}$ corresponds to the atom j with the translational vector $\boldsymbol{T}$ of the crystal lattice. The object function is the volume Ω of the unit cell. The mathematical programming is given by

$$
\begin{aligned}
\text{minimize} \quad & \Omega \\
\text{subject to} \quad & x_{ij\boldsymbol{T}} \geq d_{ij\boldsymbol{T}}^{(\min)}, \\
& x_{ij\boldsymbol{T}} \leq d_{ij\boldsymbol{T}}^{(\max)}
\end{aligned}
\qquad (2)
$$

where $d_{ij\boldsymbol{T}}^{(\min)}$ and $d_{ij\boldsymbol{T}}^{(\max)}$ are the minimum and maximum distances between atoms i and $j\boldsymbol{T}$, respectively. This model finds structural candidates of crystals without calculating the total energies.

To solve the problem, the equality and inequality constraints are relaxed as follows: The constraint of minimum distance between atoms i and $j\boldsymbol{T}$:

$$x_{ij\boldsymbol{T}} \geq d_{ij\boldsymbol{T}}^{(\min)}, \qquad (3)$$

is relaxed to the hard-spherical potential $U_{\min}(x_{ij\boldsymbol{T}})$ defined to be

$$U_{\min} \equiv \begin{cases} -k_\downarrow \left(x_{ij\boldsymbol{T}} - d_{ij\boldsymbol{T}}^{(\min)} \right) & x_{ij\boldsymbol{T}} < d_{ij\boldsymbol{T}}^{(\min)} \\ 0 & d_{ij\boldsymbol{T}}^{(\min)} \leq x_{ij\boldsymbol{T}} \end{cases}, \qquad (4)$$

where $k_\downarrow$ is a common constant for the minimum constraints. Similarly, the constraint of maximum distance between atoms i and $j\boldsymbol{T}$:

$$x_{ij\boldsymbol{T}} \leq d_{ij\boldsymbol{T}}^{(\max)}, \qquad (5)$$

is relaxed to the hard-spherical potential $U_{\max}(x)$ defined to be

$$U_{\max} \equiv \begin{cases} 0 & x_{ij\boldsymbol{T}} \leq d_{ij\boldsymbol{T}}^{(\max)} \\ k_\uparrow \left(x_{ij\boldsymbol{T}} - d_{ij\boldsymbol{T}}^{(\max)} \right) & d_{ij\boldsymbol{T}}^{(\max)} < x_{ij\boldsymbol{T}} \end{cases}, \qquad (6)$$

where $k_\uparrow$ is a common constant for the maximum constraints. The relaxations transform the problem of Eq. (2) into the minimization problem of the enthalpy $H(\{\boldsymbol{x}_i, \boldsymbol{a}_i\})$ per unit cell formulated as

$$\text{minimize} \quad H(\{\boldsymbol{x}_i, \boldsymbol{a}_i\}) = E(\{\boldsymbol{x}_i, \boldsymbol{a}_i\}) + P\Omega, \qquad (7)$$

where P is the pressure and $E(\{\boldsymbol{x}_i, \boldsymbol{a}_i\})$ is given by

$$E(\{\boldsymbol{x}_i, \boldsymbol{a}_i\}) = \sum_{i \leq j} \sum_{\boldsymbol{T}} \left[U_{\min}(x_{ij\boldsymbol{T}}) + U_{\max}(x_{ij\boldsymbol{T}}) \right], \quad (8)$$

and $\boldsymbol{a}_i$ is the primitive translation vectors of crystal lattice. The displacement of each atom is calculated from the derivative of the enthalpy H per unit cell as

$$\Delta\boldsymbol{x}_i = -k_i \frac{\mathrm{d}H}{\mathrm{d}\boldsymbol{x}_i}. \qquad (9)$$

The constant k_i is scaled to ensure so that the norm of displacement does not exceed the given maximum value, $\Delta x_{\max}$, for each optimization step. Similarly, the primitive lattice vectors $\boldsymbol{a}_i$ is optimized as

$$\Delta\boldsymbol{a}_i = -k_{\mathrm{L}} \frac{\mathrm{d}H}{\mathrm{d}\boldsymbol{a}_i}, \qquad (10)$$

where k_{L} is set to satisfy the condition:

$$\sqrt{|\Delta\boldsymbol{a}_1|^2 + |\Delta\boldsymbol{a}_2|^2 + |\Delta\boldsymbol{a}_3|^2} \leq \Delta A_{\max}. \qquad (11)$$

To reach the local optima of the problem given by Eq. (7), we gradually reduce $\Delta x_{\max}$ and $\Delta A_{\max}$ as

$$
\begin{aligned}
\Delta x_{\max}^{(n)} &= \alpha^{n-1} \Delta x_{\max}^{(1)}, \\
\Delta A_{\max}^{(n)} &= \frac{\Delta x_{\max}^{(n)}}{4000} \sum_i \frac{4\pi}{3} \left(r_i^{(\mathrm{I})} \right)^3,
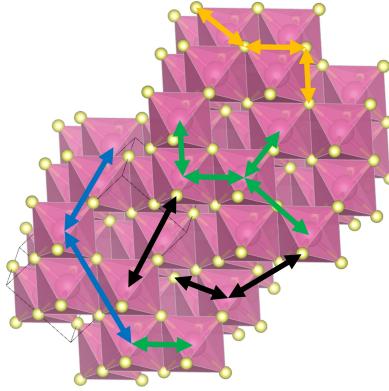\end{aligned}
\qquad (12)
$$

FIG. 1. The five kinds of geometrical constraints with the optimal solutions of Hd$_4$O$_8$ corresponding to anatase structure. Cylinders connecting to two atoms correspond to chemical bonds (CBs). Black, orange, blue, and green arrows correspond to non-bonding constraints (NBCs), anionic constraints (ACs), cationic constraints (CCs), and polyhedral constraints (PCs), respectively.

where α is a constant and n is the number of geometrical optimization steps. The values of pameters such as α, $k_\uparrow$, $k_\downarrow$, and P are discussed in Supplementary Information [33]. One may consider that structures are never relaxed due to the discontinuity in the first derivative of the potential. However, the structure converges to an optimal solution by gradually reducing the maximum $\Delta x_{\max}$ and $\Delta A_{\max}$, even though there is the discontinuity. The effectiveness of linear relaxations of inequality constraints has already been shown in the previous studies on the densest sphere packings as iterative-balance method [34–36]: The method enables the structures to reach a local optima precisely enough to calculate packing fractions. The inequality constraints are widely approximated by the logarithmic barrier functions [37], but the advantage of the linear potentials consists of the two folds: One is that the computational cost is the lowest and the other is that the potentials can impose hard penalties only when the constraints are not satisfied.

Next we turn our discussion to the second object function in the paired mathematical programming formalism. The problem is how to determine the values of $d_{ij\boldsymbol{T}}^{(\min)}$ and $d_{ij\boldsymbol{T}}^{(\max)}$. In the following discussion, we focus on the crystal structures of oxides. We introduce the five kinds of geometrical constraints as listed in Table I and illustrated in Fig. 1. The minimum and maximum distances are determined as

$$d_{ij\boldsymbol{T}}^{(\min)} = \max_{\mathrm{p}} \left\{ s_{ij\boldsymbol{T}}^{(\mathrm{p})} d_{ij\boldsymbol{T}}^{(\mathrm{p,min})} \right\}$$
$$d_{ij\boldsymbol{T}}^{(\max)} = \max_{\mathrm{p}} \left\{ s_{ij\boldsymbol{T}}^{(\mathrm{p})} d_{ij\boldsymbol{T}}^{(\mathrm{p,max})} \right\}, \quad (13)$$
$$\sum_{p} s_{ij\boldsymbol{T}}^{(\mathrm{p})} = 1$$

where $s_{ij\boldsymbol{T}}^{(\mathrm{p})} \in \{0,1\}$ is a switch variable selecting one of

the geometrical constraints, and p runs chemical bond (CB), non-bonding constraint (NBC), anionic constraint (AC), cationic constraint (CC), and polyhedral constraint (PC). The geometrical constraints for every pair of atoms are assigned depending on the network of CBs. CBs or NBCs are formed between every pair of an anion and a cation. ACs are formed between every pair of anions. Besides, if two coordination polyhedra around cations share some bridging anions, PC is formed between the two cations. Finally, CCs are formed between remained pairs of cations. To assist making coordination polyhedra, NBCs spatially separate the anion and cation when CB is not formed between them. ACs impose the hard-spherical constraint for every pair of anions; oxygen ions often constitute the Barrow-packings if the cationic radii are small. CCs are necessary to spatially separate two cations enough to avoid the condensation of cations, while a minimum distance of PC is smaller than that of CC to share a common edge or face as

$$d_{ij\boldsymbol{T}}^{(\mathrm{PC,min})} = \begin{cases} r_i^{(\mathrm{M})} + r_j^{(\mathrm{M})} & N_{ij}^{(\mathrm{CBA})} = 1 \\ D_i^{(\mathrm{shared})} + D_j^{(\mathrm{shared})} & \text{otherwise} \end{cases}, \quad (14)$$

where $D_i^{(\mathrm{shared})}$ is the minimum distance from the cation i to the shared common edge or face as detailed in Supplementary Information [33].

Since any cation i has the desirable coordination number $N_i^{(\mathrm{CB})}$ for CBs, the feasible linked-polyhedra must satisfy the conditions of

$$s_i^{(+)} \left(N_i^{(\mathrm{CB})} - n_i^{(\mathrm{CB})} \right) = 0, \quad (15)$$

where $n_i^{(\mathrm{CB})}$ is the coordination number, which is counted by the number of CBs for atom i, and $s_i^{(+)}$ is a switch function for the atom i given by

$$s_i^{(+)} = \begin{cases} 1 & \text{if the atom } i \text{ is cation} \\ 0 & \text{if the atom } i \text{ is anion} \end{cases}. \quad (16)$$

Most of random initial structures are infeasible solutions, and the analytic optimization methods such as the steepest-descent method cannot escape from the infeasible solution to a feasible solution. Therefore, we periodically update the values of $d_{ij\boldsymbol{T}}^{(\min)}$ and $d_{ij\boldsymbol{T}}^{(\max)}$ depending on the network of CBs that are also renewed periodically. The network is determined from the maximization problem of the number of CBs given by

$$\begin{aligned} \text{maximize} \quad & s_1^{(+)} \varepsilon_1 n_1^{(\mathrm{CB})} + \cdots + s_M^{(+)} \varepsilon_M n_M^{(\mathrm{CB})} \\ \text{subject to} \quad & n_i^{(\mathrm{CB})} \le N_i^{(\mathrm{CB})} \\ & n_{ij\boldsymbol{T}}^{(\mathrm{CBA})} \le N_{ij}^{(\mathrm{CBA})} \\ & s_{ij\boldsymbol{T}}^{(\mathrm{CB})} x_{ij\boldsymbol{T}} \le 2 d_{ij\boldsymbol{T}}^{(\mathrm{CB,max})} \\ & \sum_{j,\boldsymbol{T}} s_i^{(+)} s_{ij\boldsymbol{T}}^{(\mathrm{CB})} x_{ij\boldsymbol{T}} \le \min \sum_{q=1}^{n_i} x_{ij_q\boldsymbol{T}_q} \end{aligned} \quad (17)$$

TABLE I. The list of geometrical constraints. If a constraint is inactive, the mimimum and maximum distances are 0 and ∞, respectively. $r^{(\mathrm{I})}$ and $R^{(\mathrm{I})}$ are minimum and maximum ionic radii, respectively. $r^{(\mathrm{C})}$ are cationic repulsion radii.

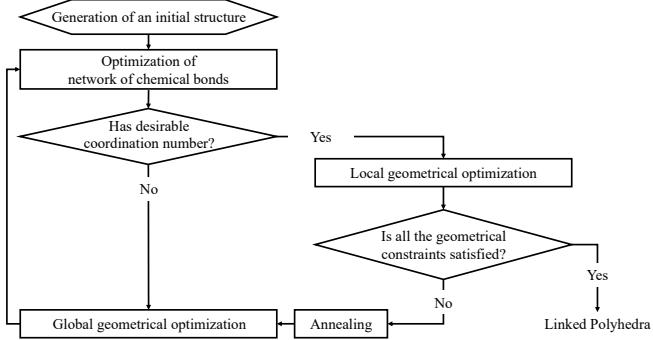| Constraint type	Abbreviation	Minimum distance if active	Maximum distance if active
Chemical Bond	CB	$d_{ij\boldsymbol{T}}^{(\mathrm{CB,min})} = r_i^{(\mathrm{I})} + r_j^{(\mathrm{I})}$	$d_{ij\boldsymbol{T}}^{(\mathrm{CB,max})} = R_i^{(\mathrm{I})} + R_j^{(\mathrm{I})}$
Non-Bonding Constraint	NBC	$d_{ij\boldsymbol{T}}^{(\mathrm{NBC,min})} = 1.2\left(R_i^{(\mathrm{I})} + R_j^{(\mathrm{I})}\right)$	$d_{ij\boldsymbol{T}}^{(\mathrm{NBC,max})} = \infty$
Anionic Constraint	AC	$d_{ij\boldsymbol{T}}^{(\mathrm{AC,min})} = R_i^{(\mathrm{I})} + R_j^{(\mathrm{I})}$	$d_{ij\boldsymbol{T}}^{(\mathrm{AC,max})} = \infty$
Cationic Constraint	CC	$d_{ij\boldsymbol{T}}^{(\mathrm{CC,min})} = r_i^{(\mathrm{C})} + r_j^{(\mathrm{C})}$	$d_{ij\boldsymbol{T}}^{(\mathrm{CC,max})} = \infty$
Polyhedral Constraint	PC	Depending on $N_{ij}^{(\mathrm{CBA})}$ as discussed in main text	$d_{ij\boldsymbol{T}}^{(\mathrm{PC,max})} = \infty$

FIG. 2. The flow chart of the optimization scheme to find the optimal solutions of the paired mathematical programming.

where $0 \leq \varepsilon_i$ is the fixed bonding affinity, $n_{ij\boldsymbol{T}}^{(\mathrm{CBA})}$ is the number of common bridging atoms between the pair of atoms i and $j\boldsymbol{T}$, $N_{ij}^{(\mathrm{CBA})}$ is the default maximum number of common bridging atoms defined in Eq. (18), and q is an index for selecting n_i indices of anions $j\boldsymbol{T}$ randomly. The third constraints define the formation ranges of CBs. The fourth constraints force cations to create CBs with the nearest n_i anions. The optimization problem can be solved as follows: First, all the cations create as many CBs as possible with neighboring anions, and seconds, we erase CBs with maximizing the object function of Eq. (17) until all the number of common bridging atoms satisfy the second constraints. Note that two polyhedra are linked by sharing a common vertex, a common edge, or a common face that corresponds to sharing one, two, or more than two common bridging atoms, respectively [24].

To find the optimal solutions, we solve the problems of Eqs. (2) and (17) alternately. The optimization scheme is shown in Fig. 2. A random initial structure can be infeasible solution, in fact, not all cations may not be surrounded by oxygen ions and may not have desirable coordination numbers. Besides, even if all the atoms have the desirable coordination numbers, the structure does not necessarily satisfy all the geometrical constraints such as the maximum number of common bridging atoms. Therefore, the global geometrical optimization is aimed at transforming the structure largely enough to create a different network of CBs by large $\Delta x_{\max}$. After the small number of global geometrical optimization steps,

we solve the problem of Eq. (17) once again to update the network of CBs. The mutable network and the other geometrical constraints assist cations making coordination polyhedra, and accordingly, the structure is transformed into a feasible solution. If the structure satisfies the condition of Eq. (15), the structure is locally optimized by small $\Delta x_{\max}$ to identify whether the structure can be the optimal solusion of the geometrical optimization problem given by Eq. (2). If so, the structure is the optimal solution of the paired optimization problems. If not, the structure is globally optimized again after annealing. The parameters of global and local optimization are detailed in Supplementary Information [33].

To estimate the number of optimal solutions and coarsely classify the optimal solutions, we define a structural fingerprint as the bundle of the fingerprints by the lists of geometrical constraints (LGCs) in dictionary order. LGCs are defined as what kinds of geometrical constraints are formed with the adjacent atoms. The fingerprint of LGC enumerates the element symbols of the adjacent atoms linked by CBs, PCs, and ACs in dictionary order, respectively. Note that there is a possibility that the same structure fingerprint is assigned to different structures. For example, the fingerprint for the rutile structure and α-PbO$_2$ is the case. See also Supplementary Information [33].

The paired mathematical programming approach is aimed at showcasting capabilities of simple rules to search structural prototypes of crystals. In this study, we introduce 11 kinds of cations shown in Table II to reproduce a wide variety of coordination polyhedra. The 11 cations are selected by considering closed packed polyhedra consisting of the centered cation and surrounding oxygen ions, and defined by $r^{(\mathrm{I})}$, $R^{(\mathrm{I})}$, $N^{(\mathrm{CB})}$, and $N^{(\mathrm{CBA})}$, where $N^{(\mathrm{CBA})}$ is the default maximum number of common bridging atoms. We determine $N_{ij}^{(\mathrm{CBA})}$ in the second constraint of Eq. (17) as

$$N_{ij}^{(\mathrm{CBA})} = \min\left\{N_i^{(\mathrm{CBA})}, N_j^{(\mathrm{CBA})}\right\}. \qquad (18)$$

Cationic radius C_n ($n \in \mathbb{N}$) corresponds to the minimum radius so that the cation can connect n oxygen ions, while C_O correponds to the ionic radius of oxygen ion. They are given in Supplementary Information [33]. If the minimum and maximum cationic radii are the same, the cation is a hard sphere; if not, the cation is a soft sphere,

TABLE II. 11 kinds of cations. C_n ($n \in \mathbb{N}$) corresponds to the minimum radius so that the cation can connect n oxygen ions, and C_O corresponds to the ionic radius of oxygen ion. The values of the radii are listed in Supplementary Information [33]. A symbol represents the atomic radius and/or the maximum number of common bridging atoms. For example, the minimum and maximum cationic radii of the cation Op are C_7 (sePta) and C_8 (Octa), respectively, while the maximum number of common bridging atoms Ht is three (Tri). The color is given if the cation is illustrated in Figures.

Symbol	$r^{(I)}$	$R^{(I)}$	$N^{(CB)}$	$N^{(CBA)}$	color
T	C_4	C_4	4	1	sky blue
Pe	C_5	C_5	5	2	-
Hd	C_6	C_6	6	2	rose
Ht	C_6	C_6	6	3	red
Sh	C_7	C_7	6	2	-
Op	C_7	C_8	7	4	dark blue
Eo	C_8	C_9	8	4	purple
Uo	C_7	C_{11}	8	4	blue
Do	C_7	C_O	8	4	-
De	C_O	C_O	9	4	light green
D	C_O	C_O	12	4	green

because the bonding lenths can vary within the two radii.

We apply our mathematical programming model to several compositions of cations, and find that a wide variety of real crystals are discovered in the optimal solutions by the paired mathematical programming. Table III shows the number of discovered optimal solutions for each composition, which seems to be much smaller than the number of local minima in the total energy by *ab-initio* simulations, and corresponding real crystals found in them. We discuss here results by focusing on aspects of the generated coordination polyhedra.

The optimal solutions of $Ht_l T_m O_n$ form a wide variety of linked-polyhedra composed of tetrahedra and octahedra as shown in Fig. 3. The cations T and Ht are placed in the tetrahedral and octahedral sites, respectively, in the Barrow-packings consisting of oxygen ions. Many structures are rejected from the optimal solutions by the constraints of the maximum number of common bridging atoms of T. The spinel structure has the lowest number of LGCs and highest symmetry, found as one of the optimal solutions for both the $Ht_4 T_2 O_8$ and $Ht_8 T_4 O_{16}$, while the $Ga_2 O_3$ structure, which is composed of 5 LGCs with the $C2/m$ symmetry, has more LGCs and lower symmetry than the structure shown in Fig. 3(d). Real crystal structures are not necessarily the optimal solution composed of the lowest number of LGCs and highest symmetry, however, they are generally good indicators to assess expectations of the structures to be realized by real crystals. The optimal solution shown in Fig. 3(c) is similar to the crystal structure of $Li_2 WO_4$; the difference comes from the sizes of the unit cells. The crystal structure of $Li_2 WO_4$ is composed of large tetrahedra and octahedra around lithiums, but this structure may be reproduced by using cations Ht and T, which are placed
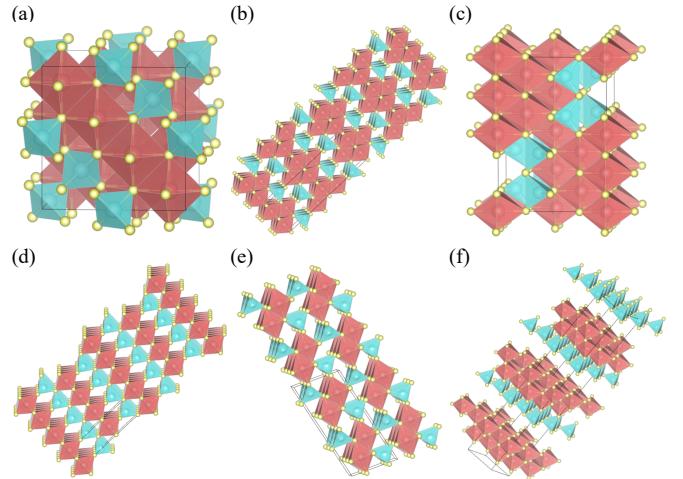

FIG. 3. (a) $Ht_8 T_4 O_{16}$ corresponding to the spinel structure (3 LGCs). Yellow balls correspond to oxygen ions. (b) $Ht_8 T_4 O_{16}$ with the $Cmmm$ symmetry (7 LGCs). (c) $Ht_8 T_4 O_{16}$ with the $Imma$ symmetry (6 LGCs). The structure is similar to the crystal of $Li_2 WO_4$ (6 LGCs). The difference comes from the sizes of the unit cells. (d) $Ht_4 T_4 O_{12}$ with the $I\bar{4}m2$ symmetry (4 LGCs). (e) $Ht_4 T_4 O_{12}$ corresponding to the $Ga_2 O_3$ structure (5 LGCs). (f) $Ht_5 T_3 O_{12}$ with the $R3$ symmetry (7 LGCs). The illustrated structures are symmetrized and their space groups are determined by the code *SPGLIB* [38] with correcting the structural distortions. All the structural figures in this study are generated by *VESTA* [39].

in the octahderal and tetrahedral sites in the Barrow-packings consisting of oxygens, as the optimal solution of $Ht_{16} T_8 O_{32}$. Figure 3(f) shows the layered structure terminated by tetrahedra and octahedra. This optimal solution is composed of the lowest number of LGCs in $Ht_5 T_3 O_{12}$.

All the optimal solutions of $D_4 Ht_4 O_{12}$ correspond to the real crystals shown in Figs. 4(a), 4(b), 4(c), and 4(d). The cation D makes coordination polyhedra of anticuboctahedron or cuboctahedron. Besides, we find that $D_3 Ht_6 O_{15}$ has the optimal solution corresponding to the $BaTi_2 O_5$ structure shown in Fig. 4(f). The optimal solution consists of 11 LGCs, but we find another optimal solution consiting of only 5 LGCs with the $Cmmm$ symmetry shown in Fig. 4(e).

Our results indicate that $De_4 Ht_2 O_8$ and $De_8 Ht_4 O_{16}$ have only one optimal solution corresponding to the crystal structure of $La_{2-x} Sr_x CuO_4$ as shown in Fig. 5 (a). The cation De makes coordination polyhedron of capped square antiprism. Besides, our calculations indicate that $D_1 De_2 Ht_2 O_7$ and $D_2 De_4 Ht_4 O_{14}$ also have a unique optimal solution corresponding to the crystal structure of $La_{2-x} Sr_{1+x} Cu_2 O_7$ which consists of three kinds of coordination polyhedra as shown in Fig. 5(b).

In the pyrochlore-α structure found as one of the optimal solutions for $Do_4 Hd_4 O_{14}$ and $Uo_4 Hd_4 O_{14}$, one oxygen ion is placed in the center of the tetrahedron composed of large cations, and the distance of ionic bond

TABLE III. The number of the discovered optimal solutions for each composition and the confirmed correspondence of optimal solutions with crystal structures. Note that if the packing fraction is less than the minimum, 60%, which is given in Table II of Supplementary Information [33], we reject the structure from the optimal solutions in this study.

Composition per unit cell	Number of discovered optimal solutions	Confirmed real crystal structures
$Ht_2T_2O_6$	21	Ga_2O_3
$Ht_3T_5O_{12}$	508	-
$Ht_4T_4O_{12}$	597	Ga_2O_3
$Ht_4T_2O_8$	29	Spinel
$Ht_8T_4O_{16}$	216	Spinel
$D_3Ht_6O_{15}$	14	$BaTi_2O_5$
$D_4Ht_4O_{12}$	4	Perovskite, $BaNiO_3$, $BaMnO_3$, Cs_2NaCrF_6
$D_1De_2Ht_2O_7$	1	$La_{2-x}Sr_{1+x}Cu_2O_7$
$D_2De_4Ht_4O_{14}$	1	$La_{2-x}Sr_{1+x}Cu_2O_7$
$De_4Ht_2O_8$	1	$La_{2-x}Sr_xCuO_4$
$De_8Ht_4O_{16}$	1	$La_{2-x}Sr_xCuO_4$
$Do_4Hd_4O_{14}$	6	Pyrochlore-α
$Uo_4Hd_4O_{14}$	20	Pyrochlore-α
$Eo_4Ht_4O_{12}$	21	-
Eo_6O_{12}	944	Zirconia
$Op_4Ht_4O_{12}$	95	$YFeO_3$
$Op_4Ht_8O_{16}$	15	-
$Pe_2Sh_1O_4$	4	$InGaZnO_4$
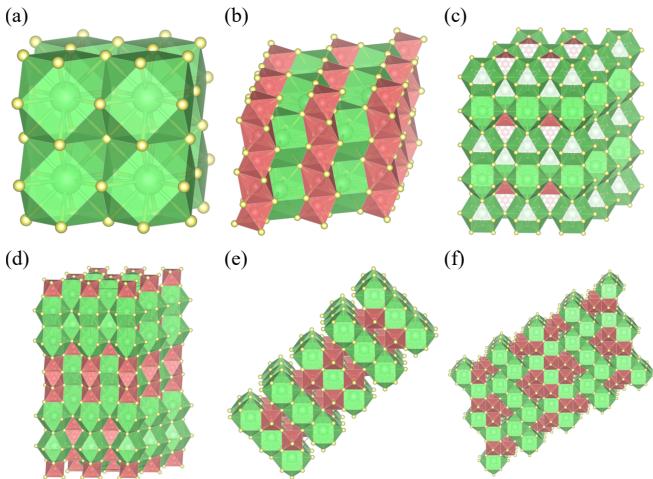
FIG. 4. (a) $D_4Ht_4O_{12}$ corresponding to the perovskite structure (3 LGCs). (b) $D_4Ht_4O_{12}$ corresponding to the crystal of $BaNiO_3$ (3 LGCs). (c) $D_4Ht_4O_{12}$ corresponding to the crystal of $BaMnO_3$ (4 LGCs). (d) $D_4Ht_4O_{12}$ corresponding to the crystal of Cs_2NaCrO_6 (5 LGCs). (e) $D_3Ht_6O_{15}$ with the $Cmmm$ symmetry (5 LGCs). (f) $D_3Ht_6O_{15}$ corresponding to the crystal of $BaTi_2O_5$ (11 LGCs).

FIG. 5. (a) $De_8Ht_4O_{16}$ corresponding to the crystal of $La_{2-x}Sr_xCuO_4$ (4 LGCs). (b) $D_2De_4Ht_4O_{14}$ corresponding to the crystal of $La_{2-x}Sr_{1+x}Cu_2O_7$ (6 LGCs). (c) $Uo_4Hd_4O_{14}$ corresponding to the pyrochlore-α structure (4 LGCs). (d) $Uo_4Hd_4O_{14}$ with the $P1$ symmetry (6 LGCs). (e) Eo_6O_{12} corresponding to the zirconia structure (2 LGCs). (f) $Eo_4Ht_4O_{12}$ with the $Cmcm$ symmetry (4 LGCs).

must be smaller than those of the other ionic bonds, and thereby a large cation makes coordination polyhedron of distorted cubic as shown in Fig. 5(c). In fact, our calculations indicate that the pyrochlore-α structure is difficult to obtain without soft sphere, while both the $Do_4Hd_4O_{14}$ and $Uo_4Hd_4O_{14}$ have the optimal solution corresponding to the pyrochlore-α structure, which has the lowest number of LGCs with highest symmetry. Note that the interatomic distances can be as small as possible with-

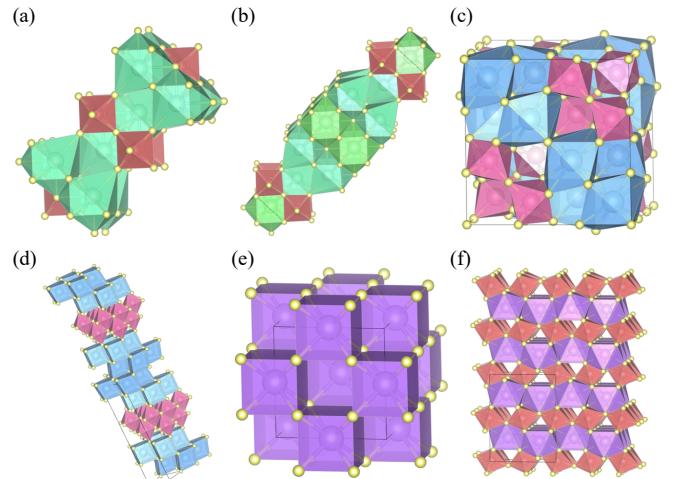out attractive forces since the volume of the unit cell is minimized. The optimal solution consisting of the second lowest number of LGCs for the $Uo_4Hd_4O_{14}$ is the layer-by-layer structure of cubics and octahedra with the $P1$ symmetry as shown in Fig. 5(d).

The cubic coordination can also be realized by the cation Eo. Eo_6O_{12} has the optimal solution corresponding to the zirconia structure shown in Fig. 5(e). The cation Eo can also make the coordination polyhedra of
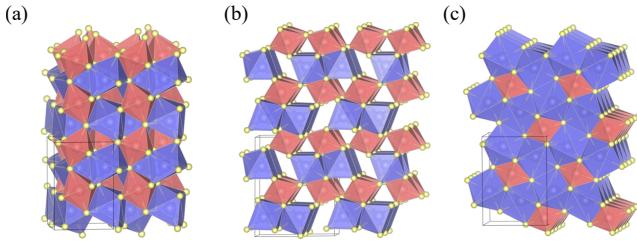
FIG. 6. (a) $Op_4Ht_4O_{12}$ corresponding to the crystal of $YFeO_3$ (4 LGCs). (b) $Op_4Ht_4O_{12}$ with the *Pnma* symmetry (5 LGCs). (c) $Op_4Hd_8O_{16}$ with the *Pbam* symmetry (4 LGCs).

square antiprism as shown in Fig. 5(f).

$Op_4Ht_4O_{12}$ has the optimal solution corresponding to the crystal structure of $YFeO_3$ which is composed of coordination polyhedra of capped trigonal prisms as shown in Fig. 6(a). $YFeO_3$ and the optimal solution shown in Fig. 6(b) have the same symmetry, but the latter structure is composed of 5 LGCs, and capped trigonal prisms in the structure share the faces. The manner of atomic distribution in $YFeO_3$ is the same as $LaMg_xTa_{1-x}O_{1+3x}N_{2-3x}$ which can be employed for overall water splitting at wavelengths of up to 600 nm [40]. The correspondence implies the applicability of our method to mixed anion compounds. On the other hand, we could not find the optimal solution corresponding to the crystal structure of Sr_2PdO_4 in the optimal solutions of $Op_4Hd_8O_{16}$. However, the optimal solution, which has 4 LGCs and the *Pbam* symmetry shown in Fig. 6(c), is similar to the crystal structure of Sr_2PdO_4; they have the same symmetry, and the difference only comes from the kind of linking between octahedra and capped trigonal prisms. In the optimal solution, octahedra and capped trigonal prisms share the faces, while in the crystal structure of Sr_2PdO_4, they share edges. In general, our algorithm tends to generate polyhedra sharing as many common bridging atoms as possible.

Finally, $Pe_2Sh_1O_4$ has the optimal solution corresponding to the crystal structure of $InGaZnO_4$. However, it may be difficult to form trigonal bipyramidal coordination polyhedra using the cation Pe, which possesses a hard cationic radius of C_5. Note that in the hexagonal closest packing of oxygen ions, the center of trigonal bipyramid is identical with the interstice between three atoms in the hexagonal layer, and accordingly, the axial atoms of the bipyramid are 41% more distant than the equatorial atoms from the central atoms. Also, note that C_5 is same as C_6.

In summary, we propose a novel method to enumerate crystal structure prototypes using the paired math-

ematical programming, subject to a set of constraints on atomic distances, and demonstrate its applicability to find a broad range of crystal structures of oxides. The method consists of two optimization problems. The first is the minimization problem of the volume of the unit cell under the geometrical constraints that are the minimum and maximum distances between every pair of atoms, while the second is the maximization problem of the number of chemical bonds. The constraints of the two problems make every optimal solution satisfy a series of empirical rules systematized in inorganic structural chemistry [24, 25]. The two optimization problems are solved alternately to find a proper structure where every cation satisfies its bonding needs by transpositions of geometrical constraints as *sillium* approach for amorphous silicons [28–31]. We find that the linear relaxations of inequality constraints are effective, and accordingly, the small computational cost enables the exhaustive search for the optimal solutions of large-scale systems. We apply the mathematical programming to cases with 18 compositions, find the optimal solutions for each case successfully, and identify the corresponding real crystal structures. Our result strongly implies that the number of optimal solutions in the mathematical programming seems to be much smaller than the number of local minima in the total energy by *ab-initio* simulations. A small number of optimal solutions, identified through the mathematical programming, can be easily validated using *ab-initio* simulations to assess their stability. We anticipate that these optimal solutions may lead to the discovery of novel materials. The successful reproduction of a broad range of oxide crystal structures may suggest the emergence of a universal rule, potentially resulting from the precise refinement of the Pauling rules as demonstrated in this study. Given the proven success with oxide cases, it is anticipated that this method could be extended to other cases such as chalcogenides, mixed anionic compounds, intermetallic compounds, and borides, provided that effective principles are established.

[1] A. R. Oganov, G. Saleh, and A. G. Kvashnin, eds., *Computational Materials Discovery* (The Royal Society of Chemistry, 2019).

[2] A. R. Oganov and C. W. Glass, The Journal of Chemical

Physics **124**, 244704 (2006).

[3] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, Computer Physics Communications **184**, 1172 (2013).

[4] X. Liu, H. Niu, and A. R. Oganov, npj Computational Materials **7**, 199 (2021).

[5] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Phys. Rev. B **82**, 094116 (2010).

[6] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Computer Physics Communications **183**, 2063 (2012).

[7] H. Wang, Y. Wang, J. Lv, Q. Li, L. Zhang, and Y. Ma, Computational Materials Science **112**, 406 (2016).

[8] F. Peng, Y. Sun, C. J. Pickard, R. J. Needs, Q. Wu, and Y. Ma, Phys. Rev. Lett. **119**, 107001 (2017).

[9] H. Liu, I. I. Naumov, R. Hoffmann, N. W. Ashcroft, and R. J. Hemley, Proceedings of the National Academy of Sciences **114**, 6990 (2017).

[10] Z. M. Geballe, H. Liu, A. K. Mishra, M. Ahart, M. Somayazulu, Y. Meng, M. Baldini, and R. J. Hemley, Angewandte Chemie International Edition **57**, 688 (2018).

[11] 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. Eremets, Nature **569**, 528 (2019).

[12] M. Somayazulu, M. Ahart, A. K. Mishra, Z. M. Geballe, M. Baldini, Y. Meng, V. V. Struzhkin, and R. J. Hemley, Phys. Rev. Lett. **122**, 027001 (2019).

[13] V. Struzhkin, B. Li, C. Ji, X.-J. Chen, V. Prakapenka, E. Greenberg, I. Troyan, A. Gavriliuk, and H.-k. Mao, Matter and Radiation at Extremes **5**, 028201 (2020).

[14] I. A. Troyan, D. V. Semenok, A. G. Kvashnin, A. V. Sadakov, O. A. Sobolevskiy, V. M. Pudalov, A. G. Ivanova, V. B. Prakapenka, E. Greenberg, A. G. Gavriliuk, I. S. Lyubutin, V. V. Struzhkin, A. Bergara, I. Errea, R. Bianco, M. Calandra, F. Mauri, L. Monacelli, R. Akashi, and A. R. Oganov, Advanced Materials **33**, 2006832 (2021).

[15] E. O. Pyzer-Knapp, J. W. Pitera, P. W. J. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith, and A. Curioni, npj Computational Materials **8**, 84 (2022).

[16] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Nature **624**, 80 (2023).

[17] C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, S. Shysheya, J. Crabbé, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, R. Tomioka, and T. Xie, arXiv preprint arXiv:2312.03687 (2024).

[18] C. Chen, D. T. Nguyen, S. J. Lee, N. A. Baker, A. S. Karakoti, L. Lauw, C. Owen, K. T. Mueller, B. A. Bilodeau, V. Murugesan, and M. Troyer, arXiv preprint arXiv:2401.04070 (2024).

[19] C. J. Pickard and R. J. Needs, Phys. Rev. Lett. **97**, 045504 (2006).

[20] C. J. Pickard and R. J. Needs, Journal of Physics: Condensed Matter **23**, 053201 (2011).

[21] 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, Nature **619**, 68 (2023).

[22] T. Yokoyama, K. Ichikawa, and H. Naito, Crystal Growth & Design **24**, 2168 (2024).

[23] L. Pauling, Journal of the American Chemical Society **51**, 1010 (1929).

[24] U. Müller, *Inorganic Structural Chemistry, Second Edition* (John Wiley & Sons, Ltd, 2007).

[25] A. F. Wells, *Structural Inorganic Chemistry, Fifth Edition* (Oxford University Press, 1984).

[26] J. George, D. Waroquiers, D. Di Stefano, G. Petretto, G.-M. Rignanese, and G. Hautier, Angewandte Chemie International Edition **59**, 7569 (2020).

[27] T. C. Hales, Annals of Mathematics **162**, 1065 (2005).

[28] W. Zachariasen, J. Am. Chem. Soc. **54**, 3841 (1932).

[29] F. Wooten, K. Winer, and D. Weaire, Phys. Rev. Lett. **54**, 1392 (1985).

[30] F. Wooten and D. Weaire, Solid State Phys. **40**, 1 (1987).

[31] G. Barkema and N. Mousseau, Phys. Rev. B **62**, 4985 (2000).

[32] P. Keating, Phys. Rev. **145**, 637 (1966).

[33] See supplementary information for geometrical constraints, cationic radii, the way to generate initial structures, and numerical aspects such as the parameters for geometrical optimizations.

[34] R. Koshoji, M. Kawamura, M. Fukuda, and T. Ozaki, Phys. Rev. E **103**, 023307 (2021).

[35] R. Koshoji and T. Ozaki, Phys. Rev. E **104**, 024101 (2021).

[36] R. Koshoji and T. Ozaki, Journal of Physics Communications **6**, 075002 (2022).

[37] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).

[38] A. Togo and I. Tanaka, arXiv preprint arXiv:1808.01590 (2018).

[39] K. Momma and F. Izumi, Journal of Applied Crystallography **44**, 1272 (2011).

[40] C. Pan, T. Takata, M. Nakabayashi, T. Matsumoto, N. Shibata, Y. Ikuhara, and K. Domen, Angewandte Chemie International Edition **54**, 2955 (2015).

Supplementary Information for Mathematical Crystal Chemistry

I. METHODS

Our mathematical programming formulation is based on inorganic structural chemistry that describes crystal sructures. As references for our basic idea, see the textbooks [U. Müller, *Inorganic Structural Chemistry, Second Edition* (John Wiley & Sons, Ltd, 2007)] and/or [A. F. Wells, *Structural Inorganic Chemistry, Fifth Edition* (Oxford University Press, 1984)].

A. Geometrical constraints

Figure S1(a) shows the optimal solution of Hd_4O_8 correponding to the rutile structure. Every oxygen makes three chemical bonds with Hd, while every Hd makes six chemical bonds with oxygens. Parallel linear strands of edge-sharing octahedra are joined by common octahedron vertices. Oxygen ions constitute the hexagonal closest-packings, and the structure is reproduced by the anionic constraints between every pair of oxygen ions. Figure S1(b) shows the optimal solution of Hd_4O_8 correponding to the α-PbO_2 structure. The zigzag chains of edge-sharing octahedra are also joined by common vertices. If a pair of oxygen and Hd is not connected by a chemical bond, a non-bonding constraint connects them. Every Hd makes ten polyhedral constraints, and two of them correspond to edge-sharing. These two optimal solutions, which correspond to the rutile structure and the α-PbO_2 structure, respectively, have the same structural fingerprint, because the two structures have the same number of LGCs of the same type, where a fingerprint of a LGC enumerates the element symbols of the adjacent atoms linked by chemical bonds, polyhedral constraints, and anionic constraints in dictionary order, respectively. Figures S1(c) shows the optimal solution Ht_4O_6 correponding to the corundum structure. There are pairs of face-sharing octahedra, and every octahedron shares three edges within a layer and three vertices with octahedra from the adjacent layer to which it has no face-sharing connection.

To share edges or faces of coordination polyhedra, the mimimum distance between two cations is shortened compared to that of the cationic constraints. Suppose σ_i is the minimum bonding distance given by

$$\sigma_i = r_i^{(I)} + C_O, \qquad (S1)$$

and then the minimum distance $D^{(\text{shared})}$ from the cation

TABLE S-I. List of cationic radii. C_n ($n \in \mathbb{N}$) corresponds to the minimum radius so that the cation can connect n oxygen ions, and C_O correponds to the ionic radius of oxygen ion.

Symbol	Cationic radius
C_4	0.314643
C_5	0.579899
C_6	0.579899
C_7	0.827755
C_8	0.903460
C_9	1.02487
C_{10}	1.165450
C_O	1.4

i to the shared common edge or face is given by

$$D_i^{(\text{shared})} = \begin{cases} \sqrt{\sigma_i^2 - C_O^2} & N_{ij}^{(\text{CBA})} = 2 \\ \sqrt{\sigma_i^2 - \dfrac{4}{3}C_O^2} & N_{ij}^{(\text{CBA})} = 3 \\ \sqrt{\sigma_i^2 - 2C_O^2} & N_{ij}^{(\text{CBA})} = 4 \end{cases} . \qquad (S2)$$

B. Initial structures

Initial structures are generated as follows: Suppose $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ are the primitive lattice vectors. First, they are given by

$$(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3) = \begin{pmatrix} l_1 & l_2 \cos\theta_2 & l_3 \sin\varphi_3 \cos\theta_3 \\ 0 & l_2 \sin\theta_2 & l_3 \sin\varphi_3 \sin\theta_3 \\ 0 & 0 & l_3 \cos\varphi_3 \end{pmatrix}, \qquad (S3)$$

where $1 \le l_i \le 2$, $\dfrac{\pi}{3} \le \theta_i \le \dfrac{2\pi}{3}$, and $-\dfrac{\pi}{6} \le \varphi_i \le \dfrac{\pi}{6}$ are randomly set. Second, the lattice is expanded until the sum of the volume of the atomic spheres becomes 70% of the volume of the unit cell, where the anionic and cationic radii are set to be 0.6 and 1.4, respectively. The cartesian coordinate of an atom is set to be

$$\boldsymbol{x} = q_1 \boldsymbol{a}_1 + q_2 \boldsymbol{a}_2 + q_3 \boldsymbol{a}_3 \qquad (S4)$$

where $0 \le q_i \le 1$ are random values.

C. Cationic Radii

The 11 cations, which are listed in Table II of the main article, are selected by considering closed packed polyhedra consisting of the centered cation and surrounding oxygen ions. Cationic radius C_n ($n \in \mathbb{N}$) corresponds to the minimum radius so that the cation can connect n oxygen ions, while C_O correponds to the ionic radius of oxygen ion. Table S-I lists the value of C_n and C_O.
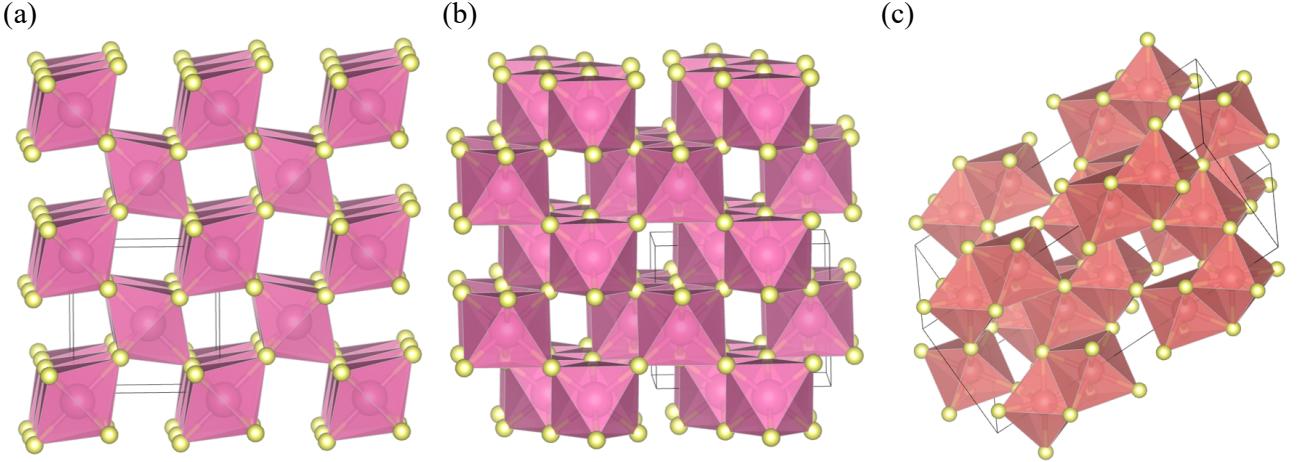
FIG. S1. (a) The optimal solution of Hd_4O_8 corresponding to the rutile structure. The space group is $P4_2/mnm$. (b) The optimal solution of Hd_4O_8 corresponding to the α-PbO_2 structure. The space group is $Pbcn$. (c) The optimal solution of Ht_4O_6 corresponding to the corundum structure. The space group is $R\bar{3}c$.

TABLE S-II. Common parameters for global and local geometrical optimizations.

Parameter	Value
Pressure	1.0
Attractive force constant $k_\uparrow$	30.0
Repulsive force constant $k_\downarrow$	-100.0
Maximum error of geometrical constraints	5%
Minimum packing fraction	60%
Maximum number of total optimization steps	40000

TABLE S-III. Specific parameters for global geometrical optimization.

Parameter	Value
Initial $\Delta x_{\max}$	6.0
Final $\Delta x_{\max}$	0.55
Maximum number of optimization steps	40000
Update frequency of geometrical constraints	20

All cations have the same cationic repulsion radius $r^{(C)} = 1.4$ which is the same as the ionic radius of oxygen ion to avoid the condensation of cations.

II. NUMERICAL ASPECTS

Table S-II lists the common parameters for the global and local geometrical optimizations. The geometrical constraints are given by the minimum and maximum distances between atoms as

$$x_{ij\boldsymbol{T}} \geq d_{ij\boldsymbol{T}}^{(\min)}, \qquad x_{ij\boldsymbol{T}} \leq d_{ij\boldsymbol{T}}^{(\max)}, \qquad \text{(S5)}$$

but when we confirm the feasibility of the structure after local geometrical optimization, we permit 5% error as

$$x_{ij\boldsymbol{T}} \geq 0.95\, d_{ij\boldsymbol{T}}^{(\min)}, \qquad x_{ij\boldsymbol{T}} \leq 1.05\, d_{ij\boldsymbol{T}}^{(\max)}. \qquad \text{(S6)}$$

TABLE S-IV. Specific parameters for local geometrical optimization. The maximum displacement is gradually decreased to find optimal solutions.

Parameter	Value
Initial $\Delta x_{\max}$	0.1
Final $\Delta x_{\max}$	0.02
Maximum number of optimization steps	1000

Besides, in this study, if the packing fraction is less than the minimum, we reject the sparse structure from the optimum solutions. α is calculated as

$$\alpha = \left(\frac{\Delta x_{\max,f}}{\Delta x_{\max,i}}\right)^{\frac{1}{N_{\text{opt}}}}, \qquad \text{(S7)}$$

where $\Delta x_{\max,i}$ and $\Delta x_{\max,f}$ are the initial and final $\Delta x_{\max}$, respectively, and N_{opt} is the maximum number of optimization steps.

Table S-III lists the specific parameters for the global geometrical optimizations (GGO). The network of the chemical bonds is updated per 20 steps of GGO. We monitor the crystal lattice per 1000 steps of GGO, and we reject the structure if the lattice is too distorted or too sparse. To reduce the computational costs, the neighboring list of pairs of atoms is updated per 60 steps of GGO.

Table S-IV lists the specific parameters for the local geometrical optimizations (LGO). The geometrical constraints are fixed through LGO. To relax the structure, the maximum displacements of atoms and crystal lattice are gradually decreased.

We generate 1000000 samples for $Ht_4T_2O_8$ and $Ht_8T_4O_{16}$. Table S-V shows the number of $Ht_4T_2O_8$ and $Ht_8T_4O_{16}$ samples optimized to the spinel structure; about 10% of $Ht_4T_2O_8$ are optimized to spinel structure, while 0.01% of $Ht_8T_4O_{16}$ are optimized to spinel struc-
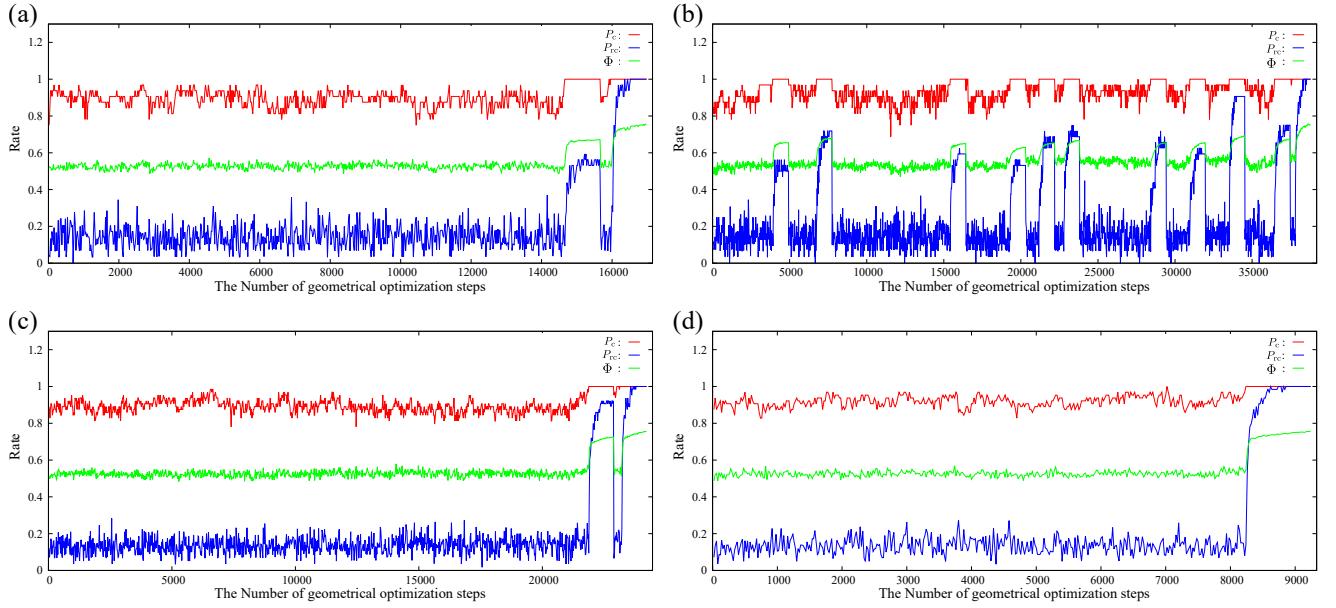
FIG. S2. The examples of the changes of the rate of chemical bonds and relaxed chemical bonds, and the packing fraction through the optimization processes. (a) The exaples of $Ht_4T_2O_8$ that is optimized to the spinel structure. In the first local optimization process, P_{rc} cannot reach 1.0. (b) The exaples of $Ht_4T_2O_8$ that is optimized to the structure with the $Cmmm$ symmetry. This structure tends to require many local optimization processes. (c) The exaples of $Ht_8T_4O_{16}$ that is optimized to the spinel structure. In the first local optimization process, P_{rc} cannot reach 1.0. (d) The exaples of $Ht_8T_4O_{16}$ that is optimized to the structure with the $Imma$ symmetry. This sample reaches the optimal solution in the first local optimization process.

TABLE S-V. The number of samples reaching optimal solutions. We generate 1000000 samples for $Ht_4T_2O_8$ and $Ht_8T_4O_{16}$. Note that 746250 samples of $Ht_4T_2O_8$ and 938542 samples of $Ht_8T_4O_{16}$ cannot reach optimal solutions, respectively.

Optimal Solution	Number of samples
$Ht_4T_2O_8$ (spinel structure)	126048
$Ht_4T_2O_8$ ($Cmmm$ symmetry)	6061
$Ht_8T_4O_{16}$ (spinel structure)	906
$Ht_8T_4O_{16}$ ($Imma$ symmetry)	800

ture. 938542 samples of $Ht_8T_4O_{16}$ cannot reach optimal solutions if we use the optimization parameters shown in Tables S-II, S-III, and S-IV.

Generally, not all the cations have the maximum number of chemical bonds. The global geometrical optimization is aimed at transforming the structure largely enough to create a different network of chemical bonds so that every cation satisfies its bonding needs. To analyze the optimization history, we monitor the rate of chemical bonds P_c given by

$$P_c \equiv \frac{\sum_i s_i^{(+)} n_i^{(CB)}}{\sum_i s_i^{(+)} N_i^{(CB)}}, \tag{S8}$$

where the numerator is the number of chemical bonds,

and the denominator is the total number of bonding needs. If every cation satisfies its bonding needs, the structure is locally optimized to identify whether the structure can be the optimal solusion of the geometrical optimization problem. If so, all the interatomic distances satisfy the condition of Eq. (S6), and we regard the chemical bonds satisfying Eq. (S6) as the relaxed chemical bonds. We monitor the rate of relaxed chemical bonds P_{rc}, where the numerator and the denominator are the total number of the relaxed chemical bonds and the total number of the chemical bonds of the constraint, respectively. We also monitor the packing fraction Φ defined to be

$$\Phi = \frac{1}{\Omega} \sum_i \frac{4\pi}{3} \left(r_i^{(I)} \right)^3. \tag{S9}$$

Figure S2 shows the changes of P_c, P_{rc} and Φ during the optimization process. The fluctuation of P_c indicates that a structure is transformed largely enough to change the network of chemical bonds. The small rate of P_{rc} comes from the large displacement through the global optimization. If P_c reaches 1.0, the structure is locally optimized to identify whether the structure can be the optimal solution of the geometrical optimization problem within the 5% error. In local optimization, P_{rc} shows larger value due to the small displacement, however, in some cases, P_{rc} cannot reach 1.0 due to the contradictions in the geometrical constraints. Φ has the value around 0.5 through the global optimization process.