Materials Research Express

PAPER

The mining of materials with similar electronic properties from the Crystallographic Open Database (COD)

G Carbajal-Francó, E Rendón-Lara, I M Abundez-Barrera and A Vásquez-Aguilar

Division of Graduate Studies and Research, Tecnológico Nacional de México/Instituto Tecnológico de Toluca, Av. Tecnológico s.n. Metepec Estado de México. C.P. 52149 México

E-mail: gcarbajal@toluca.tecnm.mx

Keywords: properties data mining, materials informatics, crystallographic similarity, materials integration, first derivative

Abstract

The finding of a material with the precise properties needed to solve a specific issue is the first topic that needs unraveling when an application is projected. One approach to find a material with a specific property value is to study a different but linked property. The aim of this research is to find materials with similar Electronic Band Structures (EBS); which in a simulation typically contain more than 1,000 ordered pairs of data. Our approach is, instead of calculating the similarities between the EBS of different materials, to calculate the similarity between their crystalline structures, and then the similarity between the EBS of the resulting similar compounds is tested. The software system developed in this research finds materials with similar crystallography, then the similarity of the compounds is tested by comparing the DFT modeled Electronic Band Diagrams (EBDs). The crystallographic data was mined from the Crystallography Open Database (COD) in the form of CIF files; that were used to calculate the x-ray diffraction (XRD) data using REFLEX, a component of Materials Studio. The plane presence, position and intensity of the peaks from the XRD data, were used to calculate the similarity between materials. With the list of similar materials from the previous process and the correspondent CIF files, the CASTEP code (from Materials Studio) was used to calculate the EBDs. In this work, three different materials were analyzed: CdTe, CdSe and GaAs. As results, 2D maps showing 50 compounds with the highest similarities are shown and for the EBD calculation of properties data mining, materials informatics, crystallographic similarity, materials integration, first derivative

1. Introduction

Materials Science is a multidisciplinary field of study. In the beginning, only chemists and physicist worked on this area but as the number of new materials, applications and requirements increased, the spectrum of scientists expanded to new areas that time ago seemed to be far away from the original fields of research. On the other hand, Computational Sciences were limited to areas such as data management and storage, control of processes, communications, and the administration of large amounts of information stored in enormous data bases. Materials scientist produce vast amounts of data as result of synthesis, study and characterization of materials; information that is often stored in large databases, normally used for identification of materials, comparison and calculation of properties [1], and recently for Materials Integration and data mining, being the later a strong bridge between Materials Science and Computational Science.

There already are efforts on mining data from large data-bases for the study of organic [2] and inorganic [3] compounds, as there are studies about complex molecules such as proteins [4]. Most of these researches use data mining for prediction of structures and analysis of properties based on the atomic constitution and distances [5, 6], and applying space group theory along with formulas for structure prediction of materials and phases that have not yet been discovered [3], or finding materials with specific properties, based on ab initio calculations,
grouped according to some related behavior of the bravais lattice, space group or atomic position within the crystal cell [7–9].

Materials with similar atomic structures are good candidates to exhibit similar properties, so there have been studies about similarity since the decade of the seventies [10]. Among the studied properties, the EBS and electronic properties have been studied by ab initio calculations of all the electronic properties of interest of a large number of inorganic materials (22,000), obtained from the Inorganic Crystal Structure Database (ICSD) [7], then the similarity of the simulated properties was calculated by data-mining the outcome data; this approach resulted in a list of novel materials that present the desired properties. Although this method is very effective, it is also very time and computational-resources consuming due to the large set of materials and simulations realized. In this work, the EBDs were only calculated for materials that showed to be similar according to our data mining system.

In this research, the EBDs of three different compounds were analyzed and compared with the EBDs of materials presenting similar crystallography, all of them mined from the large amounts of data stored in the Crystallography Open Database (COD) [11, 12]. Our premise is that the solutions of the Schrödinger equation, for an electron confined in a periodic distribution of potentials, are based on the Kronig–Penney model [13–16] and follow the Bloch’s theorem [17–20], both based on the crystallography of the compound. One standard tool to analyze the crystallography of the materials is x-ray diffraction (XRD), because it relates the inter-planar distance \(d_{hkl}\) to the angle of incidence \(\theta\), where the height of a diffraction peak is related to the atomic planar density (and structure factor) [21, 22]. To calculate the similarity between the compounds during the data mining process, the space group theory and related theories were not taken into account, only the atomic configuration of the material. On these bases, we postulate that materials with similar XRD diffractograms present a high probability of having similar electronic properties, even if they do not belong to the same space group, being the only difference the electronic configuration of the atoms in the compound.

2. Methodology

The system developed in this research is composed by two blocks (figure 1). Block 1 is dedicated to download and process the data from the COD. Once the data is extracted from the COD, it is possible to mine the database looking for compounds with a specific stoichiometry such as AB, AB\(_2\), A\(_2\)B\(_3\), and so forth. The XRD diffractograms with the relative intensity \(I_{rel} (I/I_{max})\), the Bragg angle \(\theta\) and the crystallographic planes information, are calculated from the .CIF files [23] by the Reflex software, part of the Materials Studio suite from Dassault Systèmes-BIOVIA.
Block 2 is composed by the interface that takes the input from the user and feeds it into the core of the Data Mining system; this code measures the distance between peaks and organize the final list starting with the compounds with the shortest summation of Euclidean distances between peaks, and extends up for a given number of materials defined by the user.

2.1. Block 1
This block is composed by automatons that download the entire COD and process the .CIF files [23] with the Reflex module of Materials Studio. To calculate the crystalline planes of the mined compounds, the 2\(\theta\) interval for the XRD calculations in Reflex was set from 10\(^\circ\) to 90\(^\circ\). Diffraction peaks produced by Cu K\(\alpha\)1 and K\(\alpha\)2 wavelengths of 1.540562 Å and 1.54439 Å respectively, were taken into account. We use these wavelengths because Cu is the most common experimental source of x-ray radiation for XRD, and the position of the peaks could be significant for the user. If preferred, the user can define the wave length of the x-ray source to perform the XRD calculations depending on the specific interest of the case. The results of the XRD computation for each material were refined by removing systematic absent peaks and stored in 2\(\theta\) versus \(I_{rel}\).CSV files. Values of \(I_{rel}\) are normalized to 100% for the highest XRD peak, and only compounds with 10 or more diffraction peaks were taken into account.

2.2. Block 2
Block two is the core of the data mining system; it includes the interface that gets the data from the user for a specific search. The data required from the user is the COD .CIF number or the compound formula, the number of peaks that will be used to calculate the similarity between the given compound and the rest of the database, and the required number of similar materials. Compounds with fewer peaks than the number of peaks given by the user are ignored.

To calculate the similarity between compounds, the data in the .CVS files is sorted from highest to lowest value of \(I_{rel}\). The similarity between compounds, defined as \(S\), is the Euclidean distance between peaks according to equation (1):

\[
S = \sum_{i} \sqrt{(2\theta_{C_{i}} - 2\theta_{T_{i}})^{2} + (I_{rel-C_{i}} - I_{rel-T_{i}})^{2}}
\]

where \(C\) corresponds to the compound of interest, \(T\) to the compound under analysis, and \(i\) corresponds to the peak that is being tested.

This definition of similarity results in a list of compounds ordered accordingly to the absolute value of \(S\), which can only be plotted in one dimension. With the purpose of having a second descriptor that includes the increase or decrease of the 2\(\theta\) and \(I_{rel}\) values, we used the angle given by equation (2):

\[
\phi = \arctan\left(\frac{\sum_{i} (I_{rel-C_{i}} - I_{rel-T_{i}})}{\sum_{i} (2\theta_{C_{i}} - 2\theta_{T_{i}})}\right) = \arctan\left(\frac{\sum_{i} \Delta I_{rel}}{\sum_{i} \Delta 2\theta}\right)
\]

With these two descriptors, we defined a point in a plane that can be located by the vector of magnitude \(S\) and angle \(\varphi\). In this plot, the abscissa is related to the total deviation of the inter-planar distances \(d_{hkl}\) and the ordinate is related to the difference of the atomic density of the planes and the structure factor \(F_{hkl}\) of the tested compound [14, 17] (figure 2).

2.3. Validation of results
For validation of the similarity, the Electronic Band Structure (EBS) and Density of States (DOS) of each material was calculated using CASTEP code. EBS and DOS data were calculated with an x-x-LDA functional [24, 25], an SCF tolerance of 2 \(\times\) \(10^{-6}\) eV/atom, a k-point separation of 0.8 Å\(^{-1}\), and the norm conserving pseudopotential scheme. The EBD and the DOS were calculated with 12 empty bands, a separation of 0.025 Å\(^{-1}\), and a band energy tolerance of \(10^{-5}\) eV. The similitude of the location of the maximums of the valence band (VB), the minimums of the conduction band (CB) and the shape of the bands will be consider as the similarity between compounds. In this regard, the derivative of the CB and VB are calculated and plotted. These plots do not take into account the \(E_{g}\) (as it is a constant) but the shape of the bands because the derivative is defined as the slope of a function at a given point. Moreover, the value of 0 in a derivative indicates a local minimum or maximum of the function, so the crossing of the abscissa in the plots is used to identify the maximum of the VB and the minimum of the CB.

The tested materials for this work were the result of a search for the most rapidly-interest growing AB materials in Google scholar [26]. The percentage of increment on research-interest was calculated by dividing the hit results since 2017 by the total hits from any time.
3. Results and discussion

3.1. Materials mined from the database
For this research, the three AB compounds with fastest-growing research-interest found since 2017 were CdTe with 24.74%, CdSe with 8.89% and GaAs with 4.60%. The entire COD database was downloaded from the website containing 375,000 compounds (February 2017), the number of AB mined compounds is 2,644 and after refinement only 2,565 compounds were taken into account because they present 10 or more peaks.

Figure 3 shows the map of 50 compounds similar to CdTe. Most of the compounds present a larger \( d_{hkl} \) inter-planar distance, they are distributed between quadrants II and III and there is not a clear tendency on the structure factor. Only another CdTe phase and three HgTe different phases presented a smaller \( d_{hkl} \) distance and smaller planar density (quadrant IV). The material with almost the same \( d_{hkl} \) value and the largest S is CdPo, which is on the negative y-axis between quadrants II and IV indication of less dense planes. The largest S value in the map is not greater than 65 and the smallest value is around 12.

Figure 2. 2D plot showing the similarity S and the relation between of the tested compound (●) and the compound of interest (at origin).

Figure 3. 2D map of 50 compounds similar to CdTe.
Compounds similar to CdSe are more homogenously distributed on the four quadrants (figure 4), being the more similar materials in quadrant III, indicating that they exhibit larger $d_{hkl}$ distances and smaller planar density. For this material, the largest $S$ value is around 175, and the smallest value is $\sim 45$.

Most of gallium arsenide similar compounds are in quadrant II, and around the positive x-axis in quadrant I (figure 5); indicating that the majority of the materials present a smaller $F_{hkl}$. For this material the largest $S$ value on the map is close to 75 and a smallest value near 4, being GaAs the material with the most similar compounds.
In order to process the information and test the crystallographic similarities by comparing the EBDs, the first 7 + items of each list were taken into account. Tables 1, 2 and 3 present the materials and their similar compounds.

From the tables of the three materials, it can be seen that one mined compound can present more than one CIF file similar to the material of interest; this is due to the crystallographic variations that the compound can present due to factors such as external pressure; an example of this arises from investigating CdO. Table 4 lists the 30 .CIF codes of the most similar-to CdO compounds, and only 5 of them are not CdO.

The first three CIF codes in the table 4 were reported in three different references, but from entry number 3 up to the end of the list (except entry 7) there are 22 different CIF codes that correspond to a CdO variation. The 22 different codes were reported by Jianzhong Zhang [30] in a study that relates the applied pressure and the volume of the material to the crystalline structure of the CdO. In figure 6, it is evident the effect of the pressure on the structure of the cadmium oxide, all the CdO structures reported on table 4 are aligned in quadrant II with an increasing value of \( S \) and a slightly increase of the difference between their inter-planar distances; an expected result of the increasing pressure of compression applied to the CdO [30].

Findings like this are important for Materials Integration because, in this specific case, the results present data from experiments realized with real materials that are slightly different in terms of their crystallinity, data that can be used to investigate, for example, the tuning of the electronic properties of the materials without changing its composition.

### 3.2. Electronic band structure

#### 3.2.1. Cadmium telluride

In the case of CdTe, we only used the CIF file number 1540817 for the EBD simulation because it was not our intention, in this work, to show the differences in EBD obtained for the same material with different crystallinity (CIF file 1010539). Figure 7 shows the EBDs and DOS of the six different compounds listed in table 1.
The distribution of the electronic bands is similar for all the compounds except for the CuI, which presents two higher DOS bands below the Fermi level. In order to study the VB, CB and Eg, the bands were extracted from the band diagrams dataset, and their derivatives were calculated and plotted in figure 8.

The six most similar compounds to CdTe exhibit a similar shaped-band structure, presenting the VB maximums at R and G wave vectors; meanwhile the minimums of the CB are at R and G wave vectors. All the compounds present a direct electronic transition on the wave vector G, and an indirect electronic transition between wave vectors G and R. Most of the materials present a structure that privilege the direct transitions (on G), except for GaSb where the two minimums of CB (at R and G) present a similar value, raising the possibility of having a higher number of indirect electronic transitions. In both cases the values of the electronic levels also depends on the electronic conformation of the atoms within the compound. The calculated value of the forbidden band gap (Eg) for CdTe is 1.79951 eV (figure 8b), meanwhile the values of Eg for the rest of the compounds span from 5.71 × 10⁻⁶ eV up to 2.63646 eV, making possible to have a material that present a similar crystallographic structure but a higher or lower Eg, depending on the application.

3.2.2. Cadmium sulfide
The EBDs of CdSe and its similar materials follow two different shapes, making necessary to plot the compounds as two different figures in order to appreciate the similarities of the EBDs; these diagrams are presented in figure 9.

Table 4. CIF codes of materials similar to CdO. Entry 0 corresponds to the CIF code that was used as basis for the data mining process.

| No. | CIF   | S   | Comp.                  | Reference                          |
|-----|-------|-----|------------------------|------------------------------------|
| 0   | 1011051 | 0   | CdO                    | Das Raumgitter des Kadmiumoxys = [27]. |
| 1   | 1011096 | 0.649 | CdO                    | Soluzioni solide, isomorfolismo e simmorfismo tra gli ossidi dei metallici bivalenti. Sistemi: CaO-CdO, CaO-MnO, CaO-CoO, CaO-NiO, CaO-MgO [28]. |
| 2   | 1011003 | 4.047 | CdO                    | The Structure of the Smoke Particles From a Cadmium Arc [29]. |
| 3   | 9006672 | 4.323 | CdO                    | Room temperature compressibilities of MnO and CdO: further examination of the role of cation type in bulk modulus systematics [30] Sample from Run #1: P = 0.00 GPa at end of experiment |
| 4   | 9006671 | 4.402 | CdO                    | Ibid. Sample from Run #1: P = 0.00 GPa at beginning of experiment |
| 5   | 9006685 | 4.494 | CdO                    | Ibid. Sample from Run #2: P = 0.00 GPa at end of experiment |
| 6   | 9006684 | 4.533 | CdO                    | Ibid. Sample from Run #2: P = 0.00 GPa at beginning of experiment |
| 7   | 9008609 | 4.533 | CdO                    | Crystal Structures. Rocksalt structure [31] |
| 8   | 9006673 | 5.232 | CdO                    | Room temperature compressibilities of MnO and CdO: further examination of the role of cation type in bulk modulus systematics [30] Sample from Run #1: P = 0.68 GPa |
| 9   | 9006686 | 6.117 | CdO                    | Ibid. Sample from Run #2: P = 1.25 GPa |
| 10  | 9006674 | 6.276 | CdO                    | Ibid. Sample from Run #1: P = 1.43 GPa |
| 11  | 9008683 | 6.633 | NbN                    | Crystal Structures. Rocksalt structure [31] |
| 12  | 9006675 | 7.231 | CdO                    | Room temperature compressibilities of MnO and CdO: further examination of the role of cation type in bulk modulus systematics [30] Sample from Run #1: P = 2.16 GPa |
| 13  | 9006687 | 7.776 | CdO                    | Ibid. Sample from Run #2: P = 2.56 GPa |
| 14  | 9006676 | 8.469 | CdO                    | Ibid. Sample from Run #1: P = 3.09 GPa |
| 15  | 9006677 | 9.244 | CdO                    | Ibid. Sample from Run #1: P = 3.73 GPa |
| 16  | 9006688 | 9.284 | CdO                    | Ibid. Sample from Run #2: P = 3.59 GPa |
| 17  | 9006678 | 10.262 | CdO                    | Ibid. Sample from Run #1: P = 4.46 GPa |
| 18  | 9006689 | 10.530 | CdO                    | Ibid. Sample from Run #2: P = 4.52 GPa |
| 19  | 9006690 | 11.391 | CdO                    | Ibid. Sample from Run #2: P = 5.30 GPa |
| 20  | 9006679 | 11.459 | CdO                    | Ibid. Sample from Run #1: P = 5.41 GPa |
| 21  | 9006691 | 12.289 | CdO                    | Ibid. Sample from Run #2: P = 6.01 GPa |
| 22  | 9006680 | 12.363 | CdO                    | Ibid. Sample from Run #1: P = 6.17 GPa |
| 23  | 9006681 | 13.134 | CdO                    | Ibid. Sample from Run #1: P = 6.73 GPa |
| 24  | 1011361 | 13.715 | ZnS                    | Die Kristallsstrukturen einiger binaerer Caride und Nitride [32]. |
| 25  | 9006682 | 13.812 | CdO                    | Room temperature compressibilities of MnO and CdO: further examination of the role of cation type in bulk modulus systematics [30] Sample from Run #1: P = 7.27 GPa |
| 26  | 9006692 | 14.383 | CdO                    | Ibid. Sample from Run #1: P = 7.80 GPa |
| 27  | 9006683 | 14.751 | CdO                    | Ibid. Sample from Run #1: P = 8.09 GPa |
| 28  | 9008778 | 15.880 | ZnS                    | Crystal Structures. Rocksalt structure [30] |
| 29  | 1539488 | 17.029 | ZnS                    | Lattice parameters and thermal expansion of Ti(CxN1-x), Zr(CxN1-x), Hf(CxN1-x) and TiN1-x from 298 to 1473 K as investigated by high-temperature x-ray diffraction [33] |
| 30  | 1538058 | 19.682 | ZnS                    | Das kubische Tantalmononitrid (B 1-Typ) und seine Mischbarkeit mit den isotypen Übergangsmetallnitriden und -carbidem [34] |
Figure 8 makes evident that the EBDs are not only dependant on the geometry of the crystalline structure of the compound, but they also depend on other factors, for example the electronic configuration (EC) of the elements that compose the material. To appreciate the difference between the compounds presented in figures 9(a) and (b), table 5 summarizes the EC of the materials similar to CdSe, separated in two columns in accordance to the shape of their EBDs. The first two columns correspond to the materials with EBDs similar to CdSe (figure 9(a)) and the columns 3 and 4 correspond to the Mn containing compounds (MnS and MnSe, figure 9(b)).

All the materials listed in table 5 present a similar EC, being the cation electrons on the last full s suborbital responsible of the chemical bond. The s electrons leave the cation and accommodate into the p suborbital of the anion, completing it to 6 electrons. The difference between the two groups is the number of electrons in the d suborbital of the cation, being 10 in the first case (Cd, Cu and Zn) and 5 in the second case (Mn). For the VB and CB analysis of the materials similar to CdSe, figure 10 presents the plots of the VB, CB and their derivatives.

In these compounds, the maximum of the VB is also present at the G wave vector, with a lower value maximum at H wave vector. The minimum of the CB is also at wave vector G, making these materials a direct
band gap compounds. The value of $E_g$ for CdSe is the smallest value of the three compounds (figure 10(b)) at 1.51301 eV.

3.2.3. Gallium arsenide

In the case of GaAs, the list of materials with similar crystallography includes CuBr and ZnSe (table 3), which are also included in the list of materials similar to CdSe (table 2). The crystallographic structure of CdSe is hexagonal, with an space group number 186: P6$_3$mc, as all its similar materials; meanwhile the crystal structure of GaAs and its similar materials lie in the cubic space group number 216: Fd$ar{3}$m. Figure 9 shows the crystal cell of CuBr (CIF: 9008864) polymorph with hexagonal structure, similar to the structure of the CdSe (CIF: 9016056) tested in this case (figure 11(a)), and the polymorph of CuBr (CIF: 1541526) with a cubic structure, similar to the GaAs (CIF: 9008845) (figure 11(b)).

Table 5. Electronic configuration of the materials similar to CdSe.

| At. | Electronic configuration | At. | Electronic configuration |
|-----|-------------------------|-----|-------------------------|
| Cd  | [Kr] 5s$^2$ 4d$^{10}$   | Mn | [Ar] 4s$^2$ 3d$^5$       |
| Se  | [Ar] 4s$^2$ 3d$^{10}$ 4p$^4$ | S  | [Ne] 3s$^2$ 3p$^4$       |
| Cu  | [Ar] 4s$^2$ 3d$^{10}$   | Mn | [Ar] 4s$^2$ 3d$^5$       |
| Br  | [Ar] 4s$^2$ 3d$^{10}$ 4p$^3$ | Se | [Ar] 4s$^2$ 3d$^{10}$ 4p$^4$ |
| Zn  | [Ar] 4s$^2$ 3d$^{10}$   |    |                         |
| Se  | [Ar] 4s$^2$ 3d$^{10}$ 4p$^4$ |    |                         |

Figure 8. Plots of the (a) valence bands and their derivatives, and (b) conduction bands and their derivatives. Compounds similar to CdTe.

Figure 9. EBDs for CdSe and 4 similar materials: (a) CdSe, CuBr and ZnSe; (b) MnS and MnSe.
As expected, there is a difference between the EBDs of two different crystalline phases of the same material. Figure 12 shows the subtle difference between the EBDs of the two polymorphs of CuBr. The hexagonal phase presents two wave vector \( G \) values where the band gap is at its minimum (1.791 eV), the first minimum is at 0.0 and the second minimum is at 0.6152, both allowing direct electronic transitions between valence and conduction band. In contrast, the cubic phase presents a unique minim band gap of 1.836 eV at a value of 0.6885 wave vector \( G \), that also present a direct electronic transition.

The shape of the EBDs of GaAs and its similar materials are comparable, as it was expected. In this case, SiGe and MnSe presented difficulties when the EBDs were tried to be calculated using the sX-LDA functional, so it was changed to GGA-PBE for this materials, keeping all the parameters the same. The results are shown in figure 13.

The shape of all EBDs were apparently similar, but AgSe and MnSe presented a deviation towards negative energies indicating a metallic behavior (Fermi level within the conduction band), so in order to compare the shapes, the value of energy of the local maximum (located at the wave vector \( G \)), was subtracted to all the values of the EBD for these two materials. Figure 14 shows the VB and CB plots of these compounds.

Most of these materials present a single direct band gap at wave vector \( G \) with exception of the SiGe, which also presents a similar in magnitude minimum of the CB at wave vector \( R \) allowing the indirect transition of electrons between bands. In this case, the value of the band gap \( E_g \) for GaAs is 1.04576 eV that is almost at the middle of the \( E_g \) values for the rest of the compounds, which span from 0.10813 eV for MnSe up to 1.86229 eV for ZnSe.
4. Conclusions

The similarity between compounds, mined from an open access database, was calculated by comparing the summation of the measured Euclidean distance between their XRD peaks obtained by simulation; and do not considering in the similarity calculations the space group classification, neither other related crystallographic theories other than the Bragg's law. In this work, results for CdTe, CdSe and GaAs are presented as example. The consequent similarity of structure-dependant properties, such as electron transport, is tested by comparing the shape of the EBDs of the most similar materials. The method used in this work to calculate the similarity between materials can be considered as a new tool for materials integration, and opens the possibility for new findings in materials science. Within all the possibilities, three different scenarios are presented in this manuscript: (1) the most similar materials present EBDs with similar shapes, such as the materials similar to CdTe. This knowledge can be used to substitute a material with another one that presents the same electronic band structure, but with a different bad gap; (2) within the list of similar materials, there are compounds with similar shaped-EBDs, and
the same list, some of the materials present a dissimilar EBD such as the case of CdSe, where MnS and MnSe exhibit a completely different EBD than the one presented by CdSe, this fact attributable to the electronic-shell configuration of the Mn atoms in the compounds. In this scenario, the list of materials can help to find a compound similar in crystalline structure but with a completely different band structure, that in a heterojunction made by the interface of CdSe and MnS or MnSe (in this particular case), will block the flow of electrons from one material to the other, creating an electron-confinement inter-phase; and (3) a compound that is present in the similarity list of two or more different materials that are classified in different space groups. These compounds present different phases that are similar to the different materials, such as the case of CuBr, which has a hexagonal phase that is similar to the hexagonal phase of CdSe and a cubic phase that is similar to GaAs. This information can be used to find materials that can function as an even EBS-transition inter-phase between compounds with different crystallographic structures and atomic compositions.

Acknowledgments

This research was supported by the Tecnológico Nacional de México (TecNM), and a scholarship granted by the National Council for Sciences and Technology (CONACyT).

Conflict of Interest

The authors declare that they have no conflict of interest.

Data availability statement

The crystallographic data required to reproduce these findings are available to download from http://www.crystallography.net/cod/

ORCID iDs

G Carbajal-Franco  
https://orcid.org/0000-0001-6090-5897

References

[1] Buchsbaum C, Höhler-Schlimm S and Rehme S 2010 Data bases, the base for data mining Struct. Bond. 134 37–58
[2] Hofmann D W M 2010 Data mining in organic crystallography Struct. Bond. 134 89–134
[3] Rajan K 2010 Data mining and inorganic crystallography Struct. Bond. 134 59–87
[4] Cheng H, Sen T Z, Jerningham R I and Klöckzowski A 2010 Data mining for protein secondary structure prediction Struct. Bond. 134 455–74
[5] Pepperel C A and Willet P 1991 Techniques for the calculation of three-dimensional structural similarity using inter-atomic distances J. CADD. 5 455–74
[6] Chisholm J A and Motherwell S 2005 COMPACK: a program for identifying crystal structure similarity using distances J. App. Cryst. 38 228–31
[7] Ortiz C, Eriksson O and Klintenberg M 2009 Data mining and accelerated electronic structure theory as tool in the search for new functional materials Comput. Mater. Sci. 44 1042–9
[8] Pulido A et al 2017 Functional materials discovery using energy–structure–function maps Nature 543 657–64
[9] Zha L et al 2016 A fingerprint based metric for measuring similarities of crystalline structures J. Chem. Phys. 144 1–9
[10] Willet P 2016 Molecular Similarity Approaches in Chemoinformatics: Early History and Literature Status Frontiers in Molecular Design and Chemical Information Science - Herman Skotheim Award Symp. 2013 6 (Jurgen Bajorath: American Chemical Society) 67–89
[11] Gražulis S, Chateigner D, Downs R T, Yokochi A T, Quirios M, Lutterotti L, Manakova E, Butkus J, Moeck P and Le Bail A 2009 Crystallography open database—an open-access collection of crystal structures J. Appl. Cryst. 42 726–9
[12] Gražulis S, Daikevičius A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya N R, Moeck P, Downs R T and Le Bail A 2012 Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration Nucl. Acids Res. 40 D420–7
[13] de R, Kronig L. and Penney W G 1931 Quantum mechanics of electrons in crystal lattices Proc. R. Soc. Lond. A 130 499–513
[14] Szmulowicz F 2008 New kronig–penney equation emphasizing the band edge conditions Eur. J. Phys. 29 307–15
[15] Eldib A M, Hassan H F and Mohamed M A 1987 Generalisation of the kronig-penney model J. Phys. C: Solid State Phys. 20 3011–9
[16] Szmulowicz F 1997 Kronig–penney model: a new solution Eur. J. Phys. 18 392–7
[17] Felix B 1929 Über die Quantenmechanik der Elektronen in Kristallgittern Physik. 52 555–600
[18] Alase A, Cobanera E, Ortiz G and Viola L 2017 Functional materials discovery using energy–structure–function maps Springer Science+Business Media, LLC
[19] Kittel C 1987 Quantum Theory of Solids 2nd ed. (New York, United States of America: John Wiley & Sons, Inc) 0–471–62412–8
[20] Neamen D A 2003 Semiconductor Physics and Devices 3rd ed. (New York, United States of America: McGraw Hill) 0–07–232107–5
[21] Guinebretière R 2007 X-ray Diffraction by polycrystalline Materials 1st ed. (London, UK: ISTE Ltd) 13–978–1–905209–21–7 (https://doi.org/10.1002/9780470612408.fmatter)
[22] Suryanarayana C and Grant Norton M 1998 X-Ray Diffraction, A Practical Approach 1st ed. (New York, United States of America: Springer Science+Business Media, LLC) 978–1–4899–0150–7 (https://doi.org/10.1007/978–1–4899–0148–4)
[23] Bollinger H J, Brown J C, Gražulis I D, Hester S, McMahon J R, Spadaccini B, Westbrook N J, D and Westrip S P 2016 Specification of the crystallographic information file format, version 2.0 J. Appl. Cryst. 49 277–84
[24] Clark S J, Robertson I, Lany S and Zunger A 2010 Intrinsic defects in ZnO calculated by screened exchange and hybrid density functionals Physical Review B 81 115311
[25] Clark S J and Robertson. J 2010 Screened exchange density functional applied to solids Physical Review B 82 085208
[26] Google Scholar http://scholar.google.com/. Consulted 27/02/2018
[27] Scherrer P 2015 VIII. Das Raumgitter des Kadmiumoxys, Z. Kristallographie - Crystalline Materials 57 186–9 Retrieved 14 Nov. 2018, from
[28] Natta G and Passerini I. 1929 Soluzioni solide, isomorfismo e simmorfismo tragli ossidi dei metallic bivalenti. Sistemi: CaO–CdO, CaO–MnO, CaO–CoO, CaO–NiO, CaO–MgO. Gazzetta Chimica Italiana 59 129–54
[29] Walmseley H P 1927 The structure of the smoke particles from a cadmium arc Proc. of the Physical Society 40
[30] Jianzhong Z 1999 Room temperature compressibilities of MnO and CdO: further examination of the role of cation type in bulk modulus systematic Phys Chem Minerals 26 644–8
[31] Wyckoff R W G 1963 Crystal Structures (New York: Interscience Publishers) 85–237
[32] Becker K and Ebert F 1925 Die Kristallstrukturen einiger binaerer Carbide und Nitride Z. Phys. 31 268–72
[33] Aigner K, Lengauer W, Rafaja D and Ettmayer P 1994 Lattice parameters and thermal expansion of TiC(N1−x), ZrC(N1−x), Hf(CN1−x) and TiN1−x from 298 to 1473 K as investigated by high-temperature X-ray diffraction J. Alloys Compd. 215 121–9
[34] Gatterer J, Dufek G, Ettmayer P and Kieffer R 1973 Das kubische Tantalmononitrid (B 1-Typ) und seine Mischbarkeit mit den isotypen Übergangsmetallnitriden und -carbiden Monatshefte für Chemie - Chemical Monthly. 106 1137–47