Definition of a scoring parameter to identify low-dimensional materials components

Peter Mahler Larsen,[1,2,*] Mohnish Pandey,[1] Mikkel Strange,[1] and Karsten Wedel Jacobsen[1]

[1]*Center for Atomic-scale Materials Design (CAMD), Department of Physics, Technical University of Denmark,*
2800 Kongens Lyngby, Denmark
[2]*Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

The last decade has seen intense research in materials with reduced dimensionality. The low dimensionality leads to interesting electronic behavior due to electronic confinement and reduced screening. The investigations have to a large extent focused on 2D materials both in their bulk form, as individual layers a few atoms thick, and through stacking of 2D layers into heterostructures. The identification of low-dimensional compounds is therefore of key interest. Here, we perform a geometric analysis of material structures, demonstrating a strong clustering of materials depending on their dimensionalities. Based on the geometric analysis, we propose a simple scoring parameter to identify materials of a particular dimension or of mixed dimensionality. The method identifies spatially connected components of the materials and gives a measure of the degree of "1D-ness," "2D-ness," etc., for each component. The scoring parameter is applied to the Inorganic Crystal Structure Database and the Crystallography Open Database, ranking the materials according to their degree of dimensionality. In the case of 2D materials the scoring parameter is seen to clearly separate 2D from non-2D materials and the parameter correlates well with the bonding strength in the layered materials. About 3000 materials are identified as one-dimensional, while more than 9000 are mixed-dimensionality materials containing a molecular (0D) component. The charge states of the components in selected highly ranked materials are investigated using density functional theory and Bader analysis showing that the spatially separated components have either zero charge, corresponding to weak interactions, or integer charge, indicating ionic bonding.

I. INTRODUCTION

Low-dimensional materials with one or more characteristic lengths of the materials limited to the atomic scale have received significant attention recently. Since the discovery of graphene the world has seen intense research in 2D materials involving synthesis and investigation of mechanical, electronic, magnetic, and catalytic properties of new materials [1–4]. Also a number of computational efforts have been dedicated to the identification of new 2D materials and to the construction of computational databases with information about their stability and (photo)electronic properties [5–7]. One of the driving forces behind this research has been an interest in ultrasmall electronic components and this has also led to studies of 1D or quasi-1D materials as possible interconnects [8,9]. Furthermore, the possibility of combining materials of different dimensionality into new van der Waals bonded mixed-dimensional heterostructures has recently been discussed [10]. The realization of such structures relies on the identification of appropriate weakly interacting material components of different dimensionalities.

In the following we shall define a simple geometrical scoring parameter to identify low-dimensional components in existing materials. The scoring parameter is easy to compute and can be applied to large materials databases. We illustrate this by mining the Inorganic Crystal Structure Database [11] (ICSD) and the Crystallography Open Database [12] (COD)

to find materials with clearly identifiable low-dimensional atomic structures. The identified materials consist of weakly interacting components as we demonstrate for 2D materials by comparison with previously calculated exfoliation energies. Apart from being interesting in their own right, the materials components may also form templates for substitution of similar chemical elements to form new materials of different dimensions [5,7].

II. RESULTS AND DISCUSSION

A. Bond-length interval analysis

The definition of the scoring parameter requires, first, that we can identify the dimension(s) of a periodic solid. Given an atom in a bonded cluster, the cluster dimension is given by the rank of the subspace spanned by the atom and its periodically connected neighbors. We refer to this method as the rank determination algorithm (RDA), which is described in detail in the Methods section.

An accurate identification of bonded clusters requires a full electronic structure calculation, where the bond strength and character can be addressed. However, for purposes of screening large materials databases this approach is computationally infeasible. Instead, we use a simple geometric criterion for bonding. We describe two atoms, i and j, as bonded if the distance between them is less than a specified multiple of their covalent radius sum:

$$d_{ij} < k\left(r_i^{\mathrm{cov}} + r_j^{\mathrm{cov}}\right). \tag{1}$$
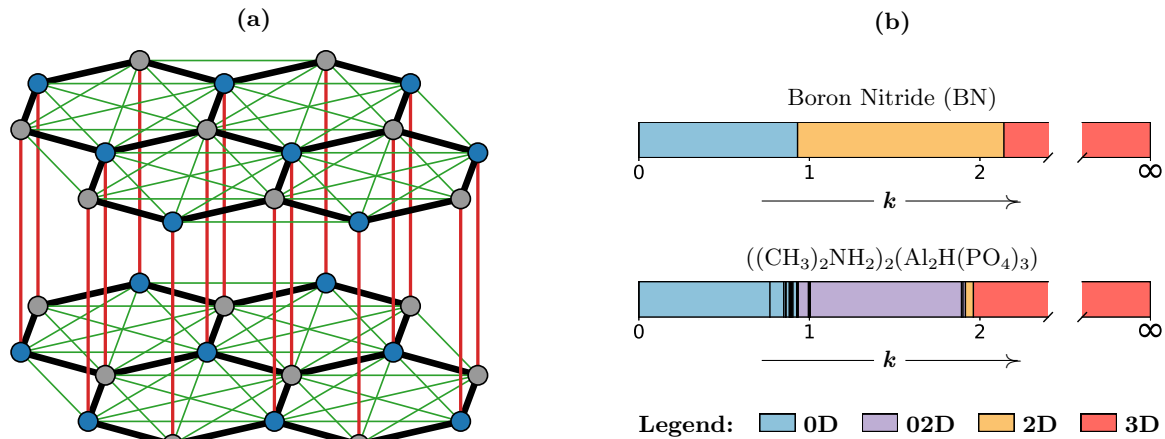
*pmla@mit.edu

FIG. 1. (a) Boron nitride (BN) in a layered structure. Edges are colored according to their effect on the dimensionality classification. Black edges are the strong covalent bonds, which result in a 2D classification. Green edges are longer bonds, which do not change the classification from 2D. Red edges are weak bonds which result in a 3D classification. (b) Illustration of the change in dimensionality classification with increasing k, for boron nitride and a layered aluminophosphate structure with intercalated organic molecules. Larger values of k increase the dimensionality. Multiple intervals with the same dimensionality can exist, though these have different numbers of components. The best classification corresponds to a wide interval starting at $k \approx 1$.

Here, d_{ij} is the distance between atoms i and j, r_i^{cov} and r_j^{cov} are the corresponding covalent radii [13], and k is a variable to be investigated. The latter choice is motivated by the strong dependence of the classification of the dimensionality of a material upon the k value; as illustrated for the boron-nitride structure in Fig. 1(a), too small a k value will underestimate the dimensionality, whereas too large a k value will overestimate it. Rather than attempt to identify a good value of k, we observe that, for any given structure, there exists a finite number of relevant k intervals to investigate.

We start by considering the set of interatomic distances in a material, sorted by increasing k value [where $k = (r_i^{\mathrm{cov}} + r_j^{\mathrm{cov}})/d_{ij}$]. Each interatomic distance corresponds to a possible bond; as shown in Fig. 1(a), bonds can be physical or not. Bonds are inserted one at a time, and at each step the RDA is used to determine the number of components and their dimensionality. Initially, every atom is a separate 0D component; as more bonds are inserted, the number of components decreases and the component dimensionalities increase. The process terminates when a single 3D component is left; i.e., all atoms are contained in the same bonded cluster. This process finds all k-value intervals in which the dimensionality classification is constant, of which there are a finite number. The interval identification method is described in more detail in the Methods section.

Figure 1(b) shows the application of this method to two different layered structures. It can be seen that different dimensionality classifications exist at different k values. Furthermore, the intervals have very different widths; the first interval is of the form $[0, k)$, whereas the last interval is of the form $[k, \infty)$.

B. Defining the scoring parameter

Figures 2(a) and 2(b) show the k intervals for all structures in the ICSD and COD with, respectively, a 2D interval and a 1D interval. In both cases there is a visible cluster of structures in the approximate region $k_1 \approx 1$ and $0.1 \leqslant k_2 - k_1 \leqslant 1.5$. The position of the cluster is intuitive from a bonding perspective. First, if the bonding model and covalent radii exactly described the actual bond lengths, the cluster would lie on the line $k_1 = 1$; the variability in the interval starting points results from the simplicity of the ball-and-stick bonding model. Second, since low-dimensional components are geometrically separated, we should expect a correspondingly wide k interval; it can be seen that easily exfoliable structures such as graphite, boron nitride, and molybdenum disulfide have wide k intervals.

We propose a scoring parameter which distills the above observations of the k-interval plots into a single number:

$$s(k_1, k_2) = f(k_2) - f(k_1), \qquad (2)$$

where

$$f(x) = \frac{c \times \max(0, x - 1)^2}{1 + c \times \max(0, x - 1)^2}. \qquad (3)$$

Here, c is a constant which determines the scale at which a bond is broken. We use $c = 1/0.15^2$, which is chosen so that $s(1, 1.15) = s(1.15, \infty) = 0.5$; slightly different values of the parameter will give similar results. Figure 3 illustrates how a k interval is transformed into a score.

The interval width increases the score, but with diminishing returns as k_1 increases above 1. This avoids the $[k, \infty)$ interval dominating unless k_1 is close to 1, in which case the structure is indeed 3D. Furthermore, k values below 1 are effectively set to 1; this avoids erroneous low-dimensional classifications when $[k_1, k_2] \approx [0, 1]$. In structures with low-dimensional components, the scoring parameter rewards large intercomponent distances. A further convenient property of the scoring scheme is that the interval scores sum to 1. We have found that the best results are achieved by merging k intervals with the same types of dimensionalities [e.g., intervals of the same color in Fig. 1(b)].
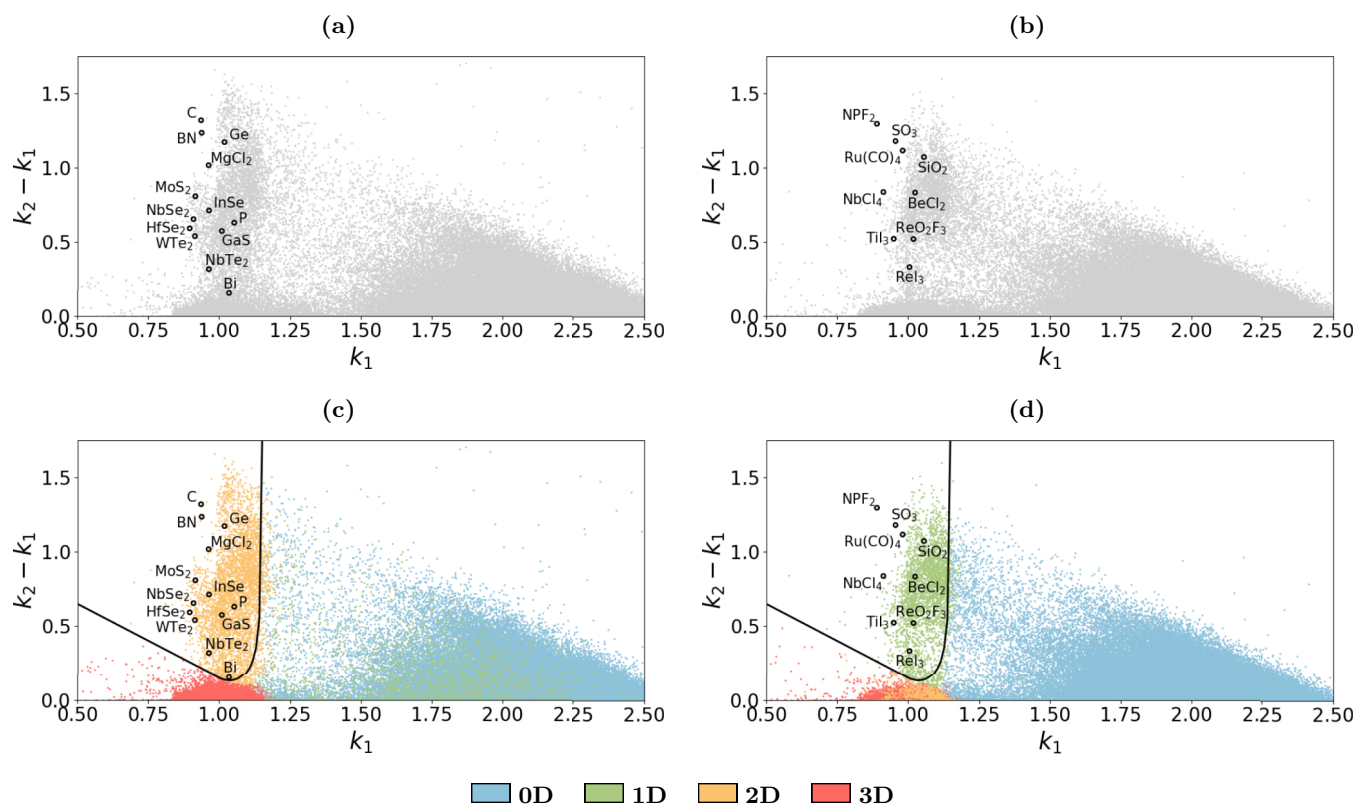
FIG. 2. (a) Interval plot of all structures in the ICSD and COD with a 2D interval, with some well-known structures marked. (b) Same as (a) but for 1D intervals. (c), (d) Same as (a), (b) but with intervals colored according to dimensionality using the scoring parameter. For clarity, mixed-dimensionality structures are not shown. The line shows the contour $s(k_1, k_2) = 0.5$.

The principal motivation of the scoring scheme is to identify the intuitively correct dimensionality classification, by determining whether a k interval lies within a cluster of the type shown in Figs. 2(a) and 2(b). Using the scoring parameter, the structures are colored in Figs. 2(c) and 2(d) according to their dimensionality classification.

The scoring scheme is demonstrated for three structures in Fig. 4. The first material, Al_2O_3, is clearly a bulk crystalline structure. If a single k-value threshold at $k \approx 1$ were used, however, it would result in a misclassification as a layered structure. Similarly, the scoring scheme also ensures that the
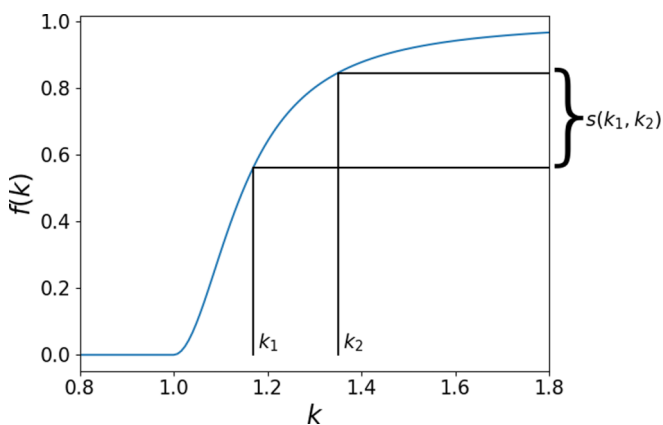

FIG. 3. Variation of $f(k)$ versus k and the dependence of the scoring parameter on the k interval.

$Ti_2Cl_2N_2$ structure is correctly identified as a layered structure. The $AuTe_2$ has an ambiguous classification, lying close to the contour $s_2(k_1, k_2) = 0.5$. In this case the dimensionality classification is sensitive to small changes in the functional form or the parameters of the scoring function. Then, the useful information contained in the scores is not in their exact values, but rather that s_2 and s_3 are approximately equal in value; this can be interpreted as a layered structure with a very small interlayer spacing.

It should be emphasized that the scoring is exclusively based on interatomic distances and atomic sizes, and that it simply assumes that longer bonds tend to be weaker than shorter ones. The physical characters of the bonds, i.e., whether they can be considered covalent, ionic, or of dispersion type, are not revealed. Nonetheless, the coarse treatment of bond lengths is justified by the cluster separation in Fig. 2(b). We will show that the scoring scheme allows for identification of interesting materials, whose properties can then be investigated experimentally or using electronic structure methods.

The scaled bonding criterion described in Eq. (1) is the same one employed by Ashton *et al.* [6] in their study of layered materials. An additive bonding criterion of the form $d_{ij} < r_i + r_j + \Delta$ is used by Mounet *et al.* [5] and Cheon *et al.* [14], using van der Waals radii and elemental radii, respectively. In these works, the material dimension is determined by sampling a range of parameter values (either k or Δ) in a fixed interval, which does not easily permit the construction of a scoring parameter. Cluster dimensionalities

COD 2300448　　COD 9011286　　COD 411179

Al_2O_3　　　　　　　$AuTe_2$　　　　　　　$Ti_2Cl_2N_2$

$s_{2D} = 0.12$　$k = [\, 0.99, 1.05 \,]$　　$s_{2D} = 0.47$　$k = [\, 1.02, 1.14 \,]$　　$s_{2D} = 0.97$　$k = [\, 0.93, 1.81 \,]$

$s_{3D} = 0.88$　$k = [\, 1.05, \infty \,]$　　$s_{3D} = 0.52$　$k = [\, 1.14, \infty \,]$　　$s_{3D} = 0.03$　$k = [\, 1.81, \infty \,]$
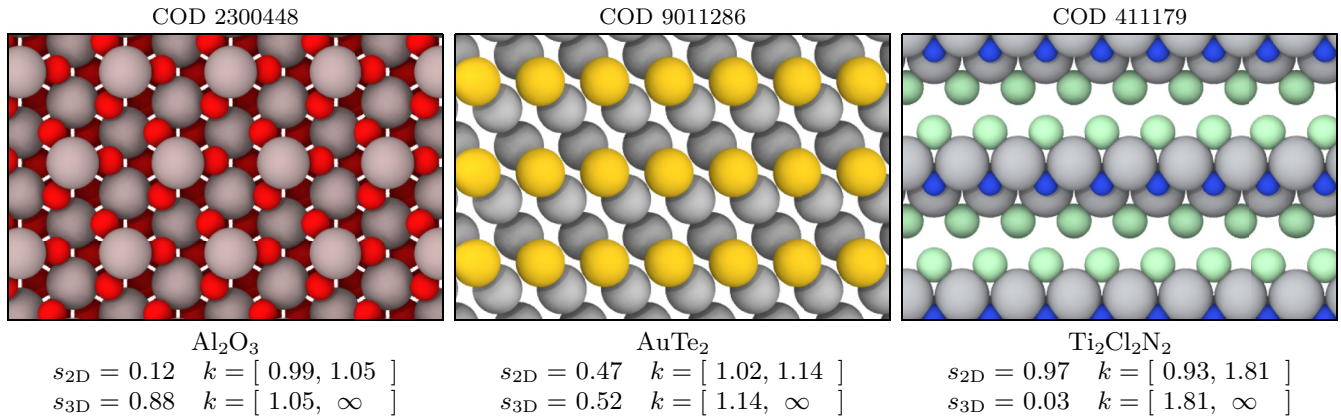
FIG. 4. Structures with successively larger interlayer spacings. All three structures contain intervals for both 2D and 3D classifications. The scoring scheme suggests the most likely dimensionality classification. In cases where multiple reasonable classifications are possible, as in the $AuTe_2$ structure, the ambiguity is reflected in the scoring scheme.

are determined using a topology-scaling algorithm (TSA) [6] (also proposed in [14]), which relates the dimension to the number of bonded clusters as a function of the size of a periodic supercell, or using the RDA [5]. Due to the use of a fixed-size supercell, the TSA and RDA can respectively underestimate and overestimate the number of bonded clusters in certain materials with complex geometries. In the methods section we describe a variant of the RDA which correctly assigns all atoms to bonded clusters without the need to specify a supercell. Except for such complicated cases, however, our definition agrees with the TSA and the supercell RDA.

Other methods for identification of layered materials include the analysis of the packing fraction [15,16], identification of layered slab structures [17], and the use of discrepancies between experimental lattice constants and those predicted by density function theory (DFT) [18]. By identifying structures with interlayer sodium atoms, Zhang *et al.* [19] have investigated promising layered cathode materials for sodium-ion batteries. McKinney *et al.* [20] have extended this search to general "ionic layered" structures.

C. Mining the ICSD and COD

We have analyzed all materials in the ICSD and COD using the proposed scoring parameter. Figure 5 shows examples of materials with different dimensionalities and high values of the scoring parameter.

The database has been filtered in standard ways [5,21] by removing incomplete and/or defective entries, structures with more than 200 atoms, structures with partial occupancies, theoretically calculated structures, and structures with missing hydrogen atoms. Duplicate structures are removed using the structure matcher function of PYMATGEN [22]. The filtering process reduces the initial set of 585 485 CIF files to 167 767 structures. The filtering statistics are shown in Table I.

An overview of the database is shown in Table II. In this table the materials have all been categorized by the dimension(s) with the largest value of the scoring parameter s. In some cases all s values may be fairly small and the classification is then rather uncertain. A large number of materials (105 199) are classified as 0D. These are mostly molecular crystals, which we shall not consider any further here.

The second largest category is the 3D materials. Most of these have a single 3D component, but some of them have two components still with large s values. Two examples $[Ag(B(CN)_4]$ and $Ca(C(CN)_3)_2]$ are shown in Fig. 5. As can be seen from the figure, both materials consist of two identical interpenetrating networks which cannot be disentangled without breaking bonds. The two networks are sufficiently spatially separated to give scoring values above 0.7. (In the figure the two networks are colored red and blue.)

4623 materials are identified as two-dimensional, which is about 2.8% of all materials. This can be compared to for example the study by Mounet *et al.* [5] where they find 1825 out of 108 423 materials (or also about 1.7%) of the materials to be easily or potentially exfoliable. About 2% of the materials are classified as 1D.

There are also some materials with several components of different dimensionality. In particular there are 9459 materials which have one or more 0D components in combination with components of higher dimensionality. These correspond to molecules or molecular ions embedded in the higher-dimensional network. Only a few materials combine 1D, 2D, and 3D components. We find 15 materials combining 1D and 2D. There are 22 materials which combine 1D and 3D components. Three of them are shown in Fig. 5.

While we have made every effort to remove inconsistent structures from the database, an automated filtering is not sufficient given the many different types of errors and partial structures present in the ICSD and COD. The numbers presented here should therefore be taken as only approximate.

A database containing the calculated scoring parameters for all dimensionalities for all compounds in the ICSD and COD is available at the Computational Materials Repository [23].

D. Physical significance of the scoring parameter

Due to the well-defined identification of the 2D materials, the scoring scheme also serves as a simple predictor of exfoliability. Mounet *et al.* [5] have calculated the exfoliation energy (i.e., the binding energy between layers) of 1535 layered materials, and they suggest an energy of 35 meV/$Å^2$ as the

1D

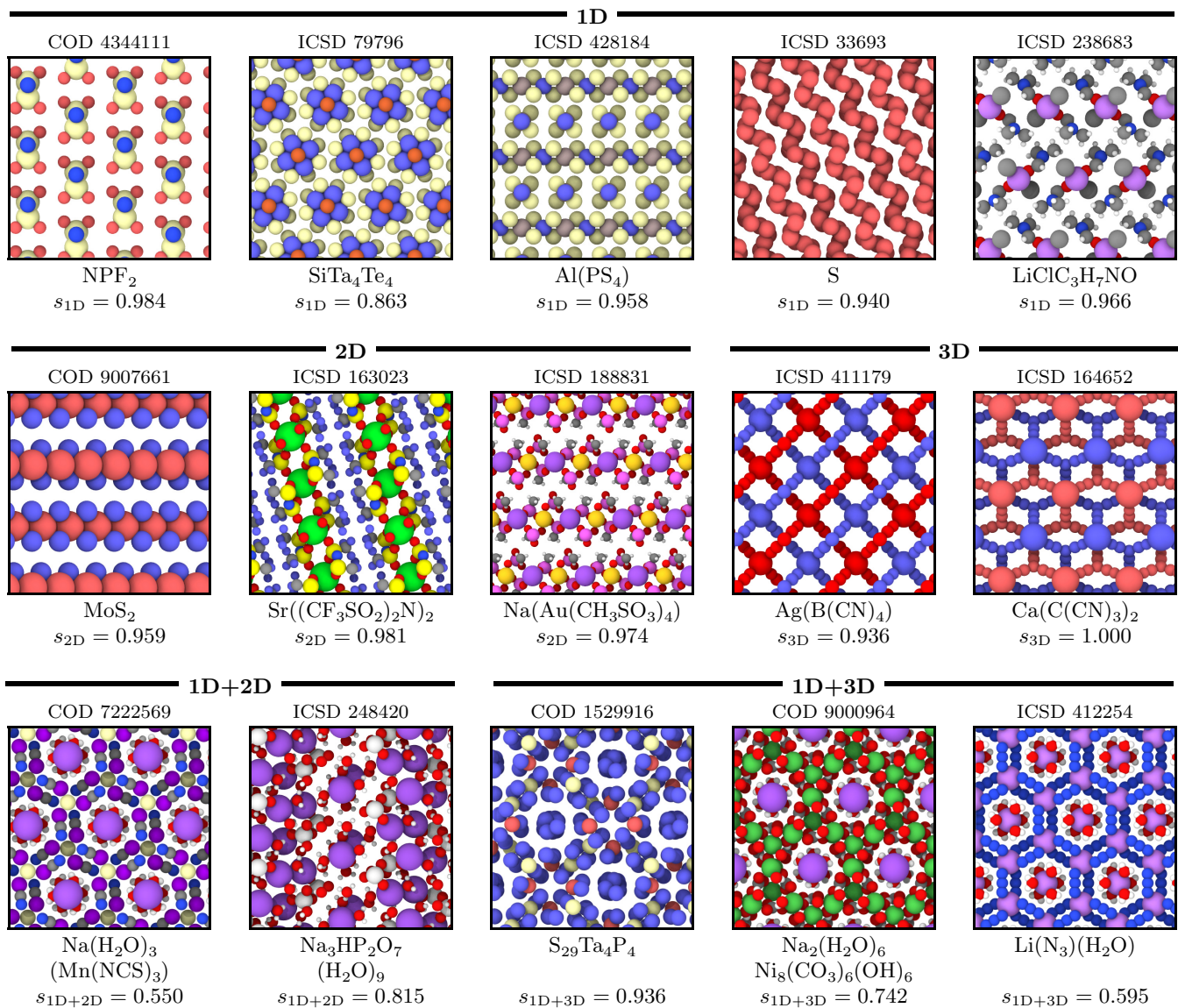| COD 4344111 | ICSD 79796 | ICSD 428184 | ICSD 33693 | ICSD 238683 |
| NPF_2 | $SiTa_4Te_4$ | $Al(PS_4)$ | S | $LiClC_3H_7NO$ |
| $s_{1D} = 0.984$ | $s_{1D} = 0.863$ | $s_{1D} = 0.958$ | $s_{1D} = 0.940$ | $s_{1D} = 0.966$ |

2D / **3D**

COD 9007661	ICSD 163023	ICSD 188831	ICSD 411179	ICSD 164652
MoS_2	$Sr((CF_3SO_2)_2N)_2$	$Na(Au(CH_3SO_3)_4)$	$Ag(B(CN)_4)$	$Ca(C(CN)_3)_2$
$s_{2D} = 0.959$	$s_{2D} = 0.981$	$s_{2D} = 0.974$	$s_{3D} = 0.936$	$s_{3D} = 1.000$

1D+2D / **1D+3D**

COD 7222569	ICSD 248420	COD 1529916	COD 9000964	ICSD 412254
$Na(H_2O)_3$ $(Mn(NCS)_3)$	$Na_3HP_2O_7$ $(H_2O)_9$	$S_{29}Ta_4P_4$	$Na_2(H_2O)_6$ $Ni_8(CO_3)_6(OH)_6$	$Li(N_3)(H_2O)$
$s_{1D+2D} = 0.550$	$s_{1D+2D} = 0.815$	$s_{1D+3D} = 0.936$	$s_{1D+3D} = 0.742$	$s_{1D+3D} = 0.595$

FIG. 5. Smorgasbord of low-dimensional and mixed-dimensional materials, identified by applying the automatic dimensionality classification method to materials in the ICSD.

threshold for "easily exfoliable" materials. They furthermore highlight 11 materials, which they denote as "well-known" 2D materials.

In Fig. 6 we show the calculated exfoliation energies versus the scoring parameter s_2. There is a clear correlation between the scoring parameter and the exfoliation energy with essentially all of the high-scoring materials (say $s_2 > 0.7$) having an exfoliation energy below the threshold. The separation of materials of different dimensionality is also clearly seen here by the low density of points in the region $s_2 \approx 0.3$–0.5. The 11 well-known 2D materials are also shown in the figure. All of them, except Bi_2Te_3, have high-scoring values with $s_2 > 0.7$. Despite its small interlayer distance, Bi_2Te_3 is nonetheless classified as a 2D material, since s_2 is larger than its other scores.

It should be noted that although the exfoliation energy is a highly relevant quantity for the exfoliation process, it is not clear whether an absolute threshold in energy is the best indicator of exfoliability. The exfoliation process involves breaking the bonds between the layers keeping the bonds within the layers intact, so the exfoliation energy should be seen relative to the intralayer bond strengths. While the scoring parameter proposed here does not explicitly involve the energetics, the high-scoring materials have a clear separation between the intra- and intercomponent bond lengths, which can be expected to be a characteristic of easily exfoliable materials.

E. Ranking of low-dimensional materials

In addition to dimensionality classification, the scoring parameter defines an order on materials. We have identified the ten materials in the ICSD and COD with the highest 2D scores, shown in Table III. Widely studied layered structures such as graphene, boron nitride, and magnesium chloride are highly ranked. Some of the remaining structures have much larger unit cells, but are nonetheless clearly van der Waals

TABLE I. Number of structures remaining after each stage of filtering, performed in the order shown. "Defective" structures encompass incorrect CIF files and theoretical structures, and manually identified entries such as misfit compounds, surface structures, and superstructures. Where duplicate structures are found across the two databases, the COD structure is kept.

	COD		ICSD	
	Removed	Remaining	Removed	Remaining
Initial		400731		184754
>200 atoms	185329	215402	7474	177280
Partial occupancy	49015	166387	75659	101621
Missing hydrogen	1470	164917	7184	94437
Defective	10219	154698	5703	88734
Duplicates	15646	139052	60019	28715
Total remaining		167767		

bonded layered structures. It should be noted that the detailed ordering of the top materials is sensitive to the detailed choice of the function $f(x)$ in Eq. (3), whereas the overall classification of the materials is more robust.

Similarly a list of the highest-scoring 1D materials is provided in Table IV. We shall not discuss these materials in depth here, but briefly characterize the top five entries with two or three different chemical elements. For all of these the one-dimensional or chainlike character has already been recognized. NPF_2 consists of chains of alternating nitrogen and phosphorous atoms with the fluorine atoms bound to the phosphorous. The chains can also close on themselves forming ring-shaped molecules. SO_3 is an asbestos-like structure made up of corner-linked SO_4 tetrahedra forming spiraling chains, while the chains in the SiS_2 structure consists of edge-sharing tetrahedra. The two ruthenium compounds form chains of ruthenium atoms. $Ru(CO)_4$ is constructed from planar units with ruthenium in the middle and CO molecules attached with a fourfold rotation symmetry. These units are then stacked forming chains of ruthenium. $RuCl_3$ is in the β phase also called the $ZrCl_3$ structure. Here again the ruthenium atoms form linear chains, but with the chlorine atoms connecting two adjacent ruthenium atoms. It can be noted that $RuCl_3$ also appears as a strongly layered material ($s_2 = 0.933$) in the α phase with prototype $RhBr_3$. We have performed density functional calculations for these highly 1D compounds using the

TABLE II. Number of entries of each dimensionality type found in the ICSD and COD. In the diagonal the number of materials with a single dimension are shown while the off-diagonal entries indicate materials with components of two different dimensionalities. In addition to the single- and bidimensional materials counted here, we have found 16 tridimensional structures with 0D, 1D, and 2D components.

Dimensions	0	1	2	3
0	105199			
1	3503	3285		
2	2946	15	4623	
3	3010	22	0	45148

GPAW code [24,25] and the Atomic Simulation Environment (ASE) [26,27]. The three compounds NPF_2, SO_3, and SiS_2 are found to be nonmagnetic large band gap semiconductors.

In the two compounds with ruthenium chains the distance between the ruthenium atoms are in fact comparable to the bond distance in ruthenium bulk metal. However, the strong couplings to the attached atoms and molecules lead to opening of band gaps. According to the DFT calculations $Ru(CO)_4$ is nonmagnetic while $RuCl_3$ is found to be antiferromagnetic. Details of the calculations can be found in the Supplemental Material [28].

The scoring approach also allows for identification of materials with several components of different dimensionality. Five of these are shown in Fig. 5. The two 1D+2D materials and the last two 1D+3D materials all involve alkali atoms (Na or Li) decorated with water molecules as the 1D component. Note that in the case of $Na(H_2O)_3(Mn(NCS)_3)$ the chains penetrate the 2D layers in the framework, while in the case of $Na_3HP_2O_7(H_2O)_9$ the chains run parallel to the 2D components.

In these materials charge transfer takes place with approximately one electron per alkali atom donated to the 2D or 3D framework. We have investigated this by performing DFT calculations followed by a Bader analysis, [29,30] where an electronic charge is associated with each atom based on a natural per-atom division of the electronic density. For the compounds $Na(H_2O)_3(Mn(NCS)_3)$ and $Na_2(H_2O)_6Ni_8(CO_3)_6(OH)_6$ the chains consist of $Na(H_2O)_3$ units with an electron transfer of 0.85 and 0.90 electrons per unit, respectively. Similarly in the case of $Li(N_3)(H_2O)$, the charge transfer is 0.85 per $Li(H_2O)_3$ unit. $Na_3HP_2O_7(H_2O)_9$ contains chains of $Na(H_2O)_4$ molecular ions with a charge transfer of 0.80 electrons per unit.

The charge transfer in these systems illustrates that the geometrically defined scoring parameter does not only identify components which are exclusively van der Waals bonded to each other. Charge transfer may take place between spatially separated components giving rise to bonding of a more ionic character.

The last compound in Fig. 5 of mixed dimensionality is $S_{29}Ta_4P_4$. It has a very intriguing structure. It consists of spiraling sulfur chains penetrating a 3D network constructed of $TaPS_6$ building blocks. The 3D network itself consists of two identical interpenetrating components. The ability of Ta-P-S compounds to form tunnels has previously been reported [31] and sulfur spirals appear in several compounds. The present compound does not exhibit any charge transfer between the components. Figure 7 shows the calculated density of states for the $S_{29}Ta_4P_4$ compound projected onto the 1D and 3D components. The two components are seen to exhibit different band gaps. This opens the possibility for selectively exciting electrons in one of the components using light with an appropriate wavelength.

III. CONCLUSION

We have defined a simple geometric scoring parameter to identify materials of particular dimensionality. The parameter provides an estimate of the degree to which a given dimensionality is present in the compound. The parameter is easy to calculate and can be applied to large materials databases.
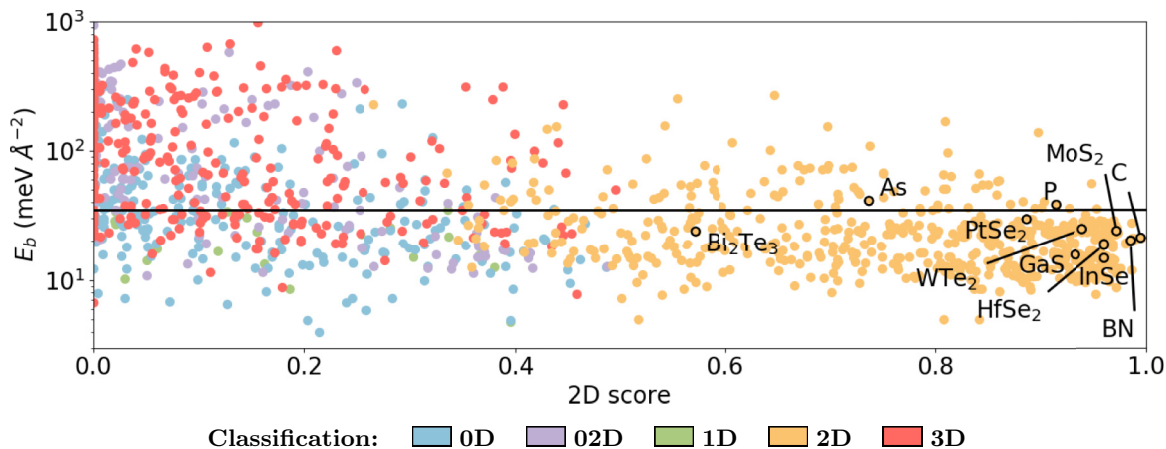
FIG. 6. Binding energies (E_b) vs 2D scores for 1535 layered materials identified by Mounet *et al.* [5], colored according to dimensionality classification. Structures with $E_b < 35$ meV/Å^2 are classified by Mounet *et al.* as *easily exfoliable*. For clarity, 44 01D, five 03D, and four 012D structures are not shown here.

As mentioned in the introduction several computational 2D materials databases are presently under construction while 1D materials and materials of mixed dimensionality have received much less attention. The present approach allows for simple identification of existing 1D or mixed-dimensional materials, which can form templates that can be used to construct larger computational databases for materials of a given dimensionality.

IV. METHODS

A. Component dimensionality

A material will in general consist of several clusters of bonded atoms. Such clusters we term the *components* of the material. The components may have different dimensionalities and they should therefore be investigated separately.

Our definition of material dimensionality of a component is as follows: select an atom in the component, with atomic coordinates $\mathbf{x}_1$. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_i\}$ denote the set of atoms to which the first atom is bonded, and which have the

same fractional coordinates but in different unit cells, i.e., $\mathbf{x}_i = \mathbf{x}_1 + \mathbf{C}^T \mathbf{h}_i$, where $\mathbf{C}$ is the unit cell description and $\mathbf{h}_i$ is an integer vector. Then, the component dimensionality is the rank of the subspace spanned by $\mathbf{X}$:

$$\dim(\mathbf{X}) = \text{rank}(\{\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \dots, \mathbf{x}_i - \mathbf{x}_1\}). \quad (4)$$

This definition (illustrated in Fig. 8) accommodates both corrugation and thickness. While $\mathbf{X}$ is an infinite set for all but 0D components, $\dim(\mathbf{X})$ can be determined in a finite number of steps by exploiting the periodicity of a material.

As described above, determination of the dimension of a material requires an analysis of its constituent bonded clusters, or *components*. To find the dimension of a component, the rank determination algorithm (RDA) of Mounet *et al.* [5] uses a supercell of fixed size with open boundary conditions. If the supercell is too small, the number of components might be overestimated.

Conversely, the topological scaling algorithm (TSA) of Ashton *et al.* [6] uses periodic unit cells, which can underestimate the number of components by forming improper connections between them. By improper connections, we mean components which are disconnected in the infinite crystal but are connected due to the periodic cell chosen.

TABLE III. Top ten "most 2D" materials in the ICSD and COD, as ordered by the interval scoring method in Eq. (2). Since the databases contain many layered polymers with large interlayer spacings, we have not included structures containing hydrogen in this list.

Source	ID	Compound	s	k_1	k_2
COD	1000410	TlAlF$_4$	0.987	0.992	2.330
COD	9000046	C	0.986	0.933	2.251
ICSD	27987	BN	0.984	0.935	2.170
ICSD	248325	C$_3$N$_4$	0.983	0.910	2.155
ICSD	187384	Rb(Au(CF$_3$SO$_3$)$_4$)	0.982	1.010	2.313
ICSD	163023	Sr((CF$_3$SO$_2$)$_2$N)$_2$	0.981	1.014	2.516
COD	1525422	K$_3$Mn(CN)$_6$	0.980	0.935	2.039
COD	1534338	MgCl$_2$	0.977	0.959	1.976
ICSD	161278	B$_3$C$_{10}$N$_3$	0.977	0.965	1.968
COD	2242431	Cs(N(SO$_2$CF$_3$)$_2$)	0.977	0.965	1.968

TABLE IV. Top ten "most 1D" hydrogen-free materials in the ICSD and COD, as ordered by the interval scoring method in Eq. (2).

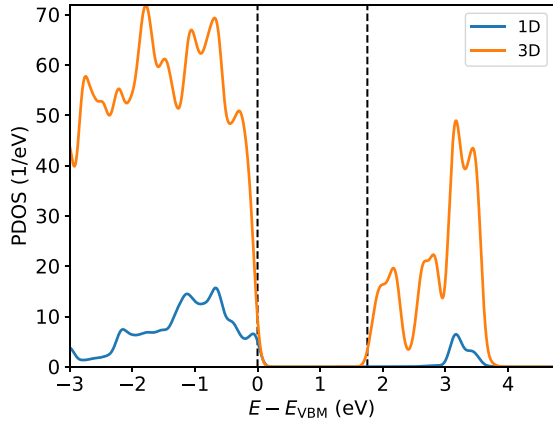| Source	ID	Compound	s	k_1	k_2
COD	4344111	NPF$_2$	0.984	0.889	2.184
COD	9010982	SO$_3$	0.983	0.954	2.133
ICSD	72577	Ru(CO)$_4$	0.982	0.979	2.095
ICSD	47183	CuCo(CO)$_4$	0.980	0.926	2.044
ICSD	291211	SiS$_2$	0.967	0.986	1.815
ICSD	22090	RuCl$_3$	0.967	0.940	1.812
ICSD	415951	V(AlCl$_4$)$_2$	0.965	0.990	1.784
ICSD	78778	CrF$_4$	0.964	0.985	1.780
ICSD	428185	AlPS$_4$	0.964	0.995	1.771
ICSD	419661	CrF$_5$	0.962	0.996	1.751

FIG. 7. Density of states projected onto the 1D and 3D components of the $S_{29}Ta_4P_4$ compound. The energy is relative to the valence band maximum (VBM) and the vertical dashed lines indicate the band edges of the smallest band gap. The electronic spectrum is calculated using the GLLB-SC exchange-correlation functional.

The problem of improper connections is illustrated with a contrived example in Fig. 9, which shows the side view of a selection of periodic helical structures. We define an n-helix as a structure which has n components, whose jth component has coordinates given by

$$x_j = \sin \frac{2\pi(t+j)}{n}, \qquad y_j = \cos \frac{2\pi(t+j)}{n}. \qquad (5)$$

The number of components is dependent on the size (along the t axis) of the periodic cell. In the formulation given in Eq. (5) any integer is a valid cell length.

Figure 9 shows how the number of components changes with varying cell periodicity. In general, the number of components for an n-helix with periodicity m is given by $\gcd(n, m \mod n)$. In order to avoid improper connections between components, a periodic cell of size n is needed. This is further complicated for cells containing multiple n-helices of different sizes. In this case, the correct size of the periodic cell is given by $\text{lcm}(n_1, n_2, n_3, \ldots)$, where n_i denotes the

number of components in the ith n-helix. For example, a structure containing a 5-helix, a 6-helix, and a 7-helix requires a periodic cell of length 210. This cell is so large that it is unlikely that it would be tested using the existing methods.

While the example described here is contrived, self-penetrating helical networks have been assembled experimentally [32,33]. Furthermore, the problem illustrated has practical consequences: an incorrect periodic cell (such as the use of a primitive unit cell) causes the interpenetrating polymer networks shown in Fig. 5 to be misclassified as 2D materials.

B. Algorithm

Component dimensionalities can be identified using a modified breadth-first-search (BFS) algorithm, shown in Algorithm 1. In standard BFS, the search terminates when all nodes have been visited. Here, we terminate the search when the rank of the subspace spanned by a component (i.e., the dimensionality) can no longer increase. The rank of a set of points is defined as

$$\text{rank}(\mathbf{v}) = \begin{cases} -1, & \text{if } \mathbf{v} = \emptyset, \\ \text{rank}_M(\mathbf{v} - \vec{v}_1), & \text{otherwise,} \end{cases} \qquad (6)$$

where rank_M denotes the standard matrix rank.

In this algorithm, components in the aperiodic primitive unit cell are the graph vertices, and connections between components (across unit cell boundaries) are graph edges. We note that, by definition, no edges exist between components within the same cell.

The input to the algorithm (line 1) is a set of graph edges ($\mathbf{E}$) and a component (c) whose dimension we wish to determine. We maintain a set of visited or *seen* vertices (line 2) and a set of visited vertices for each of the n components in the aperiodic primitive unit cell (line 3). A vertex queue is maintained whose elements consist of a component index and cell coordinates. The queue is initialized with the component c in the cell with coordinates $O = [0, 0, 0]$ (line 4). The algorithm runs until the queue is empty (line 5). The first element in the queue is extracted and removed (lines 6 and 7). If the element has already been visited it is skipped (line 8); otherwise it is added to the set of visited elements (line 11). If the addition of the vertex serves to increase the rank of the set of visited vertices (line 12), it is added to the set (line 13).

New vertices in adjacent cells are generated from the edge list. For a component i, the edge list $\mathbf{E}_i$ (line 15) specifies the neighboring components (j) and the cell offset ($\vec{\Delta}$), from which the coordinates of the neighboring cell can be calculated (line 16). If the neighbor element has already been visited it is either skipped (line 17), or added to the queue (line 21) if it serves to increase the rank of the set of visited vertices (line 20). When the queue is empty, the rank of the component is returned (line 25).

C. Interval identification

The purpose of the modified method (described in Algorithm 2) is to identify intervals in k in which the dimensionality classification is constant.
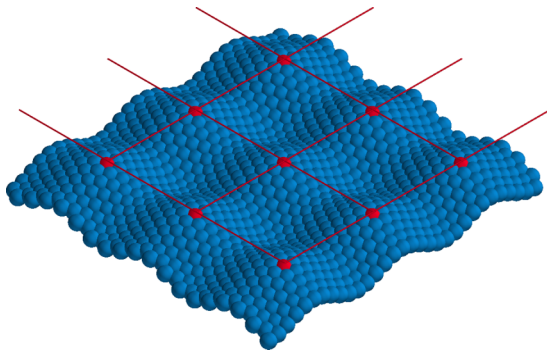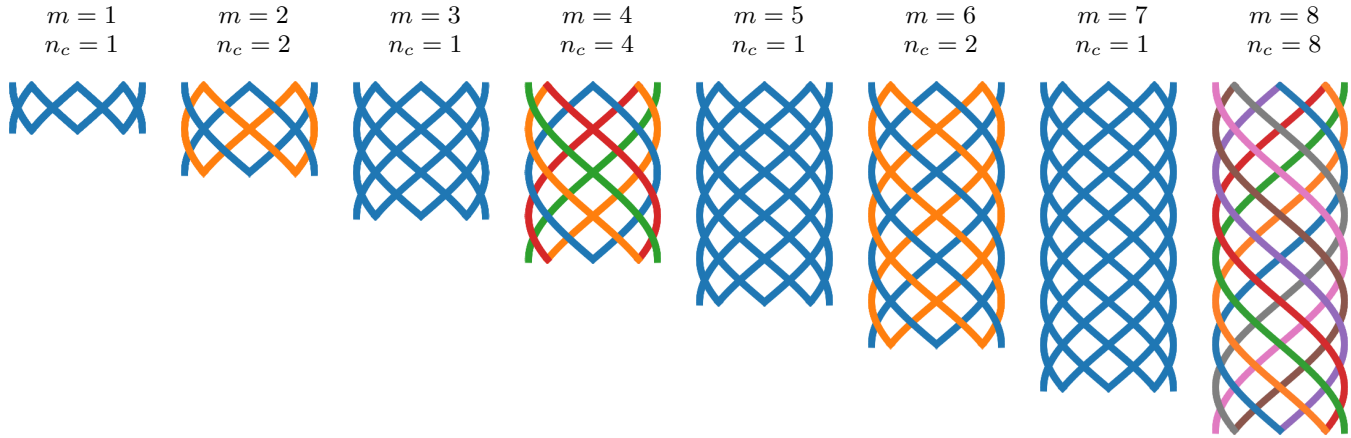

FIG. 8. Cut-out of a periodic corrugated 2D component. The component dimensionality can be found by selecting any atom in the component, and identifying all other atoms in the bonded cluster with the same fractional coordinates, shown here in red. The rank of the subspace spanned by these atoms (here, a plane) determines the dimensionality.

FIG. 9. Improper connections between components in n-helix structures, here for $n = 8$. The number of components is denoted by n_c. The infinite structure contains 8 components. Any number of repetitions, $m \geqslant 1$, of the cell for which $m \mod 8 \neq 0$ results in improper connections between components.

The input to the algorithm (line 1) is the set of all possible edges, sorted according to their k values, from lowest to highest. Each element in this set, $(k, i, j) \in \mathbf{E}$, contains the k value of the edge and the indices, i and j, of the vertices it connects. Periodic boundary conditions must be taken into account when generating this set. Due to the periodicity this set is infinitely large; the relevant (finite) subset, however, can be generated incrementally.

The algorithm proceeds by inserting edges from $\mathbf{E}$ into the graph, one by one (line 5). For every edge in the primitive cell, the corresponding number of edges are inserted into the supercell. Connected components in both the primitive cell and the supercell are identified after each edge insertion (line 6), from which a *component histogram* is calculated (line 8). The histogram, $\mathbf{h}$, contains the number of 0D, 1D, 2D, and 3D components present. Prior to any edge insertion, only 0D components are present, which is reflected in the initial state of the histogram (line 3). If an edge insertion produces a change in the component histogram (line 9), the k-interval is added (line 10) to the set of results (line 4). The algorithm terminates when the histogram consists only a single 3D component (line 12). A 3D interval is added to the results (line 13), which implicity contains all uninserted edges in $\mathbf{E}$: once the dimensionality is fully 3D, no further edge insertions can change the classification.

The algorithms developed here are included in the ASE [27] library.

Algorithm 1. Pseudocode for calculating component dimensionality.

```
1:  procedure CALCULATEDIMENSIONALITY(E, c)
2:      s := ∅
3:      v := {∅  ∀i ∈ 1…n}
4:      Q := {{c, O}}
5:      while Q ≠ ∅ do
6:          {i, p⃗} := Q₁
7:          Q := Q \ Q₁
8:          if {i, p⃗} ∈ s then
9:              Continue
10:         end if
11:         s := s ∪ {i, p⃗}
12:         if rank (vᵢ ∪ {p⃗}) > rank (vᵢ) then
13:             vᵢ := vᵢ ∪ {p⃗}
14:         end if
15:         for {j, Δ⃗} ∈ Eᵢ do
16:             q⃗ := p⃗ + Δ⃗
17:             if {j, q⃗} ∈ s then
18:                 Continue
19:             end if
20:             if rank (vⱼ ∪ {q⃗}) > rank (vⱼ) then
21:                 Q := Q ∪ {j, q⃗}
22:             end if
23:         end for
24:     end while
25:     return rank (v_c)
26: end procedure
```

Algorithm 2. Pseudocode for finding all dimensionality intervals.

```
1:  procedure FINDINTERVALS (E)
2:      k_prev := 0
3:      h_prev := [n_atoms, 0, 0, 0].
4:      R = ∅
5:      for (k, i, j) ∈ E do
6:          Add edge between vertices i and j
7:          Identify connected components
8:          Update h
9:          if h ≠ h_prev then
10:             R := R ∪ {(k_prev, k, h_prev)}
11:         end if
12:         if h = [0, 0, 0, 1] then
13:             return R ∪ {(k, ∞, h)}
14:         end if
15:         k_prev := k
16:         h_prev := h
17:     end for
18: end procedure
```

[1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).

[2] G. R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M. S. Strano, V. R. Cooper, L. Liang, S. G. Louie, E. Ringe, W. Zhou, S. S. Kim, R. R. Naik, B. G. Sumpter, H. Terrones, F. Xia, Y. Wang, J. Zhu, D. Akinwande, N. Alem, J. A. Schuller, R. E. Schaak, M. Terrones, and J. A. Robinson, ACS Nano **9**, 11509 (2015).

[3] A. C. Ferrari, F. Bonaccorso, V. Fal'ko, K. S. Novoselov, S. Roche, P. Bøggild, S. Borini, F. H. L. Koppens, V. Palermo, N. Pugno, J. A. Garrido, R. Sordan, A. Bianco, L. Ballerini, M. Prato, E. Lidorikis, J. Kivioja, C. Marinelli, T. Ryhänen, A. Morpurgo, J. N. Coleman, V. Nicolosi, L. Colombo, A. Fert, M. Garcia-Hernandez, A. Bachtold, G. F. Schneider, F. Guinea, C. Dekker, M. Barbone, Z. Sun, C. Galiotis, A. N. Grigorenko, G. Konstantatos, A. Kis, M. Katsnelson, L. Vandersypen, A. Loiseau, V. Morandi, D. Neumaier, E. Treossi, V. Pellegrini, M. Polini, A. Tredicucci, G. M. Williams, B. Hee Hong, J.-H. Ahn, J. Min Kim, H. Zirath, B. J. van Wees, H. van der Zant, L. Occhipinti, A. Di Matteo, I. A. Kinloch, T. Seyller, E. Quesnel, X. Feng, K. Teo, N. Rupesinghe, P. Hakonen, S. R. T. Neil, Q. Tannock, T. Löfwander, and J. Kinaret, Nanoscale **7**, 4598 (2015).

[4] M. Zeng, Y. Xiao, J. Liu, K. Yang, and L. Fu, Chem. Rev. **118**, 6236 (2018).

[5] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, and N. Marzari, Nat. Nanotechnol. **13**, 246 (2018).

[6] M. Ashton, J. Paul, S. B. Sinnott, and R. G. Hennig, Phys. Rev. Lett. **118**, 106101 (2017).

[7] S. Haastrup, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen, A. C. Riis-Jensen, J. Gath, K. W. Jacobsen, J. J. Mortensen, T. Olsen, and K. S. Thygesen, 2D Mater. **5**, 042002 (2018).

[8] M. A. Stolyarov, G. Liu, M. A. Bloodgood, E. Aytan, C. Jiang, R. Samnakay, T. T. Salguero, D. L. Nika, S. L. Rumyantsev, M. S. Shur, K. N. Bozhilov, and A. A. Balandin, Nanoscale **8**, 15774 (2016).

[9] A. Geremew, M. A. Bloodgood, E. Aytan, B. W. K. Woo, S. R. Corber, G. Liu, K. Bozhilov, T. T. Salguero, S. Rumyantsev, M. P. Rao, and A. A. Balandin, IEEE Electron Devices Lett. **39**, 735 (2018).

[10] D. Jariwala, T. J. Marks, and M. C. Hersam, Nat. Mater. **16**, 170 (2017).

[11] G. Bergerhoff, R. Hundt, R. Sievers, and I. Brown, J. Chem. Inf. Comput. Sci. **23**, 66 (1983).

[12] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N. R. Serebryanaya, P. Moeck, R. T. Downs, and A. Le Bail, Nucleic Acids Res. **40**, D420 (2012).

[13] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, Dalton Trans. **2008**, 2832.

[14] G. Cheon, K.-A. N. Duerloo, A. D. Sendek, C. Porter, Y. Chen, and E. J. Reed, Nano Lett. **17**, 1915 (2017).

[15] T. Björkman, A. Gulans, A. V. Krasheninnikov, and R. M. Nieminen, Phys. Rev. Lett. **108**, 235502 (2012).

[16] S. Lebègue, T. Björkman, M. Klintenberg, R. M. Nieminen, and O. Eriksson, Phys. Rev. X **3**, 031002 (2013).

[17] P. Gorai, E. S. Toberer, and V. Stevanović, J. Mater. Chem. A **4**, 11110 (2016).

[18] K. Choudhary, I. Kalish, R. Beams, and F. Tavazza, Sci. Rep. **7**, 5179 (2017).

[19] X. Zhang, Z. Zhang, S. Yao, A. Chen, X. Zhao, and Z. Zhou, npj Comput. Mater. **4**, 13 (2018).

[20] R. McKinney, P. Gorai, S. Manna, E. Toberer, and V. Stevanović, J. Mater. Chem. A **6**, 15828 (2018).

[21] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, npj Comput. Mater. **1**, 15010 (2015).

[22] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Comput. Mater. Sci. **68**, 314 (2013).

[23] Computational Materials Repository, https://cmr.fysik.dtu.dk/lowdim/lowdim.html.

[24] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).

[25] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Hakkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, J. Phys.: Condens. Matter **22**, 3202 (2010).

[26] S. R. Bahn and K. W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002).

[27] A. Larsen, J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. Groves, B. Hammer, C. Hargus, E. Hermes, P. Jennings, P. Jensen, J. Kermode, J. Kitchin, E. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, J. Phys.: Condens. Matter **29**, 273002 (2017).

[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.3.034003 for DFT-calculated electronic spectra.

[29] R. F. W. Bader and R. F. Bader, *Atoms in Molecules: A Quantum Theory*, International Series of Monographs on Chemistry (Clarendon Press, Oxford, 1990).

[30] W. Tang, E. Sanville, and G. Henkelman, J. Phys.: Condens. Matter **21**, 084204 (2009).

[31] M. Evain, S. Lee, M. Queignec, and R. Brec, J. Solid State Chem. **71**, 139 (1987).

[32] D.-R. Xiao, Y.-G. Li, E.-B. Wang, L.-L. Fan, H.-Y. An, Z.-M. Su, and L. Xu, Inorg. Chem. **46**, 4158 (2007).

[33] G.-P. Yang, L. Hou, X.-J. Luan, B. Wu, and Y.-Y. Wang, Chem. Soc. Rev. **41**, 6992 (2012).