An intrinsic representation of atomic structure: From clusters to periodic systems

Xiao-Tian Li,[1] Shao-Gang Xu,[1] Xiao-Bao Yang,[1,2] and Yu-Jun Zhao[1,2,a)]
[1]*Department of Physics and School of Materials Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China*
[2]*Key Laboratory of Advanced Energy Storage Materials of Guangdong Province, South China University of Technology, Guangzhou, Guangdong 510640, China*

(Received 23 July 2017; accepted 27 September 2017; published online 10 October 2017)

We have improved our distance matrix and eigen-subspace projection function (EPF) [X.-T. Li *et al.*, J. Chem. Phys. **146**, 154108 (2017)] to describe the atomic structure for *periodic systems*. Depicting the local structure of an atom, the EPF turns out to be invariant with respect to the choices of the unit cell and coordinate frame, leading to an intrinsic representation of the crystal with a set of EPFs of the nontrivial atoms. The difference of EPFs reveals the difference of atoms in local structure, while the accumulated difference between two sets of EPFs can be taken as the distance between configurations. Exemplified with the cases of carbon allotropes and boron sheets, our EPF approach shows exceptional rationality and efficiency to distinguish the atomic structures, which is crucial in structure recognition, comparison, and analysis. *Published by AIP Publishing.* https://doi.org/10.1063/1.4997292

I. INTRODUCTION

The importance of atomic structure is incontrovertible in the kaleidoscope of materials, leading to the crucial role of structure recognition, comparison and analysis, characterized in the age of Big Data. Actually, fast structure processing is highly desirable in structure prediction,[1–11] high-throughput calculations,[12–15] and material genome initiative[16–18] since there are often a vast number of configurations to be dealt with. The process will be staggeringly improved had the identical or similar configurations be categorized for unified treatment. Structural similarity is determined by the definition and depends on the structural representation to a great extent. An ideal representation will be (i) independent of the coordinate frame and atomic ordering, (ii) competent to discriminate configurations completely, (iii) robust enough against noise, preferably with a reasonable mechanism to quantify the difference between structures.

The conventional generalized coordinate is not favorable in structure comparison since it relies on the coordinate frame and atomic ordering. A cluster, for instance, can have distinct generalized coordinates in different coordinate frames and atomic orderings, making it hard to define a reasonable distance between structures. Obviously, the Frobenius norm of generalized coordinates is incapable of quantifying the difference between clusters since it varies with the translation, rotation, and atomic reordering of the configurations.[19] Likewise, a similar problem occurs in periodic structures as crystals. Typically, a crystal structure can be specified by its unit cell and the atomic positions within the cell, which obviously have various representations.[20] As a result, it is difficult to define an invariant distance between crystal structures irrespective of their representations. Admittedly, the unit cell of a crystal can be transformed to a unified form by some algorithms, such as the Niggli reduction.[21,22] The reduced unit cell, however, may vary significantly with little lattice distortions,[23] confining its application in structure comparison.

With respect to the generalized coordinate, the structural invariants prevail in configuration characterization due to their uniqueness. Some general structural invariants, such as the symmetry, principal inertia axis, and atomic coordination number, provide us with the outline of a structure. These fragmentary descriptors, however, cannot specify a configuration completely without the description of detailed structures. It turns out that the so-called fingerprints constructed by a series of structural invariants would be a better choice to depict the structures with the consideration of both uniqueness and completeness.

The powder diffraction patterns can serve as the characters of crystals with related mechanisms[24,25] quantifying their structural differences. Unfortunately, it has been demonstrated that similar powder diffraction patterns may belong to quite different crystals,[26] which impairs their ability to discriminate structures. Besides, there are two types of geometrical fingerprints popular in the structure prediction community: the radial distribution function (RDF)[27] and the rotationally invariant combinations of spherical harmonics.[28] The RDF contains the distribution information of the distances between atom pairs, showing an aspect of the local structure. In particular, Oganov and Valle[10,20,29,30] proposed a normalized fingerprint function based on the distance distributions, along with the discussions of several types of distances between crystal structures. The fingerprint function has been applied to the structure prediction software USPEX, which enhances the

a)Author to whom correspondence should be addressed: zhaoyj@scut.edu.cn. Tel.: +86-20-87110426. Fax: +86-20-87112837.

efficiency vastly by structure clustering[20,30] and the antiseed technique.[10] As to the combinations of spherical harmonics, they focus on the orientation of the bonds and sensitive to structural symmetry, which has been utilized to eliminate similar configurations in the structure prediction software CALYPSO.[8,9]

Distance matrix,[31–33] which is more comprehensive than the RDF with definite affiliations of the distances, is also widely used in structure recognition and comparison. The difference of distance matrices, however, cannot be taken as the distance between structures since it varies with the atomic orderings. Typically, the eigenvalue spectrum of the distance matrix can serve as the fingerprint of a structure with a metric[19,34] to quantify their differences. We have presented a representation for molecular structures based on distance matrix recently, which was proved competent and efficient.[35] This representation is more detailed than the previous ones with not only the information of eigenvalue spectrum but also the information of atomic projections on the eigen-subspaces.

In this paper, we generalize our approach to periodic systems. In order to describe the crystal structures, we propose two types of revisionary distance matrices, from which the eigen-subspace projection functions (EPFs) can be derived. The EPF depicts the local atomic structure of an atom, while a set of EPFs of the atoms in a unit cell constitutes an intrinsic representation of the crystal. The EPF distances are then presented accordingly for structure comparison and analysis, showing excellent performance in application to carbon (C) allotropes and boron (B) sheets.

II. MODEL AND METHODOLOGY

As is well known, the distance matrix of a cluster specifies the relative position of each atom by its distances to all the other atoms, containing all the structural information except the overall chirality.[31] Each atom here locates in reference to all the other atoms. An infinite system as crystal, however, cannot be represented by the conventional distance matrix straightforward since there are infinite atoms as references. In essence, it is unnecessary to contain too many referenced atoms to determine the position of an atom. Actually, dozens of referenced atoms are sufficient in general, while the infinite atoms far away can be discarded.

We present two types of revisionary distance matrices to represent the atomic structures of crystals, and the first type is termed to be the decreasing distance matrix. For each atom in a crystal, dozens of referenced atoms are selected according to their distances to the atom, i.e., the surrounding atoms within a specific r_{cut} are contained as references while the others discarded. The cutoff for the referenced atoms settles the issue of infinite references but leads to a discontinuity in the distance matrix. For instance, the referenced atoms can vary with a minor distortion of the crystal (while there are some atoms across the cutoff boundary), leading to a significant change of the distance matrix. In order to diminish the effect of the discontinuity as much as possible, we replace the interatomic distance d in the distance matrix with a decreasing function of d,

$$f(d)=\left(1-\frac{d}{d_{\mathrm{cut}}}\right)^4, \quad d\le d_{\mathrm{cut}},$$
$$=0, \qquad\qquad d>d_{\mathrm{cut}}. \tag{1}$$

This is an ordinary power function, which decreases to zero at d_{cut} with the continuity of the first derivative. In this way, the correlations between atoms decrease with their distances, while there are no direct correlations between atoms apart by d_{cut}. As a result, the atoms at the boundary would have little effect on the central atom when r_{cut} is much larger than d_{cut}. The decreasing distance matrix of atom i then can be defined as

$$\mathbf{D}_i^{\mathrm{dec}}=\begin{pmatrix} z_i & f(d_{ij_1}) & \cdots & f(d_{ij_N}) \\ f(d_{ij_1}) & z_{j_1} & \cdots & f(d_{j_1j_N}) \\ \vdots & \vdots & \ddots & \vdots \\ f(d_{ij_N}) & f(d_{j_1j_N}) & \cdots & z_{j_N} \end{pmatrix}, \tag{2}$$

where $j_1, j_2, \ldots, j_N$ are the referenced atoms around atom i within r_{cut} (we take $r_{\mathrm{cut}}=2d_{\mathrm{cut}}$ in this paper), z_i is the atomic number (or any other character) of atom i, and d_{ij_1} represent the Cartesian distance between atoms i and j_1. Such a matrix contains the information of the atomic structure around atom i, while a set of matrices belonging to the atoms in a unit cell constitute a detailed representation of the crystal structure.

The decreasing distance matrix contains only the short-range geometric correlations between atoms, without the long-range ones. It turns out that atoms apart by d_{cut} have no direct interactions in the decreasing distance matrix. In order to describe the crystal structures better, we introduce another type of matrix, the minimum distance matrix. We define the minimum distance between atoms i and j in a crystal to be

$$d_{ij}^{\mathrm{min}}=\min_{n_1,n_2,n_3}\left\|\mathbf{r}_i-\left(\mathbf{r}_j+n_1\mathbf{a}_1+n_2\mathbf{a}_2+n_3\mathbf{a}_3\right)\right\|, \tag{3}$$

where $\mathbf{r}_i$ and $\mathbf{r}_j$ represent the position vectors of atoms i and j, respectively, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are the translation vectors of the crystal, while n_1, n_2, and n_3 can be arbitrary integers. d_{ij}^{min} appears to be the minimum distance between atom i and all the equivalent atoms of atom j under the translational symmetry. The minimum distance matrix of a crystal then can be defined as

$$\mathbf{D}^{\mathrm{min}}=\begin{pmatrix} z_1 & d_{12}^{\mathrm{min}} & \cdots & d_{1n}^{\mathrm{min}} \\ d_{12}^{\mathrm{min}} & z_2 & \cdots & d_{2n}^{\mathrm{min}} \\ \vdots & \vdots & \ddots & \vdots \\ d_{1n}^{\mathrm{min}} & d_{2n}^{\mathrm{min}} & \cdots & z_n \end{pmatrix}, \tag{4}$$

where $1, 2, \ldots, n$ represent the set of atoms in a unit cell, and z_1, z_2, …, z_n are their atomic numbers (or any other characters), respectively. This matrix describes the geometric correlations between the nontrivial atoms under the translational symmetry, with the information of the lattice.

The decreasing distance matrix and the minimum distance matrix describe the crystal structures complementarily:

the former suits to crystals of small cells while the latter fits those of large cells. Of note, both the decreasing distance matrix and the minimum distance matrix can be applied to non-periodic structures like clusters, which can be regarded as of infinite unit cells. In particular, the minimum distance matrix of a cluster appears to be the extended distance matrix we proposed for clusters.[35] The decreasing distance matrix and the minimum distance matrix are both irrelevant to the unit cells and coordinate frames, beneficial in structure comparison. As is mentioned above, however, they still rely on the atomic ordering, so cannot be compared directly. We have derived a fingerprint function from the distance matrix for the recognition and comparison of structures, which has been discussed in detail previously.[35] The approach is briefly described below.

By the spectral factorization of the distance matrix, we define the eigen-subspace projection array (EPA) of atom i to be

$$\mathbf{s}_i = \left(s_{i,\lambda_1}, s_{i,\lambda_2}, \ldots, s_{i,\lambda_s}\right), \tag{5}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_s$ are all the distinct eigenvalues of the distance matrix in ascending order, and s_{i,λ_k} is the norm of orthogonal projection of atom i on the eigen-subspace associated with λ_k. Here the atoms can be viewed as the unit vectors in the eigen-space, with their segments projected to the eigen-subspaces since $\sum_{k=1}^{s} (s_{i,\lambda_k})^2 = 1$. The EPAs turn out to be geometric invariants of the atoms, specifying their positions in the eigen-space with respect to the eigen-subspace framework. Moreover, a visual eigen-subspace projection function (EPF) can be derived from the EPA, describing the relationship between the proportions of the atomic vector ($S \in [0, 1]$) and the eigenvalues that the projected eigen-subspaces are associated with. Based on the function, an EPF distance between atoms i and j can be defined as

$$d_{i,j}^{\mathrm{EPF}} = \int_0^1 \left|\Lambda_i - \Lambda_j\right| dS, \tag{6}$$

where Λ_i and Λ_j represent the EPFs of atoms i and j, respectively. Accordingly, an EPF distance between configurations p and q (both with n atoms) can be defined to be

$$d_{p,q}^{\mathrm{EPF}} = \frac{1}{n} \min_{\{i,j\}} \sum_{i=j=1}^{n} d_{i,j}^{\mathrm{EPF}}, \tag{7}$$

where atoms i and j belong to configurations p and q, respectively, and the summation represents taking the accumulated difference of all the atoms. Here the atoms of the two configurations should be matched in such a way that makes the accumulated distance minimal, which can be approached by the Hungarian algorithm.[36]

It is guaranteed that the defined EPF distance between configurations satisfies the triangle inequality of metric space, as we prove mathematically in the supplementary material, where we also discuss the homometric structure problem. The homometric structures, of identical atoms with the same set of interatomic distances, can be distinguished naturally by our EPF fingerprints as the distance matrix contains the affiliations of the distances. Meanwhile, we point out here that although Eq. (7) only takes account of configurations with the same atom numbers (in unit cells for crystals), it can be easily applied to the ones with different atom numbers. For instance, two configurations with n_1 and n_2 atoms, respectively, can be expanded to n atoms (n is the least common multiple of n_1 and n_2) by simple repetitions, and then their EPF distance can be achieved from the expanded structures.

In general, we can derive the atomic eigen-coordinates and EPAs by the spectral decomposition of the distance matrix following our methodology. The eigen-coordinates depend on the eigenvector basis and cannot serve as structural fingerprints, although they contain all the structural information (except the overall chirality). By contrast, the EPAs only rely on the relative positions of the atoms and turn out to be intrinsic quantities. The fingerprint functions (EPFs) then are proposed to specify the structures. We acknowledge that we cannot strictly prove our fingerprints characterizing a crystal structure uniquely so far. Nevertheless, the distance matrices contain much redundant information to determine a configuration (the degrees of freedom of distance matrices are much larger than that of a configuration), while the EPFs inherit a major part of that. In fact, we have never found two distinct structures with identical EPFs in practice and believe that structures are uniquely characterized by our fingerprints.

III. APPLICATION TO CARBON ALLOTROPES

Carbon, one of the most abundant elements on earth and the central element of life, has a variety of allotropes with staggeringly different structures and properties. For instance, Fig. 1(a) displays six common configurations of C: graphene, graphite, diamond, the zigzag carbon nanotube (CNT), the armchair CNT, and C_{60}. Here the single-walled (7, 0) zigzag CNT and (4, 4) armchair CNT are selected, with similar diameters. Seven non-equivalent atoms of all the configurations are marked in Fig. 1(a). Of note, the graphite has two non-equivalent atoms under all the symmetry operations, while each of the other configurations has only one.

We can analyze and discuss the local atomic structures of these non-equivalent atoms by our EPF approach (equivalent atoms would have identical EPFs, without the need for reduplicative analysis). We take $r_{\mathrm{cut}} = 2d_{\mathrm{cut}} = 10$ Å to construct the decreasing distance matrices $\mathbf{D}^{\mathrm{dec}}$ for the atoms and assign the diagonal elements with zeros since there is only one type of element (assigning with the characters of C here, such as the atomic numbers, makes no difference but a shift of all the EPFs). The derived Λ^{dec} functions of the non-equivalent atoms are shown in Fig. 1(b). These functions describe the atomic positions in the eigen-space and can serve as the atomic fingerprints. More details are provided in our earlier related work.[35]

Typically, the atoms of graphene, diamond, and C_{60} possess vastly distinct Λ^{dec} functions [see Fig. 1(b1)], in agreement with their obviously different local structures. In contrast, the atoms of graphene and graphite (including two types of atoms) have very similar functions [see Fig. 1(b2)], ascribed to their in-plane honeycomb structures. Of note, the plane-spacing of graphite is 3.35 Å, much larger than the nearest interatomic distance in-plane (1.42 Å). As the non-diagonal elements of $\mathbf{D}^{\mathrm{dec}}$ assigned with the

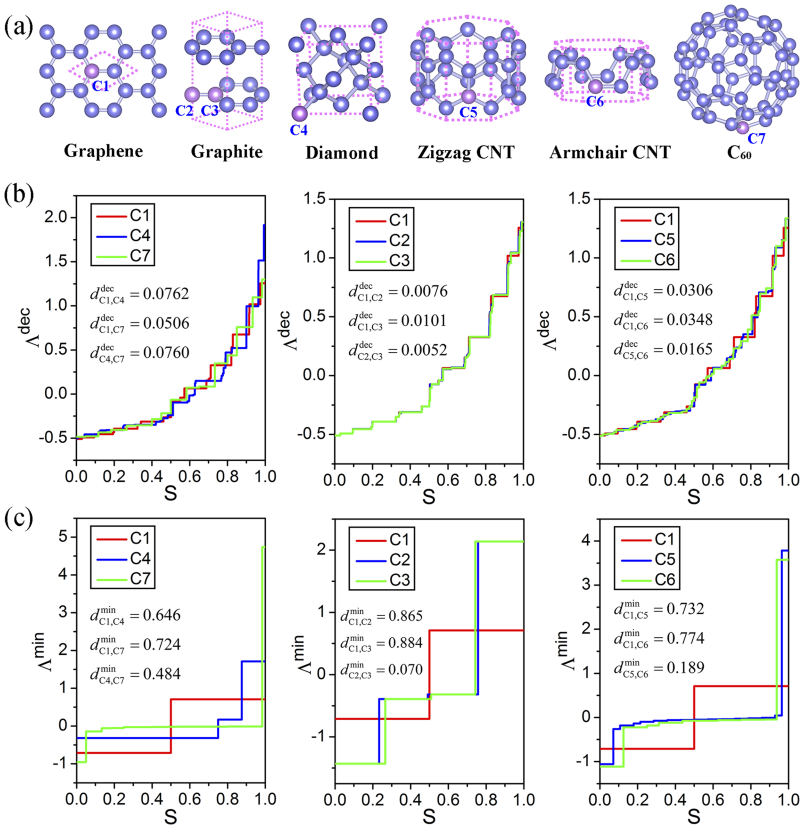

FIG. 1. (a) The atomic structures of six allotropes of C along with (b) the Λ^{dec} functions and (c) the Λ^{min} functions of their non-equivalent atoms. Seven non-equivalent atoms are marked in (a), while the unit cells of the allotropes are denoted by the dotted violet wireframes.

decreasing functions of interatomic distances [see Eqs. (1) and (2)], the in-plane structures would have a greater effect on the Λ^{dec} functions than the out-of-plane ones. In fact, the identical in-plane honeycomb structures of graphene and graphite result in the general outlines of their Λ^{dec} functions, while the distinct out-of-plane structures lead to the tiny differences. As a result, the functions of the two atoms of graphite (with similar out-of-plane structures) are closer with respect to the one of the atom of graphene (with no out-of-plane structure), which is definitely characterized by their EPF distances d^{dec} [see the definition in Eq. (6)]. Likewise, the structural similarities and differences between graphene and the CNTs can also be revealed by their Λ^{dec} functions, as shown in Fig. 1(b3). The CNTs can be seen as rolled graphenes, resulting in their roughly similar functions, while the functions of the CNTs are closer attributed to their cylindrical structures.

The minimum distance matrices $\mathbf{D}^{min}$ of the C allotropes in Fig. 1(a) have also been constructed for structure analysis. It is worthwhile to point out that $\mathbf{D}^{min}$ relies on the cell size of the crystal though irrelevant to the choice of lattice vectors [see Eqs. (3) and (4)]. Here the cells adopted for the structures are denoted by dotted violet wireframes, as shown in Fig. 1(a). The derived Λ^{min} functions of the non-equivalent atoms are displayed in Fig. 1(c). Similarly, the atoms of graphene, diamond, and C_{60} possess significantly distinct Λ^{min} functions due to their extremely different local structures, as seen in Fig. 1(c1). However, the atom of graphene has quite a different Λ^{min} function compared with the two atoms of graphite [see Fig. 1(c2)], different from the case of Λ^{dec} functions. This is attributed to the different constructions of $\mathbf{D}^{min}$. As seen in Eqs. (3) and (4), the non-diagonal elements of $\mathbf{D}^{min}$ are assigned with the minimum interatomic distances under translational symmetry, leading to a greater contribution of the distant atoms to the Λ^{min} functions than the near atoms. As a result, the two atoms of graphite have similar Λ^{min} functions due to their similar out-of-plane structures, distinct from the atom of graphene with no out-of-plane structure. Likewise, the Λ^{min} functions of graphene and the CNTs can be explained in the same way [see Fig. 1(c3)].

Recently, Xie *et al.* studied the pathways of graphite-to-diamond phase transition,[37] which contain a gradual variation of the configurations. We test our EPF representation here with the gc1 pathway composed of the initial state graphite (G), two transition states (TS1 and TS2), an intermediate state (MS), and the final state cubic diamond (CD), as shown in Fig. 2(a). Four typical atoms are selected to map into the EPF space, with their Λ^{dec} functions varying from G to CD [see Fig. 2(b)]. We take $r_{cut} = 2d_{cut} = 5$ Å here to have neat functions for comparison. As seen in the graph, the functions of C1 change slightly from G to TS2 and have a saltation from TS2 to CD, in contrast with the other three atoms. In corresponding to the variations of the atoms, we observe that the functions well reflect the gradual changes of the atomic bonding during the phase transition, with potential applications in practice.

As is shown above, the Λ^{dec} and Λ^{min} functions reflect different parts of the local structures (Λ^{dec} focuses on the near part, while Λ^{min} focuses on the distant part), resulted from the different constructions of $\mathbf{D}^{dec}$ and $\mathbf{D}^{min}$. In fact,

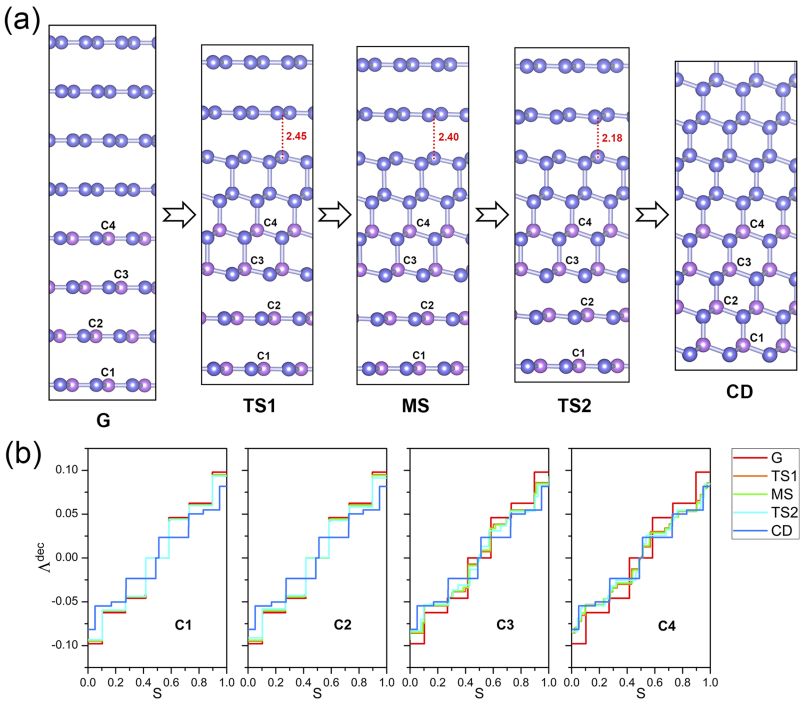

FIG. 2. (a) Snapshots along the gc1 pathway of the graphite-to-diamond phase transition (reproduced according to data from the work of Xie *et al.*[37]) and (b) the gradual variations of the Λ^{dec} functions of four representative atoms.

one can construct a variety of distance matrices to emphasize different aspects of the local structures in practice by regulating the functions in Eqs. (1) and (3). Of note, in the case of C allotropes above, each configuration has only one or two non-equivalent atoms, leading to one or two EPFs concisely. In general cases, however, there may be much more nontrivial atoms, where a set of EPFs constitute a detailed representation of the structure. The representation may be complicated, with an intuitive EPF distance [see Eq. (7)] helpful in practice, which will be illustrated by the case of B sheets in the following.

IV. APPLICATION TO BORON SHEETS

Boron, a close neighbor of carbon, possesses a rich chemistry due to its complex multicenter bonds.[38] It has been demonstrated theoretically that B has a variety of monolayer crystalline structures (buckled and unbuckled),[39–42] and confirmed by experiments later.[43,44] We search the B sheets within 1-16 times of the primitive cell completely by introducing vacancies (separate from each other in consideration of stability) in the unbuckled triangular sheet (the identical configurations have been eliminated by our EPF approach). The B sheets are performed with first-principles calculations, implemented in the Vienna *ab initio* simulation package (VASP)[45,46] with the Perdew-Burke-Ernzerhof (PBE) of Generalized Gradient Approximation (GGA) functional.[47]

We can analyze the structure-energy correlations between the B sheets by our EPF approach. The decreasing EPF distances d^{dec} between the 251 B sheets within 1-16 times of the primitive cell are obtained and projected on a two-dimensional (2D) map [see Fig. 3(a)]. The dots on the 2D map represent the structures, while their colors indicate the energies. It is obvious that there is an interesting correlation between the structures and their energies, indicating that the most stable structures get together while the less stable ones scatter around. Two extreme configurations, the triangular sheet (unbuckled) with no vacancy and the hexagonal sheet with the largest concentration (1/3) of vacancies, locate at the corners of the map, while the other sheets distribute between them logically. We notice that the configurations with high stabilities (see the blue dots on the map) have a wide distribution, revealing the polymorphism of B sheets.[41] Particularly, four typical B sheets with the best stabilities [their detailed structures are shown in Fig. 3(c)] are marked by stars on the map, separate from each other. It turns out that there are a variety of B sheets with comparable stabilities.

The B sheets can be classified according to a local structural parameter, the coordination number (CN) of atoms. For instance, the α-type B sheets contain atoms with CN = 5, 6, while the β-type ones with CN = 4, 5, 6.[42] As is mentioned above, the decreasing EPF distances d^{dec} are sensitive to the neighboring structures due to the decreasing functions assigned in the non-diagonal elements of $\mathbf{D}^{dec}$. As a result, there is a relationship between the classification and d^{dec}. For instance, among the four configurations with the best stabilities mentioned above, two α-type ones are close to each other with respect to the β-type ones, vice versa.

In addition to the local structural parameter CN, there is a global parameter η describing the vacancy concentrations of the B sheets, defined as the ratio of the number of vacancies to the number of original atomic sites within a unit cell.[39,42] There is a plot in Fig. 3(b), showing the decreasing EPF distances d^{dec} of the B sheets to two extreme configurations, the triangular sheet (unbuckled) with the minimum $\eta = 0$ and the hexagonal sheet with the maximum $\eta = \frac{1}{3}$. Note that all the other configurations situate above the straight line between the triangular and hexagonal sheets, ascribed to the triangle inequality of distance. The colors of the dots in the plot stand

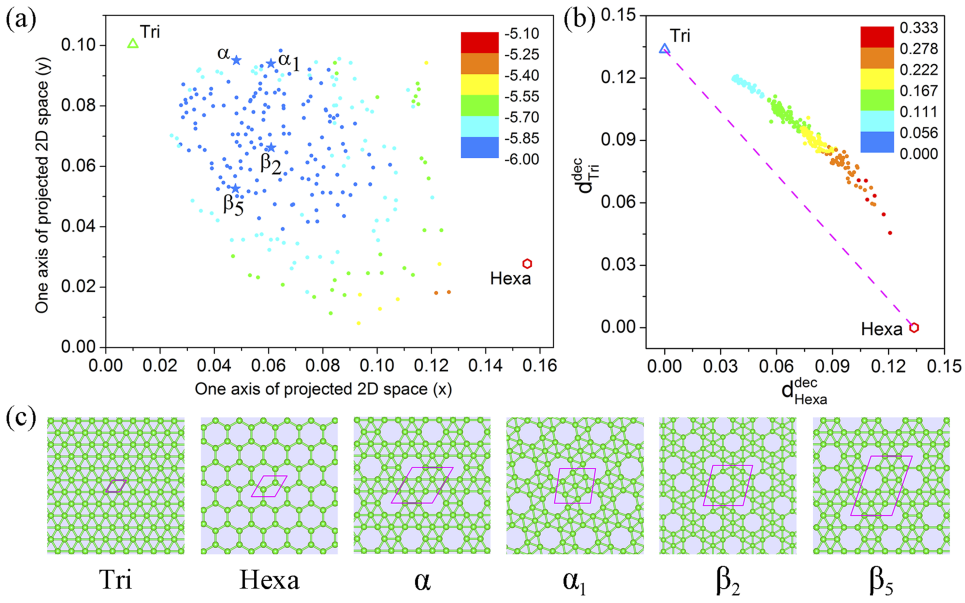

FIG. 3. (a) 2D map of the B sheets. The dots represent the B sheets within 1-16 times of the primitive cell, whose positions on the map are determined by minimizing $\sum_{p<q}\left(d_{pq}^{2D}-d_{pq}^{dec}\right)^2$. The energies of these configurations are indicated by their colors. Two extreme configurations, the triangular sheet and the hexagonal sheet, are marked by triangle and hexagon, respectively, while four typical structures with the best stabilities are marked by stars. (b) A plot showing the decreasing EPF distances of the B sheets to two extreme configurations, the triangular sheet and the hexagonal sheet. The colors of the dots stand for the vacancy concentrations η of the configurations. (c) The detailed structures of six particular configurations in (a).

for η of the structures, varying with d^{dec} from the triangular sheet side to the hexagonal sheet side gradually. It turns out that our decreasing EPF distances are closely related to the vacancy concentrations, revealing the global structural differences between the B sheets.

V. SUMMARY

An EPF representation is proposed to describe the atomic structure of clusters and periodic systems based on the revisionary distance matrices. By the construction of the matrices, the EPF can be regulated to emphasize different parts of the local structure of an atom and the corresponding lattice information, beneficial for structure analysis. An intuitive EPF distance is presented accordingly, providing a visual landscape of the correlations between structures.

SUPPLEMENTARY MATERIAL

See supplementary material for the proof of triangle inequality of the EPF distance between configurations and the discussion of the homometric structure problem.

ACKNOWLEDGMENTS

This work is financially supported by NSFC (Grant Nos. 11574088 and 51431001), the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 51621001), and Natural Science Foundation of Guangdong Province of China (Grant No. 2016A030312011).

[1] R. Martoňák, A. Laio, and M. Parrinello, Phys. Rev. Lett. **90**, 075503 (2003).
[2] S. Goedecker, J. Chem. Phys. **120**, 9911 (2004).
[3] S. T. Call, D. Y. Zubarev, and A. I. Boldyrev, J. Comput. Chem. **28**, 1177 (2007).
[4] M. A. Neumann, F. J. J. Leusen, and J. Kendrick, Angew. Chem., Int. Ed. **47**, 2427 (2008).
[5] M. Amsler and S. Goedecker, J. Chem. Phys. **133**, 224104 (2010).
[6] A. R. Oganov, A. O. Lyakhov, and M. Valle, Acc. Chem. Res. **44**, 227 (2011).
[7] Q. Zhu, A. R. Oganov, and A. O. Lyakhov, CrystEngComm **14**, 3596 (2012).
[8] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Comput. Phys. Commun. **183**, 2063 (2012).
[9] J. Lv, Y. Wang, L. Zhu, and Y. Ma, J. Chem. Phys. **137**, 084104 (2012).
[10] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, Comput. Phys. Commun. **184**, 1172 (2013).
[11] L. B. Vilhelmsen and B. Hammer, J. Chem. Phys. **141**, 044711 (2014).
[12] W. Setyawan and S. Curtarolo, Comput. Mater. Sci. **49**, 299 (2010).
[13] A. Jain, G. Hautier, C. J. Moore, S. Ping Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, Comput. Mater. Sci. **50**, 2295 (2011).
[14] S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, Comput. Mater. Sci. **58**, 218 (2012).
[15] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, Comput. Mater. Sci. **58**, 227 (2012).
[16] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 011002 (2013).
[17] A. Jain, K. A. Persson, and G. Ceder, APL Mater. **4**, 053102 (2016).
[18] M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, T. Chiang, S. Empedocles, J. M. Gregoire, A. G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt, J. V. Duren, and A. Zakutayev, Appl. Phys. Rev. **4**, 011105 (2017).
[19] A. Sadeghi, S. A. Ghasemi, B. Schaefer, S. Mohr, M. A. Lill, and S. Goedecker, J. Chem. Phys. **139**, 184118 (2013).
[20] M. Valle and A. R. Oganov, in *IEEE Symposium on Visual Analytics Science and Technology (VAST '08), 19-24 October 2008* (IEEE, 2008), pp. 11–18.
[21] L. C. Andrews and H. J. Bernstein, J. Appl. Crystallogr. **47**, 346 (2014).
[22] K. J. McGill, M. Asadi, M. T. Karakasheva, L. C. Andrews, and H. J. Bernstein, J. Appl. Crystallogr. **47**, 360 (2014).
[23] L. C. Andrews, H. J. Bernstein, and G. A. Pelletier, Acta Crystallogr., Sect. A **36**, 248 (1980).
[24] H. R. Karfunkel, B. Rohde, F. J. J. Leusen, R. J. Gdanitz, and G. Rihs, J. Comput. Chem. **14**, 1125 (1993).
[25] R. de Gelder, R. Wehrens, and J. A. Hageman, J. Comput. Chem. **22**, 273 (2001).
[26] H. Karfunkel, H. Wilts, Z. Hao, A. Iqbal, J. Mizuguchi, and Z. Wu, Acta Crystallogr., Sect. B: Struct. Sci. **55**, 1075 (1999).
[27] E. L. Willighagen, R. Wehrens, P. Verwer, R. de Gelder, and L. M. C. Buydens, Acta Crystallogr., Sect. B: Struct. Sci. **61**, 29 (2005).
[28] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Phys. Rev. B **28**, 784 (1983).
[29] A. R. Oganov and M. Valle, J. Chem. Phys. **130**, 104504 (2009).

[30]M. Valle and A. R. Oganov, Acta Crystallogr., Sect. A: Found. Crystallogr. **66**, 507 (2010).
[31]T. Havel, I. Kuntz, and G. Crippen, Bull. Math. Biol. **45**, 665 (1983).
[32]Z. Mihalić, D. Veljan, D. Amić, S. Nikolić, D. Plavšić, and N. Trinajstić, J. Math. Chem. **11**, 223 (1992).
[33]A. Kloczkowski, R. L. Jernigan, Z. Wu, G. Song, L. Yang, A. Kolinski, and P. Pokarowski, J. Struct. Funct. Genomics **10**, 67 (2009).
[34]L. Zhu, M. Amsler, T. Fuhrer, B. Schaefer, S. Faraji, S. Rostami, S. A. Ghasemi, A. Sadeghi, M. Grauzinyte, C. Wolverton, and S. Goedecker, J. Chem. Phys. **144**, 034203 (2016).
[35]X.-T. Li, X.-B. Yang, and Y.-J. Zhao, J. Chem. Phys. **146**, 154108 (2017).
[36]H. W. Kuhn, Nav. Res. Logist. Q. **2**, 83 (1955).
[37]Y.-P. Xie, X.-J. Zhang, and Z.-P. Liu, J. Am. Chem. Soc. **139**, 2545 (2017).
[38]T. Ogitsu, E. Schwegler, and G. Galli, Chem. Rev. **113**, 3425 (2013).
[39]H. Tang and S. Ismail-Beigi, Phys. Rev. Lett. **99**, 115501 (2007).
[40]X. Yang, Y. Ding, and J. Ni, Phys. Rev. B **77**, 041402 (2008).
[41]E. S. Penev, S. Bhowmick, A. Sadrzadeh, and B. I. Yakobson, Nano Lett. **12**, 2441 (2012).
[42]X. Wu, J. Dai, Y. Zhao, Z. Zhuo, J. Yang, and X. C. Zeng, ACS Nano **6**, 7443 (2012).
[43]A. J. Mannix, X.-F. Zhou, B. Kiraly, J. D. Wood, D. Alducin, B. D. Myers, X. Liu, B. L. Fisher, U. Santiago, J. R. Guest, M. J. Yacaman, A. Ponce, A. R. Oganov, M. C. Hersam, and N. P. Guisinger, Science **350**, 1513 (2015).
[44]B. Feng, J. Zhang, Q. Zhong, W. Li, S. Li, H. Li, P. Cheng, S. Meng, L. Chen, and K. Wu, Nat. Chem. **8**, 563 (2016).
[45]G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[46]G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[47]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).