Data Mining for Three-Dimensional Organic Dirac Materials: Focus on Space Group 19

R. Matthias Geilhufe1, Stanislav S. Borysov1, Adrien Bouhon2 & Alexander V. Balatsky1,3,4

We combined the group theory and data mining approach within the Organic Materials Database that leads to the prediction of stable Dirac-point nodes within the electronic band structure of three-dimensional organic crystals. We find a particular space group \( P2_12_12_1 \) (\#19) that is conducive to the Dirac nodes formation. We prove that nodes are a consequence of the orthorhombic crystal structure. Within the electronic band structure, two different kinds of nodes can be distinguished: 8-fold degenerate Dirac nodes protected by the crystalline symmetry and 4-fold degenerate Dirac nodes protected by band topology. Mining the Organic Materials Database, we present band structure calculations and symmetry analysis for 6 previously synthesized organic materials. In all these materials, the Dirac nodes are well separated within the energy and located near the Fermi surface, which opens up a possibility for their direct experimental observation.

Recently, we have witnessed growing interest in the research community in the Dirac materials where the low-energy excitations behave as massless Dirac fermions. Among the most prominent examples are the two-dimensional material graphene, the surface of bulk topological insulators like PbSn4−xTe4, Dirac-line materials and Weyl semimetals like TaAs. To date, the strong focus within electronic Dirac materials lies in the inorganic crystals. The class of organic crystals remains rather unexplored and only a few organic Dirac materials are known. One prominent example is the quasi two-dimensional charge transfer salt \( \text{α-(BEDT-TTF)}_2\text{I}_3 \) which shows a tilted Dirac cone located at the Fermi energy under high pressure. At the same time, organic crystals offer a high potential for technological applications due to low production costs, elastic properties (flexible electronics) and the opportunity to build large-area devices. We therefore will focus on exploring the space of organic materials with the specific goal of identifying Dirac materials.

Since the crystal structure plays a crucial role for hosting Dirac nodes, attempts of identifying organic Dirac materials so far are mainly based on variations of already known Dirac materials. In two dimensions, this can be done by starting with the graphene structure and replacing the carbon atoms by more complex organic molecules. A similar strategy was also discussed with respect to \( \text{α-(BEDT-TTF)}_2\text{I}_3 \). To go beyond this approach, we adopt a new strategy for the search for Dirac nodes in the class of three-dimensional organic crystals. We performed a data mining study on the basis of 5217 electronic Kohn-Sham band structures calculated using Density Functional Theory (DFT) and stored within the Organic Materials Database (OMDB). We looked for isolated linear crossings, i.e., where no other bands can be found besides the crossing within the corresponding energy range. Although all considered organic materials were previously synthesized, little attention has been paid to the electronic structure for most of them so far.

To achieve stable Dirac points within the electronic structure, symmetry or topological protection needs to be present. In this connection, crystals with non-symmorphic space groups have been widely discussed where the key role is played by high-dimensional irreducible representations at the Brillouin zone boundary. As a consequence of the Wigner-Eckart theorem, the degeneracy of an electronic state is equal to the dimension of an irreducible representation. In the context of crystals, these degeneracies were discussed in great detail during the 1960s. Recently, with a reinterpretation, degenerate electronic states in crystals again attracted attention as a host to unconventional fermions as low-energy excitation. There is also a second opportunity of hosting Dirac nodes when the crystal structure is conducive to the Dirac nodes formation.
isomorphic to 222 and the coset representatives are given by

$$\sim -$$

with the difference that at least one crossing has to occur along one of the paths having the group of pure translations (green dashed circles in Fig. 1(d)) as well as along the R-related linear crossings occur at the high-symmetry point P.

Dirac nodes are a consequence of the orthorhombic crystal structure of the space group.

The Dirac nodes are a consequence of the orthorhombic crystal structure of the space group $P_{212121}$ (#19), an almost similar situation is present, with the difference that at least one crossing has to occur along one of the paths $\Gamma X$, $\Gamma Y$ or $\Gamma Z$ within the Brillouin zone. This group will be discussed below in Results.

In this paper, we report results of a combined study using abstract group theory and data mining within the Organic Materials Database (OMDB)\textsuperscript{17}. We point out the first real material examples in the class of three-dimensional organic crystals hosting isolated Dirac nodes in the electronic structure: C$_6$H$_7$ClO$_3$\textsuperscript{25}, C$_{10}$H$_{10}$Br$_2$Cl$_3$NO$_2$\textsuperscript{26}, C$_{12}$H$_{13}$NO\textsuperscript{27}, C$_{13}$H$_{12}$N$_2$O\textsuperscript{28}, C$_{10}$H$_{10}$F$_2$NO\textsuperscript{28}, and C$_{10}$H$_{12}$BrNO\textsuperscript{30}. It will be shown that the found Dirac nodes are a consequence of the orthorhombic crystal structure of the space group $P_{212121}$ (#19). Within the band structure of the materials, two different kinds of nodes can be distinguished: 8-fold degenerate Dirac nodes protected by the crystalline symmetry and 4-fold degenerate tilted Dirac nodes protected by the band topology.

### Results

#### Data mining and electronic structure calculations.

As reported in ref. 17, most three-dimensional organic crystals are insulating. However, the doping of organic materials is extensively studied, opening the opportunity of shifting the Fermi level into the valence band (p-doping) or conduction band (n-doping)\textsuperscript{11,31}. Even though it is chemically more difficult to achieve in comparison to inorganics, for example, due to purification effects of organics, it was successfully implemented within OLEDs\textsuperscript{32,33}, solar cells\textsuperscript{34,35} or thermoelectric materials\textsuperscript{36}. Therefore, we searched for isolated linear crossings in a neighborhood of 0.5 eV above the lowest unoccupied electronic state and below the highest occupied electronic state within the Kohn-Sham band structures stored in the OMDB by explicitly focusing our attention to materials with the space group $P_{212121}$ (#19). For the 6 most promising structures tabulated in Table 1, additional refined DFT calculations were performed (see Methods for more details).

In the following, we concentrate on the discussion of the material C$_6$H$_7$ClO$_3$ (the electronic and molecular structures for the other 5 materials can be found in the supplementary material). Its orthorhombic crystallographic unit cell is build up of four copies of C$_6$H$_7$ClO$_3$ molecules as shown in Fig. 1(a). The calculated band structure along several high symmetry paths within the Brillouin zone (Fig. 1(b)) is plotted in Fig. 1(d). The isolated linear crossings occur at the high-symmetry point $R$ (green dashed circles in Fig. 1(d)) as well as along the high-symmetry path $\Gamma X$ (red dotted circles in Fig. 1(d)). A picture of the linear energy dispersion within the $\bar{b}_2$-$\bar{b}_1$ plane within reciprocal space in the vicinity of the $R$-point can be seen in Fig. 1(c). The linear crossings are well separated within the energy and as a consequence the electronic density of states grows quadratically with the energy in the vicinity of the three-dimensional Dirac crossing ($m(E) \sim E^{3/2}$, where $d = 3$) as can be seen in Fig. 1(d).

However, to justify the claim of the found materials being Dirac materials, a protection of the crossings by crystalline symmetry is necessary. Indeed, we find that the nature of the crossings forming Dirac nodes in the spectra can be explained within the framework of group theory below.

#### Group theory analysis.

The space group $P_{212121}$ (#19) itself (hereinafter denoted by $\mathcal{G}$) is an infinite group having the group of pure translations $\mathcal{F}$ as an infinite, normal and Abelian subgroup. The point group of the lattice $\mathcal{G}_0$, i.e., the group of all rotational parts of the space group elements is given by 222 ($D_2$). The factor group $\mathcal{G}/\mathcal{F}$ is isomorphic to 222 and the coset representatives are given by

$$T_1 = (E, 0, 0, 0),$$

$$T_2 = (C_{2x}, 1/2, 1/2, 0),$$

$$T_3 = (C_{2y}, 0, 1/2, 1/2),$$

### Table 1. Data-mined organic Dirac materials with the space group 19. The first four materials show crossings in the valence band, the latter two have crossings in the conduction band. References to the papers where the synthesis of the materials was reported are given in the last column.

| # | OMDB ID | COD ID | sum formula | Refs |
|---|---------|--------|-------------|------|
| 1 | 3407    | 1504134 | C$_6$H$_7$ClO$_3$ | 25   |
| 2 | 4071    | 1503207 | C$_{10}$H$_{10}$Br$_2$Cl$_3$NO$_2$ | 26   |
| 3 | 3180    | 1506714 | C$_9$H$_{10}$F$_3$NO | 27   |
| 4 | 5010    | 7153350 | C$_9$H$_7$N$_2$O | 28   |
| 5 | 3617    | 1506293 | C$_{13}$H$_{12}$N$_2$O | 29   |
| 6 | 690     | 4066669 | C$_{10}$H$_{12}$BrNO | 30   |
Here, \( E \) denotes the identity element and \( C_{2x}, C_{2y} \) and \( C_{2z} \) denote two-fold rotations (rotations by 180°) about the Cartesian \( x-, y- \) and \( z- \)-axis, respectively. In general, since \( \mathcal{G} \) is an infinite group, it has infinitely many irreducible representations. However, due to the special structure of space groups they can be indexed by the combined index \( \mathbf{k} = (k_x, k_y, k_z) \), where \( \mathbf{k} \) denotes a vector in reciprocal space and \( p \) denotes an additional index running over all the allowed representations at \( \mathbf{k} \). A degeneracy of a state with energy \( E(\mathbf{k}) \) can be expected when the associated irreducible representation \( \Gamma_{\mathbf{k}} \) has dimension \( d > 1 \) or when pairs of complex conjugate representations are present. Such a degeneracy is denoted as “protected by the crystalline symmetry”. At the \( R \) point within the Brillouin zone, the physically-irreducible representation is four dimensional. Additional spin-degeneracy leads to an 8-fold degenerate linear crossing. Such crossings were recently referred to as double Dirac crossings. Within the band structure in Fig. 1(d), these crossings are highlighted by green dashed circles.

At the \( \Gamma \) point, the group of the \( \mathbf{k} \)-vector only contains the elements \( T_1 \) and \( T_2 \). Hence, according to Table 3, the bands will be classified by their transformation behavior with respect to the two-fold rotation, namely even (\( \Gamma_1 \)) or odd (\( \Gamma_2 \)). Every state at

\[
\Gamma_1 = (C_{2x}, 1/2, 0, 1/2).
\]

**Table 2.** Character table and compatibility relations of the point group 222 (\( D_2 \)).

| \( \mathbf{A} \) | \( \mathbf{B}_1 \) | \( \mathbf{B}_2 \) | \( \mathbf{B}_3 \) |
|---|---|---|---|
| \( \mathbf{E} \) | \( \mathbf{E} \) | \( \mathbf{E} \) | \( \mathbf{E} \) |
| \( C_{2x} \) | \( C_{2x} \) | \( C_{2x} \) | \( C_{2x} \) |
| \( C_{2y} \) | \( C_{2y} \) | \( C_{2y} \) | \( C_{2y} \) |

---

Figure 1. Crystal structure and electronic structure of the orthorhombic organic crystal \( \text{C}_6\text{H}_7\text{ClO}_3 \). (a) Orthorhombic unit cell and single molecule of \( \text{C}_6\text{H}_7\text{ClO}_3 \). Grey denotes carbon, red oxygen, blue hydrogen and green chlorine. (b) Brillouin zone path for band structure calculation. (c) Energy dispersion in \( b_2-b_1 \)-plane around the \( R \) point. (d) Band structure and density of states. Crossings highlighted in green circles are 8-fold degenerate Dirac points protected by crystalline symmetry and crossings in red circles are 4-fold degenerate Dirac points protected by band topology. (e) Energy dispersion along \( \Gamma X, \Gamma Y, \Gamma Z \) and irreducible representations.
Table 3. Character table of the point group 2 (C₂).

|    | T₁ | T₂ |
|----|----|----|
| E  | 1  | 1  |
| C₂ | 1  | -1 |

Table 4. Character table of the space group P2₁2₁2₁ at X, Y and Z.

|    | T₁ | T₂ | T₃ | T₄ |
|----|----|----|----|----|
| E  | 2  | 0  | 0  | 0  |

X transforms as the 2-dimensional irreducible representation E illustrated in Table 4. Moving towards X and coming from ΓX bands of character Γ₁ and Γ₂ have to merge pairwise. In general, bands can only cross (accidental crossings) when they belong to different irreducible representations. Otherwise, they will hybridize and form a spectral gap. Clearly, a similar consideration holds along ΓY and ΓZ. Taking into account all possible permutations of the four irreducible representations at Γ as well as the possible connections to X, Y and Z, it can be verified that at least one crossing can be found along one of the 3 paths. Taking into account the C₂ rotation symmetry, a second copy of the crossing along the path Γ(−X) can also be found within the Brillouin zone. Furthermore, as soon as one slightly departs from one of the 3 paths towards the interior of the Brillouin zone, none of the symmetries is kept and only the identity element T₁ is present. Thus, there is no reason to protect the crossing at any point that is not lying on one of the 3 paths and the crossing itself is a Dirac point. Hence, by including spin-degeneracy, this crossing is 4-fold degenerate.

In the present case of C₆H₇ClO₃, the topologically protected crossing can be found along the path ΓX, as can be verified from Fig. 1(d,e). In Fig. 1(d), the topologically protected crossing is highlighted by a red dotted circle. As can be seen, the four interesting bands below the Fermi level have the ordering B₃, A, B₂, B₁. Bands originating from A and B₁ and from B₂ and B₃ are merging pairwise at X for the above discussed reason. As a consequence, a crossing of bands originating from A and B₁ can be observed (see Fig. 1(e)).

Without spin-orbit coupling the system can be split into two identical spin-polarized components. One simple (spin-polarized) point-node is naturally characterized through \( \pi_2(G_2(\mathbb{C}^2)) \cong \pi(U(2)) = \mathbb{Z} \) with the topological charge given by the Chern number, \( |C| = 1 \). By C₂ symmetry, the two simple point-nodes (located here on the ΓX-axis) must carry the same charge, say \( C = -1 \). According to the Nielsen-Ninomiya theorem, the total charge over the whole Brillouin zone must be zero. We then deduce that the double point-node at R must carry a charge \( C = +2 \). This can be confirmed both numerically and algebraically.

Interestingly, C₆H₇ClO₃ is close to a Lifshitz transition between two inequivalent global band topologies, controlled by a band inversion at Γ. Here, A and B₁ are almost degenerate having an energy of \( \approx -0.39 \) eV (see Fig. 1(e)). Indeed, the energy ordering of the irreducible representations at Γ, i.e., \{B₃, B₂, B₁, A, B₁, B₂, A, B₁\}, guarantees to have two disconnected blocks of four bands (without counting the spin degree of freedom) over the whole Brillouin zone. Alternatively, if we exchange A and B₁, the global band topology guarantees to have a single block of eight bands being nontrivially connected. Supporting this, our data mining has revealed a clear candidate of a nontrivially connected eight-band subspace which is characterized by two pairs of topologically stable simple point-nodes (one pair along ΓX and one pair along ΓY) shown in the supplementary material (see Fig. 2(b)). This is the first realistic band structure reported with this new type of global band topology.

Discussion

We presented the first 6 predicted compounds for three-dimensional organic crystals hosting isolated Dirac crossings. All 6 materials were synthesized before, so we encourage direct experimental verification of the results presented here. Within the respective band structures, we identified 8-fold degenerate Dirac nodes at the R point in the Brillouin zone together with 4-fold degenerate topologically protected Dirac points along the high-symmetry path ΓX. The crossings are well separated in energy due to the flat electronic bands of organic crystals and potentially accessible via doping or gating. In comparison to inorganic materials, the spatially sparse unit cells of organic materials characterized by van der Waals bonding between large molecules leads to an electronic structure characterized by blocks of well separated flat bands. This particular property also increases chances of finding topologically protected crossings to be isolated within the energy. We expect that more organic Dirac materials will be reported elsewhere with growth of the OMDB database. Furthermore, the slope of the crossings is usually much smaller then for similar inorganic crystals leading to potential applications as slow Dirac materials.

Methods

Materials data and data mining. We analyzed 5217 Kohn-Sham band structures of the three-dimensional organic crystals stored within the Organic Materials Database (OMDB) (http://omdb.diramaterials.org). The search algorithm for isolated Dirac crossings consisted of two major steps. On the first step, illustrated in Fig. 2, the algorithm selected all materials with either zero or tiny direct energy gaps (less than 1 meV) located up to 0.5 eV above the minimum energy of the lowest conductance band or 0.5 eV below the maximum energy of the highest valence band (for metals, the reference point was the calculated Fermi energy). The purpose of this gap is to introduce numerical tolerance since the band structure calculations were performed along a discrete mesh and the algorithm could miss the crossing point if it occurs between two mesh points. The algorithm also checked...
that no other bands can be found within the corresponding energy range of the gap, which is a necessary (but not sufficient) criterion to find an isolated Dirac crossing. The first step allowed us to significantly reduce the search space as it found only 45 gaps (45 materials or 0.9% of the initial dataset) near the lowest conductance band and 92 gaps (91 materials or 1.8% of the initial dataset) near the highest valence band, making manual inspection of the search results feasible. The statistics suggest that this simple criterion performs well in filtering out irrelevant materials and can be used for the automated search for isolated Dirac crossings.

In the second step, we applied a nearest neighbor search algorithm to arrange the selected materials according to their similarity to a pattern of two crossing straight lines. With this aim, we considered the two nearest bands within a momentum window around a direct gap detected in the previous step. We empirically set the size of the momentum window to be 0.4 units wide as an expected characteristic scale of the Dirac crossing. Finally, we used the average Euclidean distance (the root mean square error) between the pattern and these two bands (within the momentum window) linearly scaled to the same bounding box to arrange the selected materials. The second step becomes important to prioritize search results when the number of materials is large.

Electronic structure calculations. Having found a subset of perspective materials, we performed refined electronic structure calculations in the framework of the density functional theory by applying a pseudopotential projector augmented-wave method, as implemented in the Vienna Ab initio Simulation Package (VASP) and the Quantum ESPRESSO code. The exchange-correlation functional was approximated by the generalized gradient approximation according to Perdew, Burke and Ernzerhof. The structural information were taken from the Crystallography Open Database (COD) and transformed into DFT input files by applying the Pymatgen package.

Within VASP, the precision flag was set to “normal” meaning that the energy cut-off is given by the maximum of the specified maxima for the cut-off energies within the POTCAR files (for example, for carbon this value is given by 400 eV). The calculations were performed spin-polarized but without spin-orbit coupling. For the integration in $k$-space, a $6 \times 6 \times 6 \Gamma$-centred mesh according to Monkhorst and Pack was chosen during the self-consistent cycle. A structural optimization was performed by allowing the ionic positions, the cell shape and the cell volume to change (ISIF = 3). Optimized structures from the VASP calculations were used. Quantum ESPRESSO was applied to estimate the associated irreducible representations of the energy levels within the band structure. The cut-off energy for the wave function was chosen to be 48 Ry and the cut-off energy for the charge density and the potentials was chosen to be 316 Ry. The calculated band structures using VASP and Quantum ESPRESSO are in perfect agreement.

References
1. Wehling, T., Black-Schaffer, A. M. & Balatsky, A. V. Dirac materials. Adv. Phys. 63, 1–76 (2014).
2. Abergel, D., Apalkov, V., Berashevich, J., Ziegler, K. & Chakraborty, T. Properties of graphene: a theoretical perspective. Adv. Phys. 59, 261–482 (2010).
3. Fu, L., Kane, C. L. & Mele, E. J. Topological insulators in three dimensions. Phys. Rev. Lett. 98, 106803 (2007).
4. Tanaka, Y. et al. Experimental realization of a topological crystalline insulator in SnTe. Nat. Phys. 8, 800–803 (2012).
5. Geilhufe, M. et al. Effect of hydrostatic pressure and uniaxial strain on the electronic structure of Pb$_x$Sn$_{1-x}$Te. Phys. Rev. B 92, 235203 (2015).
6. Geilhufe, R. M., Bouhon, A., Borysov, S. S. & Balatsky, A. V. Three-dimensional organic dirac-line materials due to nonsymorphic symmetry: A data mining approach. Phys. Rev. B 95, 041103 (2017).
Acknowledgements
The work at Los Alamos is supported by the US Department of Energy, BES E3B7. Furthermore, the work was supported by the Swedish Research Council Grant No. 638-2013-9243, the Knut and Alice Wallenberg Foundation, and the European Research Council under the European Union’s Seventh Framework Program (FP/2007-2013)/ERC Grant Agreement No. DM-321031, Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich Foundation. The authors acknowledge computational resources from the Max Planck Institute of Microstructure Physics in Halle (Germany) and the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre at Linköping University.

Author Contributions
A.V.B. designed the study. R.M.G. performed the DFT calculations as well as the group theoretical analysis. S.S.B. developed and applied the data mining algorithms. A.B. worked on the group theoretical analysis and topological aspects. All authors worked on the text and reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-07374-7

Competing Interests: The authors declare that they have no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017