



Predicting the volumes of crystals

Iek-Heng Chu^{a,1}, Sayan Roychowdhury^{b,1}, Daehui Han^a, Anubhav Jain^{b,*}, Shyue Ping Ong^{a,*}

^aDepartment of NanoEngineering, University of California San Diego, 9500 Gilman Drive # 0448, La Jolla, CA 92093, USA

^bEnergy Technologies Area, Lawrence Berkeley National Lab, 1 Cyclotron Rd, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:

Received 4 December 2017

Received in revised form 17 January 2018

Accepted 17 January 2018

Keywords:

Crystal volume

Lattice prediction

Data mining

Structural optimization

Ionic substitution

ABSTRACT

New crystal structures are frequently derived by performing ionic substitutions on known crystal structures. These derived structures are then used in further experimental analysis, or as the initial guess for structural optimization in electronic structure calculations, both of which usually require a reasonable guess of the lattice parameters. In this work, we propose two lattice prediction schemes to improve the initial guess of a candidate crystal structure. The first scheme relies on a one-to-one mapping of species in the candidate crystal structure to a known crystal structure, while the second scheme relies on data-mined minimum atom pair distances to predict the crystal volume of the candidate crystal structure and does not require a reference structure. We demonstrate that the two schemes can effectively predict the volumes within mean absolute errors (MAE) as low as 3.8% and 8.2%. We also discuss the various factors that may impact the performance of the schemes. Implementations for both schemes are available in the open-source pymatgen software.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

To generate new materials that may potentially possess superior properties, a common strategy both experimentally and computationally is to perform partial or complete substitution of various species in a known crystal. The selection of substituents can be made either based on chemical intuition or by using quantitative data-mined substitution probabilities [1]. The derived candidates are then used for further experimental analysis (e.g., in the refinement of X-ray diffraction patterns), or as an initial guess to electronic structure calculations to determine its phase stability [2] and other application-specific properties [3–7], for example, for energy storage [4,8–11], solid-state lighting [12], thermoelectrics [13,14], catalysis [15], etc. [16,17]. In these analyses, a reasonable guess of the initial lattice parameters is necessary. For instance, the first step in the computational evaluation of any new candidate crystal involves the optimization of the lattice parameters and atomic positions to obtain the equilibrium geometry, and the closer the initially supplied lattice parameters and atomic positions are to the final equilibrium structure, the more likely the structure will converge at a reasonable speed.

For ionic-substitution-derived candidates, one often sets the initial lattice parameters and atomic positions to be identical to

those of the parent structure. In cases where there are substantial size differences between the substituent and original atoms (e.g., for anion substitutions), this suboptimal guess can lead to large errors in structure refinement, as well as slow, or even failures in, convergence. As another use case of lattice scaling, many data mining descriptors, e.g., density, packing fraction, require knowledge of the cell volume. If one is canvassing new chemical compounds with data mining and requires knowledge of a descriptor that is cell-volume dependent, schemes that can provide accurate estimates of the cell parameters are highly desirable.

In this work, we propose two prediction schemes to provide improved estimates of the lattice lengths (and hence, volume) of a candidate crystal structure. The first scheme, which relies on a one-to-one mapping of species in the candidate crystal structure to a known crystal structure, is able to achieve very low mean absolute errors (MAEs) of 3.8% in the volume, whereas the second scheme, which relies on data-mined minimum atom pair distances, can achieve a MAE of 8.2%. We will also discuss the various factors that may impact the performance of the schemes.

2. Lattice length scaling schemes

2.1. Reference lattice scaling scheme

In the first scheme, we focus on new materials that are derived from ionic substitutions of a known crystal, i.e., the atomic positions and lattice parameters of the parent structure are known

* Corresponding authors.

E-mail addresses: ajain@lbl.gov (A. Jain), ongsp@eng.ucsd.edu (S.P. Ong).

¹ These authors contributed equally to this work.

from either experiments or computations. We will henceforth refer to this scheme as the “reference lattice scaling (RLS) scheme”.

Assuming that there are no large changes in lattice angles and atomic positions, our hypothesis is that the lengths of the lattice vectors $\{a_i\}$ are proportional to the sum of the atomic density-weighted atomic radii of the species in the crystal structure, as follows,

$$a_i \propto \sum_k r_k \cdot (N_k)^{1/3},$$

where N_k and r_k are the number of atoms of specie k in the cell and the atomic radius of specie k , respectively, and the factor of $1/3$ converts the volume density to a length density. Here, the atomic radii refer to one of the commonly used definitions of ionic, covalent or Van der Waals radii. We will discuss the selection of radii in a later section. We have observed a similar relationship for the case of bournonite (CuPbSbS_3) family in our recent work, where the computed cell volume for over 300 substitutions was approximately proportional to the sum of atomic volumes determined by the composition [18].

The relationship between the lattice lengths of a derived structure $\{a_i^d\}$ can then be related to the parent structure $\{a_i^p\}$ as follows:

$$\frac{a_i^d}{a_i^p} = \frac{\sum_{k=1}^N r_k^d \cdot (N_k^d)^{1/3}}{\sum_{k=1}^N r_k^p \cdot (N_k^p)^{1/3}} = \alpha_r \quad (1)$$

where the superscripts d and p are used to label parameters for the derived or parent structures, respectively. Similarly, one can demonstrate that the ratio between the volume of the derived structure V_d and the parent structure V_p is given as follows:

$$\frac{V_d}{V_p} = \left[\frac{\sum_{k=1}^N r_k^d \cdot (N_k^d)^{1/3}}{\sum_{k=1}^N r_k^p \cdot (N_k^p)^{1/3}} \right]^3 = \alpha_r^3 \quad (2)$$

Fig. 1(a) illustrates the schematic application of RLS to a derived structure as an example, in which the initial cell parameters $\{a_i^d\}$ are scaled by the factor α_r defined in Eq. (1).

2.2. Data-mined lattice scaling scheme

Unlike RLS, the second scaling scheme for a new material does not require computational or experimental knowledge of a reference crystal. As the predicted crystal parameters are determined based on a data-mining approach, we refer to this second scheme as the “data-mined lattice scaling (DLS) scheme”

For a given crystal structure X , we scale its lattice parameters by a factor determined based on the data-mined predicted atom pair distance between two atoms in X versus their initial distance. Here, any atoms within 4 Å are considered as potential atom pairs.

In the data-mined predictor, the distance associated with two species i and j , d_{ij} , is parameterized as

$$d_{ij} = r_i + r_j + \sigma_X k_i + \sigma_X k_j \quad (3)$$

where σ_X is the standard deviation of Pauling electronegativity of all the species in structure X , called the “electronegativity spread”. The electronegativity spread is intended to be a measure of “structural ionicity”: σ_X equals zero for any pure element while σ_X is large for highly ionic compounds (e.g., $\sigma_X = 1.5$ for LiF). The parameters r_i and k_i are specie dependent, and are derived from fitting $\{(r_i, k_i)\}$ via linear regression on a large training set of observed atom pair distances. In this work, we acquired a large training set of 23,721 thermodynamically-stable (i.e., energy above hull (E_{hull}) [2,19,20] = 0 meV/atom) crystal structures from the Materials Project (MP) database and ran an iterative fitting procedure to determine the $\{(r_i, k_i)\}$ parameters. The fitting procedure and the performance of the DLS on the training set are given in [Supplementary Information \(SI\)](#). We expect the fitted r_i to be approximately equal to the atomic radius (r_c) because r_i represents the contribution of an atom to the atom pair distances in the absence of any electronegativity spread ($\sigma_X = 0$), i.e., in a pure element. The fitted k_i is an adjustment factor based on σ_X in a material that allows the atomic radius to change in more electronegative compounds to provide a continuous measure of ionic radius and we expect that k_i becomes negative for cations and positive for anions. The associated fitted values are tabulated in [Table S1](#).

After the set of parameter pairs $\{(r_i, k_i)\}$ are trained, the predicted lattice parameters for any input crystal structure with initial lattice parameters $\{a_i\}$ and atom pair distances $\{d_{ij}\}$ can be estimated as $a_i^{\text{DLS}} = \alpha_d \cdot a_i$, where the lattice scaling factor α_d is computed from the “most constrained atom distance” as follows,

$$\alpha_d = \max \left\{ \frac{d_{ij}^{\text{DLS}}}{d_{ij}} \right\} \quad (4)$$

where d_{ij}^{DLS} is the predicted minimum distance two atoms computed using parameters $\{(r_i, k_i)\}$. Thus, the algorithm simultaneously enforces two conditions: (i) no two atoms are closer than their minimum predicted distance d_{ij}^{DLS} , preventing “too small” volumes, and (ii) at least one pair of atoms are at precisely their minimum predicted distance, preventing “too large” volumes. Fig. 1(b) depicts the schematics of DLS.

3. Selection of test set

To evaluate the performance of our proposed schemes, we selected a test set of 3112 structure pairs (S_p, S_d) from 309 structural prototypes in the 2016 version of Inorganic Crystal Structure Database (ICSD) [21]. S_p and S_d refer to the parent and derived

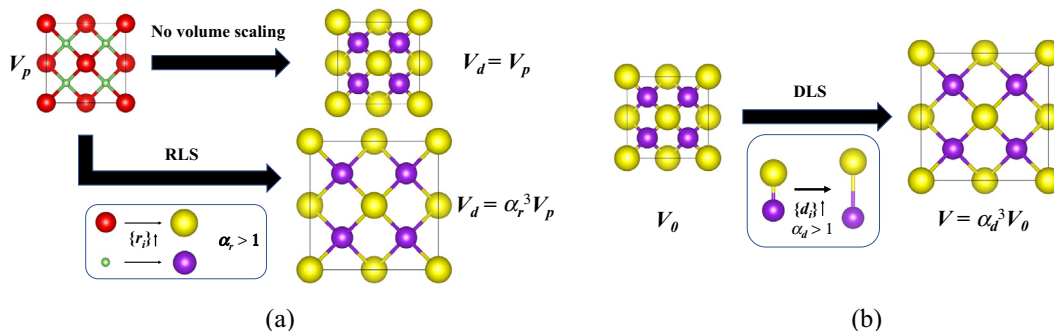


Fig. 1. Schematics of the lattice scaling from (a) reference lattice scaling (RLS) scheme, and (b) data-mined lattice scaling (DLS) scheme. The scaling factors α_r and α_d are defined in Eqs. (1) and (4), respectively.

Download English Version:

<https://daneshyari.com/en/article/7958016>

Download Persian Version:

<https://daneshyari.com/article/7958016>

[Daneshyari.com](https://daneshyari.com)