Towards novel organic high-*Tc* **superconductors: Data mining using density of states similarity search**

R. Matthias Geilhufe,^{1,*} Stanislav S. Borysov,^{1,†} Dmytro Kalpakchi,^{1,2} and Alexander V. Balatsky^{1,3,‡}

¹*Nordita, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden* ²*Department of Computer Science, KTH Royal Institute of Technology,*

SE-10044 Stockholm, Sweden

³*Institute for Materials Science, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

(Received 10 September 2017; revised manuscript received 4 December 2017; published 8 February 2018)

Identifying novel functional materials with desired key properties is an important part of bridging the gap between fundamental research and technological advancement. In this context, high-throughput calculations combined with data-mining techniques highly accelerated this process in different areas of research during the past years. The strength of a data-driven approach for materials prediction lies in narrowing down the search space of thousands of materials to a subset of prospective candidates. Recently, the open-access organic materials database OMDB was released providing electronic structure data for thousands of previously synthesized three-dimensional organic crystals. Based on the OMDB, we report about the implementation of a novel density of states similarity search tool which is capable of retrieving materials with similar density of states to a reference material. The tool is based on the approximate nearest neighbor algorithm as implemented in the ANNOY library and can be applied via the OMDB web interface. The approach presented here is wide ranging and can be applied to various problems where the density of states is responsible for certain key properties of a material. As the first application, we report about materials exhibiting electronic structure similarities to the aromatic hydrocarbon p-terphenyl which was recently discussed as a potential organic high-temperature superconductor exhibiting a transition temperature in the order of 120 K under strong potassium doping. Although the mechanism driving the remarkable transition temperature remains under debate, we argue that the density of states, reflecting the electronic structure of a material, might serve as a crucial ingredient for the observed high *T_c*. To provide candidates which might exhibit comparable properties, we present 15 purely organic materials with similar features to p-terphenyl within the electronic structure, which also tend to have structural similarities with p-terphenyl such as space group symmetries, chemical composition, and molecular structure. The experimental verification of these candidates might lead to a better understanding of the underlying mechanism in case similar superconducting properties are revealed.

DOI: [10.1103/PhysRevMaterials.2.024802](https://doi.org/10.1103/PhysRevMaterials.2.024802)

I. INTRODUCTION

Since the first discovery of superconductivity within organic materials in the late 1970s [\[1\]](#page-5-0), the study of organic conductors and their transition to a superconducting state has become an active branch in condensed matter physics. In comparison to inorganic materials, most of the threedimensional organic crystals exhibit large band gaps [\[2\]](#page-5-0), which makes the identification of organic metals a highly nontrivial task. Among the first reported and widely discussed organic superconductors (e.g., TMTSF- or BEDT-TTF-based materials) are quasi one- or two-dimensional narrow band gap semiconductors which show metallic conductivity under high pressure [\[3,4\]](#page-5-0). The commonly unconventional pairing mechanisms within organic materials lead to symmetries of the superconducting gap beyond the BCS *s*-wave, such as a *d*-wave gap in κ -(BEDTTTF)₂Cu[N(CN)₂]Br [\[5\]](#page-5-0) and κ - $(BEDT-TTF)_{2}Cu(NCS)_{2}$ [\[6\]](#page-5-0), and even odd-frequency superconductivity measured in $(TMTSF)_2CIO_4 [7]$ $(TMTSF)_2CIO_4 [7]$ and theoretically predicted in κ -(BEDT-TTF)₂X [\[8\]](#page-5-0). For both the pressurized organic semiconductors as well as the pure organic metals, the superconducting transition temperature is usually in the order of 10 K [\[9\]](#page-5-0).

In contrast to the small transition temperatures observed so far, it has been recently reported that potassium-doped p-terphenyl exhibits a superconducting transition temperature of 123 K which is comparable to transition temperatures of high-temperature superconducting cuprates [\[10\]](#page-5-0). This remarkable finding might open a path towards high-temperature organic superconductivity. To extend the discussion towards other materials candidates, we report about the results of a data mining exercise for other organic crystals with similar electronic features as p-terphenyl. To date, the mechanism driving the high transition temperature remains unclear.

The data mining is performed on electronic structure data of 11 512 organic crystals stored within the open-access organic materials database (OMDB) [\[2\]](#page-5-0), freely accessible at [http://omdb.diracmaterials.org.](http://omdb.diracmaterials.org) It contains results of band structure and density of states (DOS) calculations for previ-

^{*}geilhufe@kth.se

[†] borysov@kth.se

[‡] balatsky@hotmail.com

ously synthesized three-dimensional organic crystals, where the crystal structures are mainly taken from the crystallographic open database (COD) $[11–15]$. The web interface of the OMDB contains advanced search functionality such as pattern matching for specific features within electronic band structures [\[16\]](#page-5-0) and the newly developed DOS similarity search tool presented in this paper.

Recently, a data mining approach was successfully applied, for example, for the search of stable nitride perovskites [\[17\]](#page-5-0), thermoelectric materials [\[18\]](#page-5-0), and Dirac materials [\[19–22\]](#page-5-0). In a similar manner as reported in the current paper, potential inorganic high-temperature superconductors were previously reported by Klintenberg and Eriksson [\[23\]](#page-5-0) after comparing electronic structure features of the materials stored within the electronic structure project [\[24\]](#page-5-0) to certain prototype materials.

Even though important contributions towards *ab initio* description of superconductivity were achieved [\[25–27\]](#page-5-0), a general theory for high- T_c superconductors is still under debate. However, since the normal state electronic properties together with the structure of the material will significantly influence the occurrence of a superconducting phase, a similarity search for distinct features with respect to a prototype material represents a valid approach to guide the search for other potential high-temperature superconductors. The results presented here are based on a high-throughput approach, intended to provide guidance for a experimental verification of prospective candidates. A positive experimental verification of high-temperature superconductivity within the 15 materials suggested in the following will motivate a specific microscopic description of the superconducting state.

The paper is structured as follows. At the beginning, we give details on the implementation of the novel DOS similarity search tool within the OMDB. Afterwards, we illustrate the electronic structure of the prototype material p-terphenyl which was calculated in the framework of the density functional theory (DFT). Finally, we present results of the similarity search within the OMDB and list the 15 materials which are most similar to p-terphenyl.

II. DOS SIMILARITY SEARCH TOOL

We developed a DOS similarity search tool which is freely available online at [https://omdb.diracmaterials.org/search/dos.](https://omdb.diracmaterials.org/search/dos) It allows us to retrieve materials from the OMDB with a DOS similar to the one of a provided reference material. The DOS search tool implementation follows the same line as the previously developed pattern search tool for electronic band structures within the OMDB [\[16\]](#page-5-0). The implementation uses the approximate nearest neighbor search algorithm and is based on the ANNOY library [\[28\]](#page-5-0). Subsequently, the implementation is summarized.

A. The interface

At the current stage, the reference DOS uploaded by the user has to be either in the JSON format or in the DOSCAR format, which is the standard output of the VASP code [\[29–31\]](#page-5-0). An extension to other standard formats used within major *ab initio* codes will be provided in the nearest future. Within the reference DOS, the user specifies the part of interest, which is used as the search query. The specified part also defines the size of the moving window used for the search. Furthermore, an energy search range needs to be specified, which can either comprise the entire DOS available for all stored materials or it can be fixed to a specific value, as the minimum energy of the lowest conduction band as a lower bound or the maximum energy of the highest valence band as an upper bound of the search window.

B. Search algorithm

Given the query DOS of length *w*, we employ a moving window approach, where DOS contained within the OMDB are scanned using a moving window of the same size *w* and stride *s*, which defines the number of data points the window jumps at each scanning step. Since *s* should be small with respect to *w* (which is of the order of eV), we set $s = 0.02$ eV. In the case the user specifies a fixed search bound, e.g., corresponding to the maximum energy of the highest valence band, only one search window for each material is used.

Both the query DOS and the data points contained within the moving window [Fig. $1(a)$] are represented as highdimensional vectors of size *d* [Fig. 1(b)] where a linear interpolation is applied to achieve the required number of equidistant points. As *d* should be large enough for a vector to represent DOS patterns, we define $d = w/s + 1$ (which corresponds for example to 51 points for $w = 1$ eV). The similarity between the normalized query and window vectors is measured using the cosine distance $\sqrt{2} - 2\cos\theta$, where θ is the angle between the vectors [Fig. $1(c)$]. As θ ranges from 0 (two vectors are equivalent up to a scaling factor) to π (two vectors have opposite directions), the cosine distance ranges from 0.0 to 2.0, respectively.

Due to high computational costs, a linear brute force solution, i.e., a sequential comparison of all data vectors to

FIG. 1. Calculation of similarity between two DOS. (a) the reference DOS to be used as a query and a part of DOS of the same length in the energy space for a material from the database are represented as high-dimensional vectors (b). Similarity between these two DOS is based on the angle *θ* between the corresponding vectors (c), where $\theta = 0$ corresponds to a perfect match (two DOS are equivalent up to a scaling constant). Here, a principal component projection from the high-dimensional to a two-dimensional space is used for visualization purposes.

a search query, is not suitable for online implementations, where results are expected within a short amount of time. Therefore, we use the approximate nearest neighbor search algorithm as implemented within the ANNOY library. During an indexing step, ANNOY creates a forest of binary search trees by applying the random projection method, partitioning the *d*-dimensional search space. During the search step, the trees allow us to find the closest vectors to a query vector within logarithmic time. Since this performance improvement comes in the expense of increased memory usage, only moving windows of 1 eV and 2 eV are indexed at the current stage and available for search. More details about the approximate nearest neighbor search algorithm and the ANNOY library can be found in Ref. [\[16\]](#page-5-0).

III. REFERENCE MATERIAL P-TERPHENYL

A. Crystal structure and electronic structure

The organic compound p-terphenyl or 1,4- Diphenylbenzene crystallizes within a monoclinic crystal structure obeying the space group $P2_1/a$ (14). The molecule itself has the chemical sum formula $C_{18}H_{14}$, exhibiting the following skeletal formula:

According to Ref. [\[32\]](#page-5-0), the lattice parameters are given by $a = 8.10 \text{ Å}, b = 5.61 \text{ Å}, c = 13.61 \text{ Å}, \text{and } \beta = 92.02$ °.

The electronic structure of p-terphenyl for a nonoptimized unit cell with ionic positions according to Ref. [\[32\]](#page-5-0) is illustrated in Fig. 2. The crystal itself is a large band gap insulator with a calculated band gap of 3 eV. The band structure [Fig. $2(a)$] exhibits a localized and well isolated flat band right below the Fermi level which is also seen in the density of states shown in Fig. 2(b). Such flat bands are common within the band structure of organics due to the weak hopping of electrons between the molecular orbitals. Both below and above the gap, the DOS shows localized features within an energy range of about 2 eV. After a structural optimization of the ionic positions, the band structure and density of states remain qualitatively similar. The band gap itself slightly decreases to about 2.8 eV, whereas the spectral gap between the localized band below the Fermi level to the second energetically lower branch of bands slightly increases to about 0.8 eV.

B. Computational details

The electronic structure of the reference material p-terphenyl was calculated in the framework of the density functional theory [\[33–35\]](#page-5-0) by applying a pseudopotential projector augmented-wave method [\[36–39\]](#page-6-0), as implemented in the Vienna *ab initio* simulation package (VASP) [\[29–31](#page-5-0)[,40\]](#page-6-0). During the calculations, the exchange-correlation functional was approximated by the generalized gradient approximation according to Perdew, Burke, and Ernzerhof [\[41\]](#page-6-0) (PBE). The structural information was taken from the Cambridge Structural Database (CSD) [\[32,](#page-5-0)[42\]](#page-6-0) and transferred into POSCAR

FIG. 2. Electronic structure of p-terphenyl. (a) Band structure along various high symmetry paths within the Brillouin zone, (b) Density of states including the two energy windows W1 and W2 used for the similarity search.

files using VESTA (visualization for electronic and structural analysis) [\[43\]](#page-6-0).

The energy cutoff was set to 400 eV. Due to the light atoms involved, the calculations were performed spin polarized but without spin-orbit coupling. For the integration in *k* space, a $6 \times 6 \times 6$ Γ -centered mesh according to Monkhorst and Pack [\[44\]](#page-6-0) was chosen during the self-consistent cycle. Band structure calculations were performed in two ways. First, the crystal structure was kept fixed as provided by the experiment [\[32\]](#page-5-0), which is in correspondence to the electronic structure calculations stored within the OMDB [\[2\]](#page-5-0). Second, the ionic positions were optimized by incorporating van der Waals corrections within the framework of the density dependent dispersion correction according to Steinmann and Corminboeuf [\[45\]](#page-6-0) to cross check the stability of the electronic structure with respect to slight structural changes.

IV. SEARCH RESULTS AND DISCUSSION

To identify materials with similar electronic properties as p-terphenyl, the DOS similarity search tool was applied to 11 512 materials stored within the OMDB at the moment of writing. For each material, the search was performed separately for the region below the highest occupied state as well as above the lowest unoccupied state. We used the search window size of 2 eV with the maximum energy of the highest valence band/minimum energy of the lowest conduction band as an upper/lower bound of the search window (see also Fig. 2).

OMDB-ID	COD-ID	Chemical formula	Dist.	Space group
12336	2212018	$C_8H_8Br_2$	0.641	$P2_1/c$
261	4062949	C_8H_5O	0.693	$P2_1/c$
11968	2018118	$C_9H_9Br_3$	0.714	P ₁
12231	2203730	$C_{18}H_{18}O_2$	0.716	$P\bar{1}$
2731	4020760	$C_{12}H_{16}O_3$	0.722	$P2_1/c$
4909	7150359	$C_5H_3N_{11}$	0.725	$P2_1/c$
2692	4028771	$C_{15}H_{20}N_{2}O$	0.738	P_{21}
21766	8102335	$C_4H_4O_4Pb$	0.751	$P2_1/c$
17076	4505801	$C_9H_9I_3$	0.762	P ₁
11918	2014326	$C_3H_5NS_2$	0.768	$P\bar{1}$

TABLE I. Top search results for the DOS similarity search close to the highest occupied and the lowest unoccupied electronic levels.

The search within the window below the highest occupied state identifies 16 materials showing a minimal distance of 0.641 and a maximal distance of 0.791 to the reference DOS. The similar search above the lowest unoccupied state gives five materials having a minimal distance of 0.641 and a maximal distance of 0.767. The best search results are shown in Table I and electronic structures of the three top examples are shown in Figs. 3 and [4,](#page-4-0) for the search above the highest occupied and lowest unoccupied states, respectively. The best similarity (with a distance of 0.641) was observed for 1,4-Bis(bromomethyl)benzene (OMDB-ID 12336, COD-ID 2212018).

In general, a striking correlation between the crystal structure of the identified materials with the reference material can be observed. Eight out of 15 materials exhibit the same space group and one material belongs to the same crystal class. Additionally, 14 of the 15 materials contain aromatic rings and 10 of the 15 materials contain almost planar molecules comparable to p-terphenyl.

The mechanisms guiding the high-temperature superconductivity in p-terphenyl are still unclear. The first attempts were reported by Mazziotti *et al.* [\[46\]](#page-6-0) proposing that the driving mechanism is the quantum resonance between different superconducting gaps near a Lifshitz transition.

Considering a BCS-type mechanism to induce superconductivity, the transition temperature T_c increases with increasing DOS at the Fermi level $N(E_F)$ and coupling constant *g*,

$$
T_c \sim e^{-1/|g|N(E_{\rm F})}.\tag{1}
$$

As shown before, pristine p-terphenyl is a large band gap insulator. However, according to Ref. [\[10\]](#page-5-0), a large superconducting transition temperature of 123 K was observed after doping the material with potassium. By introducing additional electrons into the system, potassium shifts the Fermi level

(a) OMDB-ID 12336, COD-ID 2212018, C8H8Br²

(b) OMDB-ID 261, COD-ID 4062949, C₈H₅O

(c) OMDB-ID 12231, COD-ID 2203730, $C_{18}H_{18}O_2$

FIG. 3. Electronic structure of the three most promising materials with similarity to the electronic structure of p-terphenyl within the valence bands.

into the conduction band resulting in an *n*-doped system. However, as large peaks due to the flat bands can be observed in both valence and conduction bands, a similar transition temperature might be observed by introducing *p* doping. In comparison to *n* doping, *p* doping is more straightforward to

(a) OMDB-ID 20042, COD-ID 2222905, C13H9FO³

(b) OMDB-ID 21466, COD-ID 7153064, C₁₆H₁₂ClN₃O

FIG. 4. Electronic structure of the three most promising materials with similarity to the electronic structure of p-terphenyl within the conduction bands.

achieve for organic materials [\[47,48\]](#page-6-0). Similarly to p-terphenyl, localized peaks in the density of states can also be observed in the suggested materials (Table [I\)](#page-3-0). Although *p* and *n* doping can affect electronic structure, we expect that couplings will depend on its averaged properties. Indeed, the electron-phonon coupling in the Eliashberg approach is integrated over the frequency range. If this coupling is responsible for the high T_c in the original compound, we expect the same mechanisms to be in play for the proposed materials.

V. CONCLUSIONS

We described the implementation of the novel density of states similarity search tool based on the approximate nearest neighbor method implemented in the ANNOY package. In connection to the present work, the search tool was released within the web interface of the OMDB, an open-access electronic structure database for previously synthesized organic materials. The approach presented in the current work is wide ranging and can be applied to various problems where the density of states is (or might be) connected to the properties of interest.

To propose a guidance in narrowing down the search space for novel organic high-temperature superconductors, we applied the novel search tool to query for organic materials with similar electronic properties to the reference material p-terphenyl. Under strong potassium doping, p-terphenyl was recently reported to exhibit a superconducting transition temperature comparable to those of known high- T_c cuprates. Resulting from the data-mining exercise, we propose 15 materials as candidates to search for high- T_c superconductivity (Table [I\)](#page-3-0) which have strong similarities within their DOS to p-terphenyl. Additionally, the presented materials also tend to show similarities within their crystal and molecular structures. Based on the similarities found, we expect similar superconducting mechanisms to take place and suggest experimental verification of the presented results.

The experimental verification of the candidates presented within this paper under similar conditions as reported for potassium doped p-terphenyl might lead to a better understanding of the prospective high-temperature superconductivity within this material. To date, the mechanism driving the remarkable transition temperature remains under debate. Furthermore, the crystal structure and chemical composition of the final compound after p-terphenyl was exposed high temperatures as well as highly reactive potassium for several hours remains to be investigated. Therefore, we believe that the field will benefit from investigating additional materials with similar properties.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Antonio Bermejo-Gómez, Philip Hofmann, Bart Olsthoorn, and M. Berk Gedik. The work at Los Alamos is supported by the US Department of Energy, BES E3B7. Furthermore, we are grateful for support from the Swedish Research Council Grant No. 638-2013-9243, the Knut and Alice Wallenberg Foundation, the European Research Council under the European Unions Seventh Framework Program (FP/2207-2013)/ERC Grant Agreement No. DM-321031, as well as the Villum foundation through the Villum Center of Excellence for Dirac Materials. The authors acknowledge computational resources from the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre at Linköping University as well as the High Performance Computing Center North.

- [1] D. Jerome, A. Mazaud, M. Ribault, and K. Bechgaard, Superconductivity in a synthetic organic conductor $(TMTSF)_2PF_6$, [J. Phys. Lett.](https://doi.org/10.1051/jphyslet:0198000410409500) **[41](https://doi.org/10.1051/jphyslet:0198000410409500)**, [95](https://doi.org/10.1051/jphyslet:0198000410409500) [\(1980\)](https://doi.org/10.1051/jphyslet:0198000410409500).
- [2] S. S. Borysov, R. M. Geilhufe, and A. V. Balatsky, Organic materials database: An open-access online database for data mining, [PLoS ONE](https://doi.org/10.1371/journal.pone.0171501) **[12](https://doi.org/10.1371/journal.pone.0171501)**, [e0171501](https://doi.org/10.1371/journal.pone.0171501) [\(2017\)](https://doi.org/10.1371/journal.pone.0171501).
- [3] H. Kino and T. Miyazaki, First-principles study of electronic structure in α -(BEDT-TTF)₂I₃ at ambient pressure and with uniaxial strain, [J. Phys. Soc. Jpn.](https://doi.org/10.1143/JPSJ.75.034704) **[75](https://doi.org/10.1143/JPSJ.75.034704)**, [034704](https://doi.org/10.1143/JPSJ.75.034704) [\(2006\)](https://doi.org/10.1143/JPSJ.75.034704).
- [4] R. Kondo, S. Kagoshima, N. Tajima, and R. Kato, Crystal and electronic structures of the quasi-two-dimensional organic conductor α -(BEDT-TTF)₂I₃ and its selenium analog α -(BEDT-TSEF) $_2$ I₃ [under hydrostatic pressure at room temperature,](https://doi.org/10.1143/JPSJ.78.114714) J. Phys. Soc. Jpn. **[78](https://doi.org/10.1143/JPSJ.78.114714)**, [114714](https://doi.org/10.1143/JPSJ.78.114714) [\(2009\)](https://doi.org/10.1143/JPSJ.78.114714).
- [5] M. Pinterić, S. Tomić, M. Prester, D. J. Drobac, O. Milat, K. Maki, D. Schweitzer, I. Heinen, and W. Strunz, Probing the order parameter of the layered organic superconductor κ −(BEDT – TTF)₂Cu[N(CN)₂]Br by ac susceptibility measurements, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.61.7033) **[61](https://doi.org/10.1103/PhysRevB.61.7033)**, [7033](https://doi.org/10.1103/PhysRevB.61.7033) [\(2000\)](https://doi.org/10.1103/PhysRevB.61.7033).
- [6] M.-S. Nam, J. A. Symington, J. Singleton, S. J. Blundell, A. Ardavan, J. A. A. J. Perenboom, M. Kurmoo, and P. Day, Angle dependence of the upper critical field in the layered organic superconductor-(BEDT-TTF)₂Cu(NCS)₂ (BEDT-TTF bis (ethylene-dithio) tetrathiafulvalene), [J. Phys.: Condens. Matter](https://doi.org/10.1088/0953-8984/11/43/102) **[11](https://doi.org/10.1088/0953-8984/11/43/102)**, [L477](https://doi.org/10.1088/0953-8984/11/43/102) [\(1999\)](https://doi.org/10.1088/0953-8984/11/43/102).
- [7] F. L. Pratt, T. Lancaster, S. J. Blundell, and C. Baines, Low-Field Superconducting Phase of (TMTSF)₂ClO₄, [Phys. Rev. Lett.](https://doi.org/10.1103/PhysRevLett.110.107005) **[110](https://doi.org/10.1103/PhysRevLett.110.107005)**, [107005](https://doi.org/10.1103/PhysRevLett.110.107005) [\(2013\)](https://doi.org/10.1103/PhysRevLett.110.107005).
- [8] M. Vojta and E. Dagotto, Indications of unconventional superconductivity in doped and undoped triangular antiferromagnets, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.59.R713) **[59](https://doi.org/10.1103/PhysRevB.59.R713)**, [R713\(R\)](https://doi.org/10.1103/PhysRevB.59.R713) [\(1999\)](https://doi.org/10.1103/PhysRevB.59.R713).
- [9] M. Lang and J. Müller, Organic superconductors, *The Physics of Superconductors* (Springer, New York, 2004), p. 453.
- [10] R.-S. Wang, Y. Gao, Z.-B. Huang, and X.-J. Chen, Superconductivity above 120 kelvin in a chain link molecule, [arXiv:1703.06641.](http://arxiv.org/abs/arXiv:1703.06641)
- [11] R. T. Downs and M. Hall-Wallace, The american mineralogist crystal structure database, [Am. Mineral.](https://doi.org/10.2138/am-2003-0409) **[88](https://doi.org/10.2138/am-2003-0409)**, [556](https://doi.org/10.2138/am-2003-0409) [\(2003\)](https://doi.org/10.2138/am-2003-0409).
- [12] S. Gražulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, Crystallography open database–an open-access collection of crystal structures, [J. Appl. Crystallogr.](https://doi.org/10.1107/S0021889809016690) **[42](https://doi.org/10.1107/S0021889809016690)**, [726](https://doi.org/10.1107/S0021889809016690) [\(2009\)](https://doi.org/10.1107/S0021889809016690).
- [13] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quiros, N. R. Serebryanaya, P. Moeck, R. T. Downs, and A. Le Bail, Crystallography open database (COD): An open-access collection of crystal structures and platform for world-wide collaboration, [Nucleic Acids Res.](https://doi.org/10.1093/nar/gkr900) **[40](https://doi.org/10.1093/nar/gkr900)**, [D420](https://doi.org/10.1093/nar/gkr900) [\(2012\)](https://doi.org/10.1093/nar/gkr900).
- [14] S. Gražulis, A. Merkys, A. Vaitkus, and M. Okulič-Kazarinas, Computing stoichiometric molecular composition from crystal structures, [J. Appl. Crystallogr.](https://doi.org/10.1107/S1600576714025904) **[48](https://doi.org/10.1107/S1600576714025904)**, [85](https://doi.org/10.1107/S1600576714025904) [\(2015\)](https://doi.org/10.1107/S1600576714025904).
- [15] A. Merkys, A. Vaitkus, J. Butkus, M. Okulič-Kazarinas, V. Kairys, and S. Gražulis, Cod:: CIF:: Parser: An error-correcting CIF parser for the perl language, [J. Appl. Crystallogr.](https://doi.org/10.1107/S1600576715022396) **[49](https://doi.org/10.1107/S1600576715022396)**, [292](https://doi.org/10.1107/S1600576715022396) [\(2016\)](https://doi.org/10.1107/S1600576715022396).
- [16] S. S. Borysov, B. Olsthoorn, M. Berk Gedik, R. Matthias Geilhufe, and A. V. Balatsky, Online search tool for graphical patterns in electronic band structures, $arXiv:1710.11611$.
- [17] R. Sarmiento-Perez, T. F. T. Cerqueira, S. Korbel, S. Botti, and M. A. L. Marques, Prediction of stable nitride perovskites, [Chem. Mater.](https://doi.org/10.1021/acs.chemmater.5b02026) **[27](https://doi.org/10.1021/acs.chemmater.5b02026)**, [5957](https://doi.org/10.1021/acs.chemmater.5b02026) [\(2015\)](https://doi.org/10.1021/acs.chemmater.5b02026).
- [18] S.Wang, Z.Wang,W. Setyawan, N.Mingo, and S. Curtarolo, Assessing the Thermoelectric Properties Of Sintered Compounds via High-Throughput Ab-Initio Calculations, [Phys. Rev. X](https://doi.org/10.1103/PhysRevX.1.021012) **[1](https://doi.org/10.1103/PhysRevX.1.021012)**, [021012](https://doi.org/10.1103/PhysRevX.1.021012) [\(2011\)](https://doi.org/10.1103/PhysRevX.1.021012).
- [19] M. Klintenberg, J. T. Haraldsen, and A. V. Balatsky, Computational search for strong topological insulators: An exercise in data mining and electronic structure, [Appl. Phys. Res.](https://doi.org/10.5539/apr.v6n4p31) **[6](https://doi.org/10.5539/apr.v6n4p31)**, [31](https://doi.org/10.5539/apr.v6n4p31) [\(2014\)](https://doi.org/10.5539/apr.v6n4p31).
- [20] R. Matthias Geilhufe, S. S. Borysov, A. Bouhon, and A. V. Balatsky, Data mining for three-dimensional organic dirac materials: Focus on space group 19, [Sci. Rep.](https://doi.org/10.1038/s41598-017-07374-7) **[7](https://doi.org/10.1038/s41598-017-07374-7)**, [7298](https://doi.org/10.1038/s41598-017-07374-7) [\(2017\)](https://doi.org/10.1038/s41598-017-07374-7).
- [21] R. Matthias Geilhufe, A. Bouhon, S. S. Borysov, and A. V. Balatsky, Three-dimensional organic dirac-line materials due to [nonsymmorphic symmetry: A data mining approach,](https://doi.org/10.1103/PhysRevB.95.041103) Phys. Rev. B **[95](https://doi.org/10.1103/PhysRevB.95.041103)**, [041103](https://doi.org/10.1103/PhysRevB.95.041103) [\(2017\)](https://doi.org/10.1103/PhysRevB.95.041103).
- [22] Q. Yan, R. u. Chen, and J. Neaton, Data-driven discovery of new dirac semimetal materials, Bull. Am. Phys. Soc. **62** (2017).
- [23] M. Klintenberg and O. Eriksson, Possible high-temperature superconductors predicted from electronic structure and datafiltering algorithms, [Comput. Mater. Sci.](https://doi.org/10.1016/j.commatsci.2012.08.038) **[67](https://doi.org/10.1016/j.commatsci.2012.08.038)**, [282](https://doi.org/10.1016/j.commatsci.2012.08.038) [\(2013\)](https://doi.org/10.1016/j.commatsci.2012.08.038).
- [24] C. Ortiz, O. Eriksson, and M. Klintenberg, Data mining and accelerated electronic structure theory as a tool in the search for new functional materials, [Comput. Mater. Sci.](https://doi.org/10.1016/j.commatsci.2008.07.016) **[44](https://doi.org/10.1016/j.commatsci.2008.07.016)**, [1042](https://doi.org/10.1016/j.commatsci.2008.07.016) [\(2009\)](https://doi.org/10.1016/j.commatsci.2008.07.016).
- [25] M. Lüders, M. A. L. Marques, A. Floris, G. Profeta, N. N. Lathiotakis, C. Franchini, A. Sanna, A. Continenza, S. Massidda, and E. K. U. Gross, Density functional theory for superconductors, Psi-k newsletter **76**, 54 (2006).
- [26] M. Lüders, M. A. L. Marques, N. N. Lathiotakis, A. Floris, G. Profeta, L. Fast, A. Continenza, S. Massidda, and E. K. U. Gross, Ab initio theory of superconductivity. I. Density functional formalism and approximate functionals, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.72.024545)**[72](https://doi.org/10.1103/PhysRevB.72.024545)**, [024545](https://doi.org/10.1103/PhysRevB.72.024545) [\(2005\)](https://doi.org/10.1103/PhysRevB.72.024545).
- [27] M. A. L. Marques, M. Lüders, N. N. Lathiotakis, G. Profeta, A. Floris, L. Fast, A. Continenza, E. K. U. Gross, and S. Massidda, Ab initio theory of superconductivity. II. Application to elemental metals, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.72.024546) **[72](https://doi.org/10.1103/PhysRevB.72.024546)**, [024546](https://doi.org/10.1103/PhysRevB.72.024546) [\(2005\)](https://doi.org/10.1103/PhysRevB.72.024546).
- [28] ANNOY (approximate nearest neighbors oh yeah), <https://github.com/spotify/annoy> (2017), accessed: 2017-09-01.
- [29] G. Kresse and J. Hafner, *Ab initio* molecular dynamics for liquid metals, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.47.558) **[47](https://doi.org/10.1103/PhysRevB.47.558)**, [558](https://doi.org/10.1103/PhysRevB.47.558) [\(1993\)](https://doi.org/10.1103/PhysRevB.47.558).
- [30] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.54.11169) **[54](https://doi.org/10.1103/PhysRevB.54.11169)**, [11169](https://doi.org/10.1103/PhysRevB.54.11169) [\(1996\)](https://doi.org/10.1103/PhysRevB.54.11169).
- [31] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, [Comput. Mater. Sci.](https://doi.org/10.1016/0927-0256(96)00008-0) **[6](https://doi.org/10.1016/0927-0256(96)00008-0)**, [15](https://doi.org/10.1016/0927-0256(96)00008-0) [\(1996\)](https://doi.org/10.1016/0927-0256(96)00008-0).
- [32] H. M. Rietveld, E. N. Maslen, and C. J. B. Clews, An X-ray and neutron diffraction refinement of the structure of *p*-terphenyl, [Acta Crystallogr. Sect. B](https://doi.org/10.1107/S0567740870003023) **[26](https://doi.org/10.1107/S0567740870003023)**, [693](https://doi.org/10.1107/S0567740870003023) [\(1970\)](https://doi.org/10.1107/S0567740870003023).
- [33] [P. Hohenberg and W. Kohn, Inhomogeneous electron gas,](https://doi.org/10.1103/PhysRev.136.B864) *Phys.* Rev. **[136](https://doi.org/10.1103/PhysRev.136.B864)**, [B864](https://doi.org/10.1103/PhysRev.136.B864) [\(1964\)](https://doi.org/10.1103/PhysRev.136.B864).
- [34] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, [Phys. Rev.](https://doi.org/10.1103/PhysRev.140.A1133) **[140](https://doi.org/10.1103/PhysRev.140.A1133)**, [A1133](https://doi.org/10.1103/PhysRev.140.A1133) [\(1965\)](https://doi.org/10.1103/PhysRev.140.A1133).
- [35] R. O. Jones, Density functional theory: Its origins, rise to prominence, and future, [Rev. Mod. Phys.](https://doi.org/10.1103/RevModPhys.87.897) **[87](https://doi.org/10.1103/RevModPhys.87.897)**, [897](https://doi.org/10.1103/RevModPhys.87.897) [\(2015\)](https://doi.org/10.1103/RevModPhys.87.897).
- [36] D. R. Hamann, M. Schlüter, and C. Chiang, Norm-Conserving Pseudopotentials, [Phys. Rev. Lett.](https://doi.org/10.1103/PhysRevLett.43.1494) **[43](https://doi.org/10.1103/PhysRevLett.43.1494)**, [1494](https://doi.org/10.1103/PhysRevLett.43.1494) [\(1979\)](https://doi.org/10.1103/PhysRevLett.43.1494).
- [37] P. E. Blöchl, Projector augmented-wave method, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.50.17953) **[50](https://doi.org/10.1103/PhysRevB.50.17953)**, [17953](https://doi.org/10.1103/PhysRevB.50.17953) [\(1994\)](https://doi.org/10.1103/PhysRevB.50.17953).
- [38] D. Vanderbilt, Soft self-consistent pseudopotentials in a generalized eigenvalue formalism, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.41.7892) **[41](https://doi.org/10.1103/PhysRevB.41.7892)**, [7892](https://doi.org/10.1103/PhysRevB.41.7892) [\(1990\)](https://doi.org/10.1103/PhysRevB.41.7892).
- [39] G. Kresse and J. Hafner, Norm-conserving and ultrasoft pseu[dopotentials for first-row and transition elements,](https://doi.org/10.1088/0953-8984/6/40/015) J. Phys.: Condens. Matter **[6](https://doi.org/10.1088/0953-8984/6/40/015)**, [8245](https://doi.org/10.1088/0953-8984/6/40/015) [\(1994\)](https://doi.org/10.1088/0953-8984/6/40/015).
- [40] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.59.1758) **[59](https://doi.org/10.1103/PhysRevB.59.1758)**, [1758](https://doi.org/10.1103/PhysRevB.59.1758) [\(1999\)](https://doi.org/10.1103/PhysRevB.59.1758).
- [41] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, [Phys. Rev. Lett.](https://doi.org/10.1103/PhysRevLett.77.3865) **[77](https://doi.org/10.1103/PhysRevLett.77.3865)**, [3865](https://doi.org/10.1103/PhysRevLett.77.3865) [\(1996\)](https://doi.org/10.1103/PhysRevLett.77.3865).
- [42] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, The Cambridge structural database, [Acta Crystallogr. Sect. B](https://doi.org/10.1107/S2052520616003954) **[72](https://doi.org/10.1107/S2052520616003954)**, [171](https://doi.org/10.1107/S2052520616003954) [\(2016\)](https://doi.org/10.1107/S2052520616003954).
- [43] K. Momma and F. Izumi, Vesta 3 for three-dimensional visu[alization of crystal, volumetric and morphology data,](https://doi.org/10.1107/S0021889811038970) J. Appl. Crystallogr. **[44](https://doi.org/10.1107/S0021889811038970)**, [1272](https://doi.org/10.1107/S0021889811038970) [\(2011\)](https://doi.org/10.1107/S0021889811038970).
- [44] H. J. Monkhorst and J. D. Pack, Special points for brillouin-zone integrations, [Phys. Rev. B](https://doi.org/10.1103/PhysRevB.13.5188) **[13](https://doi.org/10.1103/PhysRevB.13.5188)**, [5188](https://doi.org/10.1103/PhysRevB.13.5188) [\(1976\)](https://doi.org/10.1103/PhysRevB.13.5188).
- [45] S. N. Steinmann and C. Corminboeuf, Comprehensive benchmarking of a density-dependent dispersion correction, [J. Chem. Theory Comput.](https://doi.org/10.1021/ct200602x) **[7](https://doi.org/10.1021/ct200602x)**, [3567](https://doi.org/10.1021/ct200602x) [\(2011\)](https://doi.org/10.1021/ct200602x).
- [46] M. Vittoria Mazziotti, A. Valletta, G. Campi, D. Innocenti, A. Perali, and A. Bianconi, Possible fano resonance for high-t c multi-gap superconductivity in p-terphenyl doped by k at the lifshitz transition, [Europhys. Lett.](https://doi.org/10.1209/0295-5075/118/37003) **[118](https://doi.org/10.1209/0295-5075/118/37003)**, [37003](https://doi.org/10.1209/0295-5075/118/37003) [\(2017\)](https://doi.org/10.1209/0295-5075/118/37003).
- [47] B. Lüssem, M. Riede, and K. Leo, Doping of organic semiconductors, [Phys. Status Solidi A](https://doi.org/10.1002/pssa.201228310) **[210](https://doi.org/10.1002/pssa.201228310)**, [9](https://doi.org/10.1002/pssa.201228310) [\(2013\)](https://doi.org/10.1002/pssa.201228310).
- [48] M. Pfeiffer, K. Leo, X. Zhou, J. S Huang, M. Hofmann, A. Werner, and J. Blochwitz-Nimoth, Doped organic semicon[ductors: Physics and application in light emitting diodes,](https://doi.org/10.1016/j.orgel.2003.08.004) Org. Electron. **[4](https://doi.org/10.1016/j.orgel.2003.08.004)**, [89](https://doi.org/10.1016/j.orgel.2003.08.004) [\(2003\)](https://doi.org/10.1016/j.orgel.2003.08.004).