

OPEN

DATA DESCRIPTOR

High-throughput inverse design and Bayesian optimization of functionalities: spin splitting in two-dimensional compounds

Gabriel M. Nascimento^{1,3}, Elton Ogoshi^{1,3}, Adalberto Fazzio^{1,2}, Carlos Mera Acosta¹✉ & Gustavo M. Dalpian¹✉

The development of spintronic devices demands the existence of materials with some kind of spin splitting (SS). In this Data Descriptor, we build a database of *ab initio* calculated SS in 2D materials. More than that, we propose a workflow for materials design integrating an inverse design approach and a Bayesian inference optimization. We use the prediction of SS prototypes for spintronic applications as an illustrative example of the proposed workflow. The prediction process starts with the establishment of the design principles (the physical mechanism behind the target properties), that are used as filters for materials screening, and followed by density functional theory (DFT) calculations. Applying this process to the C2DB database, we identify and classify 358 2D materials according to SS type at the valence and/or conduction bands. The Bayesian optimization captures trends that are used for the rationalized design of 2D materials with the ideal conditions of band gap and SS for potential spintronics applications. Our workflow can be applied to any other material property.

Background & Summary

The design of materials properties usually involves two main steps: (i) prediction and (ii) optimization. The first step is typically carried out by direct calculation or experimental measurement for all possible combinations of atomic identities, composition, and structures (ACS)¹. Although this process has been successfully implemented for functionalities such as ferroelectricity^{2–4} and two-dimensional materials^{5–7}, this *direct approach* is usually tedious and expensive. The second step consists in the extrapolations of numerical correlations found with approaches such as machine learning^{8–12} or cluster expansion methods¹³. However, numerical relations are not necessarily transferable (i.e., limited to the set of compounds used to train the models -training set), preventing the rational design of material candidates with the optimized property out of the training set. In this work, we propose an inverse design process that integrates the rationalized prediction and optimization of materials properties as we illustrate it for the specific case of the functionality of spin splitting (SS) in 2D compounds.

Of special interest for spintronics device applications are the SS and spin polarization (SP). These functionalities are the cornerstone of spintronics, a promising, rapidly growing area that is based on the electron spin manipulation^{14–16}. In spintronic device prototypes, the interaction between the central region (a compound with SS bands) and electrodes (spin current source and detectors) can affect the SP properties¹⁶. Indeed, although the interface states can intrinsically possess both SP and SS, not all interfaces allow the desirable controllability of the SS. Alternative device configurations minimizing the interface effects can consist of a Van der Waals heterojunction with a two-dimensional (2D) material as central region^{17–19} or 2D compounds with intrinsic topological properties protecting spin currents against backscattering^{20–22}. Unlike 3D compounds, there is no database of 2D materials with SS²³. Thus, to illustrate the proposed design and optimization, we focus on 2D compounds possessing SS prototypes. These prototypes are historically classified by some kind of symmetry breaking (Fig. 1), namely: i) Rashba SS, with the electric breaking of inversion symmetry inducing SS^{24,25}, ii)

¹Center for Natural and Human Sciences, Federal University of ABC, Santo Andre, SP, Brazil. ²Brazilian Nanotechnology National Laboratory (LNNano), CNPEM, 13083–970, Campinas, São Paulo, Brazil. ³These authors contributed equally: Gabriel M. Nascimento, Elton Ogoshi. ✉e-mail: cmeraacosta@gmail.com; gustavo.dalpian@ufabc.edu.br

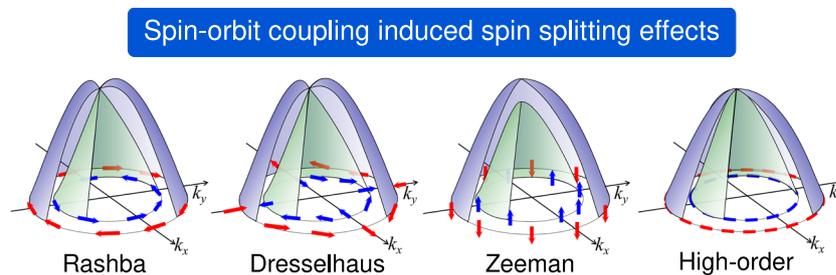


Fig. 1 Schematic illustration of the SS prototypes and corresponding SP patterns, where the *high-order* category usually corresponds to a non-trivial spin texture.

the Dresselhaus SS resulting from the non-electrical breaking of inversion symmetry²⁶, and iii) the Zeeman SS induced by the breaking of the time-reversal symmetry^{27,28}. Here, we focus only on intrinsically induced SS, i.e., the effects that are originated from the material's intrinsic electric dipoles and spin-orbit coupling interactions, and consider a fourth additional SS prototype, namely *High-Order*, as a category of observed SS effects that do not exactly follow the phenomenological SS models (see Methods). These SS prototypes are also characterized by different spin textures, i.e., the SP pattern in the reciprocal space (Fig. 1)²⁹.

We propose and implement a two-step automatic design of material properties: i) target property prediction based on the inverse design approach and ii) optimization of the target property based on Bayesian inference. Unlike the direct approach and machine learning, the inverse design approach requires first to define the design principles (DPs) (a physical mechanism) enabling the existence of the target properties, i.e., each of the SS prototypes above defined. We then develop a workflow that integrates a set of programs and scripts to automatically apply the design principles as filters, perform density functional band structure calculations including the spin-orbit coupling, and analyze the physical properties characterizing spin-polarized bands (e.g., energy position with respect to the Fermi energy, SP values, and SS magnitude) for 2D materials. We integrate this workflow to the optimization process based on Bayesian inference, in which each desirable optimization condition is maximized. Applying this approach to 2D compounds, we calculate 436 2D materials and identify more than 1200 SS at the valence and/or conduction bands, that were then classified as Rashba, Dresselhaus, Zeeman, and High-Order SS prototypes^{30,31}. The Bayesian analysis allowed us to find chemical and structural trends of three optimized properties relevant to spintronics: the Rashba parameter, the Zeeman spin splitting, and the band gap.

Methods

We propose a workflow based on the *inverse design* approach that automatically filters, selects, and designs compounds with specific functionalities³², i.e., a controllable material property with potential device applications. Remarkably, in addition to the simple material prediction, commonly performed for diverse functionalities in quantum materials^{32–34}, we integrate the proposed workflow to the automatic optimization of the target functionality based on Bayesian statistics.

Unlike numerical correlation methods, such as machine learning and cluster expansion, the inverse design approach aims first to establish the physical mechanism (design principles) enabling the target property or functionality. The second step is to seek compounds using the design principles as filters or conditions for a rationalized design. Finally, the third step is the theoretical or experimental characterization of the magnitude of the target property for the predicted/selected compounds. In principle, unlike numerical predictive methods, in the inverse design method there is no selection of false positives, i.e., compounds that are selected but do not have the target property. The main idea of the inverse design approach is to perform an optimized materials prediction process that, compared to the “direct approach”, reduces the number of necessary experiments or calculations for the target property verification. The direct approach involves the direct computation or measure of the property for all possible combinations of atomic identities, composition, and structures (ACS) attributes.

In this section, we describe the above-noted steps for the inverse design process in the context of spin splitting (SS) prototypes in two-dimensional (2D) materials, namely: A. Definition of design principles, B. Implementing design principles as filters, and C. Target property magnitude characterization, including the SS Identification Algorithm (SSIA). Additionally, we also describe the automatic Bayesian optimization of the target properties as well as its integration with the inverse design process.

Automatic inverse design approach. *A. Design principles for spin splitting prototypes.* The inverse design process starts with the definition of the physical mechanism enabling the target property. The application of the inverse design process for more than one target property, e.g., different spin splitting prototypes, requires to divide them into: 1. design principles (DPs) enabling all spin splitting types, i.e., common DPs, and 2. unique enabling DPs for each spin splitting type, as indicated in Fig. 2. This design principles division is convenient for computational implementations and was also applied for the inverse design of co-functionalities by Acosta *et al.*³². It is important to note that the understanding of the mechanism behind the existence or absence of SS has evolved over time, from an orthodox description based on the crystal point group symmetry to a description based on the atomic site symmetry³⁵. The latter not only reproduces the well-known description of the SS existence in compounds with non-centrosymmetric crystal point groups (e.g., Rashba and Dresselhaus SS), but also shows that compounds with local polar atomic site symmetry can have splitted split bands formed by orbital spatially

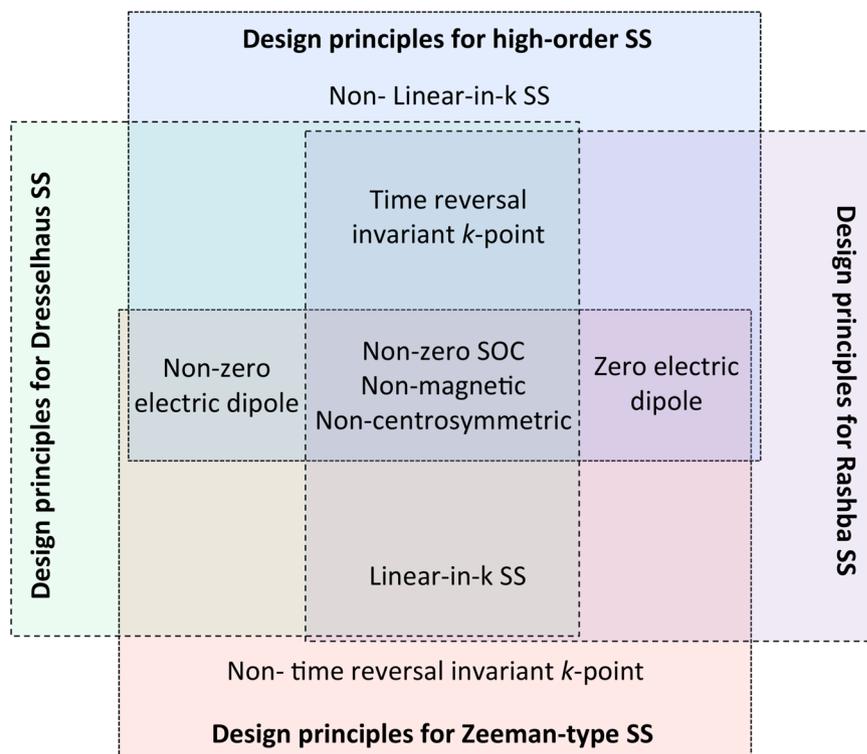


Fig. 2 Schematic illustration of the enabling design principles for SS prototypes.

localized at different material sectors³⁵. If local dipoles at different sectors cancel each other (i.e., centrosymmetric compounds), the band structure is spin degenerated. Despite the remarkable predictions and potential application of the hidden SS^{35–37}, in this work, we focus on materials that explicitly have energetically discriminating spin bands in its electronic structure. In this section, we then describe the common and unique enabling design principles for the SS prototypes classification here employed.

- Common enabling DPs for all SS types:** The common design principles for the SS prototypes are represented in the central region of Fig. 2 (i.e., the intersection region). These include: a) non-centrosymmetry crystal symmetry, which is necessary for the SOC-induced breaking of spin degeneracy, b) non-vanishing SOC and c) non-magnetic materials, which narrow the possible materials' degrees of freedom in the analysis. As it is very well established, the breaking of both the inversion symmetry and time-reversal symmetry lifts the spin degeneracy. Our work focuses on the breaking of the inversion symmetry (i.e., non-centrosymmetric compounds - DP a). The design principles a-c are used as filters to screen materials from the original data source. Additionally, we also restricted our initial data set to compounds with a finite electronic band gap (i.e., larger than 1 meV). This is of special interest for applications in spin transistor devices.
- Unique enabling DPs for SS types:**
 - DPs for Rashba spin splitting:** Emmanuel Rashba determined that the breaking of the inversion symmetry induced by external electric fields in thin films leads to a shift in the momentum space of bands with opposite spin²⁴. This linear-in- k shift near k -points preserving the time-reversal symmetry is also characterized by a splitting in the energy of the spin bands. Recently, this Rashba SS was confirmed to also exist in compounds with inversion symmetry breaking induced by the intrinsic electric dipole in 3D non-centrosymmetric bulk compounds^{38–40}. This discovery motivated the inverse design of the Rashba SS in 3D bulk compounds²³. Although intrinsic electric dipoles can also lead to the Rashba SS in 2D materials, even without external electric fields³⁵, there is not a list of 2D Rashba compounds. To organize the design and materials' screening of this type of compounds, we first defined the enabling design principles. In addition to the common DPs previously defined, the Rashba SS in 2D compounds also requires a non-zero electric dipole (allowed by at least one polar atomic site and local dipoles that add up to non-zero, as described by Zhang *et al.*³⁵). The SS type depends on the wave vector point group²⁹ and hence, the DPs also include the k -point preserving the time-reversal symmetry, as well as a linear-in- k shift. The unique DPs for the Rashba SS (i.e., non-zero electric dipoles, time-reversal symmetry, and linear-in- k SS) are illustrated in the purple quadrant of Fig. 2. We note that the *Rashba* SS prototype classification employed in this work is equivalent to the R-1 classification proposed by Zhang *et al.*³⁵, where the lack of global bulk inversion symmetry leads to non-degenerate, i.e., explicit spin split, electronic bands.
 - DPs for Dresselhaus spin splitting:** Based on the $k \cdot p$ model, Gene Dresselhaus found that the SOC Hamiltonian describing states near a time-reversal symmetry k -point in a non-polar non-centrosymmetric material

(i.e., with zero electric dipole) is given by $H(k) = \beta[(k_y^2 - k_z^2)k_x J_x + (k_z^2 - k_x^2)k_y J_y + (k_x^2 - k_y^2)k_z J_z]$, where J_i are the components of the total angular momentum operator²⁶. In 2D compounds, the Hamiltonian is $H_{2D}(k) = \beta_{2D}(\sigma_x k_x + \sigma_y k_y)$, where σ_i are the Pauli matrices. This SOC term leads to a linear-in- k SS. A list of 3D compounds having Dresselhaus SS has been recently reported²³. Based on the enabling design principles, we extend the search of Dresselhaus SS to 2D materials. Besides the common DPs, as shown in Fig. 2, enabling DPs also include zero electric dipoles (i.e., non-polar site symmetries or polar site symmetries that add up to zero, as established in ref. 35), time-reversal symmetry, and linear-in- k SS. Unlike the Rashba SS allowed by non-centrosymmetric polar point groups, the Dresselhaus SS can be found in non-centrosymmetric non-polar point groups, but it is not limited to these point groups (e.g., polar site symmetry can accidentally add up to zero). Similar to the previous case, this classification is equivalent to the D-1 effect in ref. 35

- **DPs for Zeeman-type spin splitting:** The existence of the electron spin was first elucidated by the magnetic discrimination of states⁴¹. The breaking of time-reversal symmetry via a magnetic field shifts in energy the Bloch bands with different spins, as in ferromagnetic compounds. More than one hundred years after Zeeman's discovery, it was established that even non-magnetic compounds can have a Zeeman-type SS at k -points breaking the time-reversal symmetry^{27,28}. The SOC acts as an effective magnetic field that locally breaks the time-reversal symmetry but globally preserves it. The Zeeman-type SS can exist in polar and non-polar compounds; however, it satisfies the linear-in- k SOC Hamiltonians describing the Rashba and Dresselhaus SSs. In this Zeeman-type SS, unlike the Dresselhaus and Rashba SS (Fig. 1), there are no degenerated bands. Summarizing, besides the common DPs, the Zeeman-type SS also require the local breaking of the time-reversal symmetry in the reciprocal space (i.e., wavevector point group symmetry without TR-symmetry), as indicated in Fig. 2.
- **DPs for high-order spin splitting:** We define the high-order prototype as composed of SS whose band dispersion does not exactly follow the phenomenological linear-in- k Rashba and Dresselhaus Hamiltonians but are degenerated at time-reversal symmetry invariant k -points. The design principles that apply here are polar and non-polar non-centrosymmetric crystal point groups, SS near k -points preserving the time-reversal symmetry, and non-linear SS. High-order SS is also induced by the odd terms in k^n appearing in the SOC Hamiltonian (with $n > 1$). We acknowledge that this category is broadly defined and represents a way to report materials and spin splittings that do not follow the Rashba/Dresselhaus characteristic band dispersions but still present SS that can be investigated in future theoretical works.

Note that a given compound can present more than one SS prototype at different k -points, as discussed by Acosta *et al.*²⁹, and thus the SS characterization should identify all SS types in a given compound. As both the Rashba and Dresselhaus SS prototypes share the common characteristic of having linear-in- k Hamiltonians, it will be useful to label them as a *Linear SS* (LSS) when implementing the SS identification algorithm, as described below, while not losing the symmetry criteria that differentiate them. Zeeman and High-order SS also have acronyms in the context of the algorithm: ZSS and HOSS, respectively). Additionally, the hidden SS can also constitute some of the potentially most promising classes of materials, which will be the focus of future works limited to only hidden SP in 3D and 2D materials.

B. Materials screening based on common design principles for spin-splitting types. In this section, we illustrate how design principles are translated into rules for compound selection, as well as the algorithms and computational implementations to perform the inverse design of SS prototypes in 2D compounds. The entire process detailed in this section is represented in the diagram shown in Fig. 3.

The starting point is the Computational 2D Materials Database (2020 version)⁴², containing a total of 3814 unique entries generated by elemental substitution based on known 2D structural prototypes. Even though the cohesion of a given material in a two-dimensional structure after relaxation is already checked throughout the workflow performed in the database, we re-scan all the entries with a modified rank determination algorithm⁴³, implemented with the `analysis.dimensionality` module from Pymatgen⁴⁴. Although not being strictly mandatory, this step is intended to unify the criteria of the symmetry and dimensionality analysis if similar procedures are applied to other two-dimensional databases that implement different strategies for 2D materials discovery. At this point, 3708 materials classified as 2D by the algorithm proceed in the screening workflow. For the list of 3708 2D compounds, the enabling DPs (See section A of Methods) are applied as screening filters based on the materials information provided by the C2DB database, distilling the materials landscape to be analyzed by first-principles calculations. Before implementing the common DPs, we use the DFT calculated GGA-PBE band gap values in the C2DB database⁴² to remove the entries with near-zero band gap (i.e., $E_g < 10^{-3}$ eV). This initial filtering results on 1020 non-zero bandgap materials. We note that throughout the DFT calculations workflow (see next subsection of Methods), the compound AuTe (identified in the C2DB with the uid: Au2Te2-aaafa8f843d5b) has a bandgap of 0.04 eV, but it is identified as metallic according to the GGA-PBE calculations implemented in the workflow of this work. This compound is then excluded from the SS identification analysis. The number of entries in this screening stage is then corrected to 1019.

For the selection of non-centrosymmetric materials (common DP-a - see intersection region in Fig. 2), the structure space group number is determined via the `symmetry.analyzer` module from Pymatgen⁴⁴. Here, a strict tolerance parameter for the space group classification is employed (`symprec=0.001`), to determine the structure's symmetry. Such tight parameter, although being too strict for general purposes and possibly enabling the selection of false positives SS materials, is intended to select the largest group of materials that can potentially display the aforementioned SS effects in its band edges. We note, retrospectively, that if the tolerance parameter employed at this stage was the default value (`symprec=0.01`), 21 materials that possess some

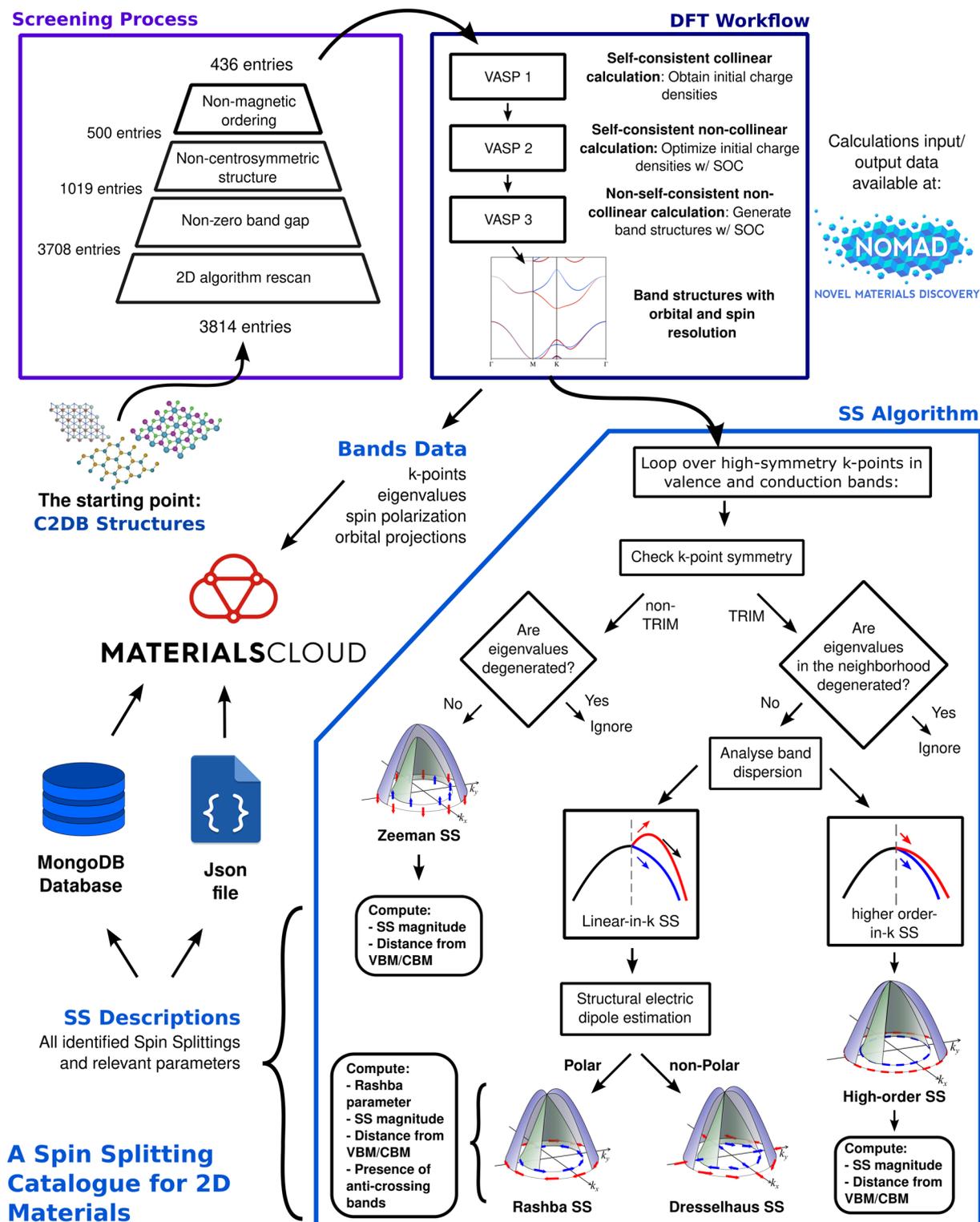


Fig. 3 Representation of the workflow outlined in section B of Methods. It starts with the 3814 materials from the C2DB (2020 version) and proceeds to the screening process (purple rectangle), sequence of DFT calculations (dark blue rectangle), and SS identification (blue pentagon), where its underlying algorithm is schematically represented. The final result is a database of SS in 2D materials, identified in valence and/or conduction bands, available in multiple formats and data types (see *Data Records* section).

type of SS in their band edges (determined by the next stages of this work) would not be identified. The number of materials that proceed at this stage of the screening process is then 500.

Filter	Entries
2D algorithm ⁴³	3708
PBE Gap > 0	1019
Non-centrosymmetric structure	500
Non-magnetic	436

Table 1. Number of entries in each screening step based on the enabling design principles.

k-point mesh	G-centered, 11 × 11 × 1 grid
Energy difference for electronic convergence	1.0 × 10 ⁻⁶ eV
Smearing	tetrahedron method with Blochl corrections (ISMear = -5)

Table 2. Relevant parameters for the DFT calculations i) and ii) in the workflow performed for each entry.

To guarantee the global preservation of time-reversal symmetry (common DP-b - see intersection region in Fig. 2), magnetic compounds are eliminated from the previous list of 500 non-centrosymmetric compounds. The time-reversal symmetry breaking can induce SS that are not necessarily induced by the spin-orbit coupling (e.g., the Zeeman effect⁴⁵ and the anti-ferromagnetic-induced SS^{46,47}). Naturally, our approach can be extended to magnetic compounds by considering a complete analysis of magnetic point group symmetry. In this case, the SOC is not necessarily a common design principle. Here, we use the magnetic ground state reported by the C2DB database, in which the anti-ferromagnetic configuration is restricted to duplicated unit cells. The detailed study of the magnetic ground states usually requires the DFT total energy study of multiple spin configurations and larger supercells⁴⁸. Alternatively, machine learning algorithms can be used to predict the magnetic ground states, as we demonstrated in ref. ⁴⁹. This filtering process based on common DPs then results in 436 non-magnetic non-centrosymmetric semiconductors. Table 1 summarizes the number of entries that proceed at each step of the screening process.

Although a large atomic SOC is usually desired, an extra filter for compounds with high atomic numbers has not been applied, so a larger set of atomic species can be analyzed in the optimization step. At this point, in principle, all the selected compounds can potentially have at least one SS type. Thus, the final materials selection according to the unique design principles is accompanied by the characterization of the magnitude of the spin splitting. We then proceed with the selected 436 compounds for the high-throughput calculations required to evaluate the unique enabling design principles (Fig. 2).

C. Target property verification: High-throughput calculations. To classify the 436 2D compounds potentially having SS according to their SS prototype, the implementation of the unique design principles is required. This requires the calculation of all band structures. Thus, in the specific case in which the target property is the SS type, it is convenient to design an algorithm that not only filters compounds using the unique design principles according to the band shape, but also extracts the magnitude of the specific SS type. A workflow of *ab-initio* calculations is then developed to generate band structures with a spin-polarization resolution that are analyzed in the SS identification algorithm. DFT + SOC calculations are performed using the Vienna Ab initio Simulation Package (VASP) with the projector-augmented wave (PAW) method^{50,51} and GGA-PBE⁵² parameterization for the exchange-correlation functional. The workflow is prepared and managed using the Atomic Simulation Environment (ASE) package^{53,54}, which is integrated with VASP.

For each entry (i.e., each of the 436 selected 2D compounds), a set of three calculations is performed to i) determine the ground state charge density in a self-consistent scheme, ii) optimize the charge density in a non-collinear scf calculation, and iii) perform a band structure, non-collinear calculation (Fig. 3). No additional relaxation procedure is performed since the available structures in the C2DB already correspond to an energy minimum. The same exchange-correlation functional is employed throughout the calculations in both the database and this work. We have verified this for a set of aleatory selected compounds at the start of the workflow. For all calculations, the cutoff energy for the plane-wave expansion was set to 520 eV. The choice of potentials employed in the workflow follows the Materials Project recommended PAW setup⁵⁵, implemented through ASE⁵⁴. Relevant parameters for the calculations i) and ii) are presented in Table 2. For the sampling of Brillouin Zone according to the specific *k*-paths in calculation iii), a density parameter of 80 *k*-points per Å⁻¹ was set for all entries.

SS Identification Algorithm (SSIA). Based on the data generated from the band structures with orbital and spin resolution, an algorithm is designed to analyze the energy dispersion at the valence and conduction bands leading to the identification of the SS type (i.e., Rashba, Dresselhaus, Zeeman, and High-order prototypes). In this process, the unique DPs are applied as criteria used by the algorithm to evaluate and identify Ss according to i) symmetry of the *k*-point where the SS occurs, ii) band dispersion in the region of the SS, and iii) estimation of structure electric dipole. The code is built upon Pymatgen and ASE functionalities, and its underlying algorithm is detailed in this section. The algorithm of SS identification is solely based on the band structure data, obtained by the DFT calculations. It consists of looping the analysis over the eigenvalues of the valence

and conduction bands and the immediate next bands (which would represent the spin degenerated copy in a non-polarized scheme) on the high-symmetry k -points, that are labeled as time-reversal invariant momentum (TRIM) or non-TRIM k -points. For the 5 different Brillouin zones in 2 dimensions, the TRIM k -points can be directly determined and passed as a list to the SS algorithm (see section *K-paths along high-symmetry lines* in the Supplementary file 1 - Band Structures material). Any high-symmetry k -point which is not in this list is treated as non-TRIM by the code, in the sense that it checks the possibility of having non-degenerate bands at this k -point (Zeeman SS). When analyzing the proximity of those points in each direction, three possibilities may arise:

1. The pair of bands are energy degenerated on the high-symmetry k -point and in its vicinity: No SS is present;
2. The pair of bands are not energy degenerated in a non-TRIM k -point: There is a SS gap, which is measured, and the SS is classified as belonging to the Zeeman prototype;
3. The pair of bands is energy degenerated in the high-symmetry k -point, but not in its vicinity: A SS happens on the k -path segment between high-symmetry k -points. The SS prototype classification will then follow the characteristics of the dispersion of the bands: if both SS bands follow the same direction (the sign of the first derivative is the same in the region next to the high-symmetry k -point) the SS is immediately classified in the high-order prototype, as it do not completely obey the phenomenological Rashba/Dresselhaus linear-in- k Hamiltonian. If the bands follow opposite directions, a Rashba/Dresselhaus-type SS is observed, and the classification of the SS in one of those groups will follow the result from the structure based estimation of electric dipole. Crystal structures displaying zero (non-zero) net electric dipole are classified in the Dresselhaus (Rashba) prototype. For the last two cases, the Rashba/Dresselhaus coefficient $\alpha_{R,D}$ is computed according to Eq. 1:

$$\alpha_{R,D} = 2 \frac{\Delta E_{R,D}}{k_{R,D}}. \quad (1)$$

where $\Delta E_{R,D}$, $k_{R,D}$, stand for the energy difference of the spin splitted bands (SS magnitude) and the k -path interval from the SS to its correspondent high-symmetry k -point in reciprocal space, respectively.

Figure 3 summarizes the SS classification heuristics adopted in the algorithm. For all SS prototypes, the energy difference between the SS bands and between the SS and the VBM/CBM is computed. The result is an extensive list of SS identified at the valence and/or conduction bands for 358 materials, presented in the tables in the Supplementary files 2–5 for the Rashba, Dresselhaus, Zeeman, and high-order SSs, respectively, whose distribution is also represented in Fig. 4. We note that single materials can have multiple, non-excluding SS prototypes at the valence/conduction bands, which are reported in the tables accordingly.

Inverse optimization process. Besides the enabling (common and unique) design principles, there are other conditions that are not required for the existence of the target property, but are important for its optimization towards specific device applications. Unlike the enabling design principles, the optimizing design principles are not necessarily physical mechanisms, but characteristics related to the chemical compositions and structure (that do not affect the existence of the target property). For SS and SP prototypes, for instance, one would like to have a compound with (i) a large enough band gap to allow gate-controllable position of the Fermi energy, and (ii) effective masses set to increase the charge carrier mobility to provide control over the performance of semiconductor devices. Additionally, it is also desirable to optimize some other properties characterizing SS prototypes and the efficiency of spintronic devices, namely: (iii) large SS (i.e., larger than 25 meV), (iv) large SP coefficient (i.e., larger than 1.3 meV), and (v) position of the SS with respect to the Fermi energy.

In an ideal scenario, properties i–v should have optimized values. However, these optimum values could be physically contradictory according to the usual trends defined by the chemical composition. For instance, while large SS are usually expected in compounds formed by atoms with large atomic numbers, large bandgaps tend to be found in compounds formed by atoms with small atomic numbers. This suggests that large bandgaps and large SSs are in some sense contradictory. The question arising from these apparent contradictions is: How to find the optimal candidate? This problem is also evident in other areas. For instance, the apparent contradiction between a high thermal insulation and electrical conductivity, which is desirable for thermoelectric applications. Here, we illustrate a strategy to address this problem for the specific case of compounds with SS prototypes for spintronic device applications, as shown in section Technical Validation. As we explain below, the inverse optimization process is based on Bayesian statistics using the materials' composition and crystal structure.

Bayesian statistics for materials property optimization. Bayesian Inference (BI) is a statistical method of inference based on Bayes' Theorem. The Bayes' Theorem specifies how one should update the probabilities when new information is given. Starting from a hypothesis H , such as a material belonging to class h , and a property of this material, one can define $P(H|A)$, i.e., the probability of an material belonging to h given that it has the property A . This is called posterior probability and it is given by the Bayes Theorem:

$$P(H|A) = \frac{P(A|H)}{P(A)} P(H) \quad (2)$$

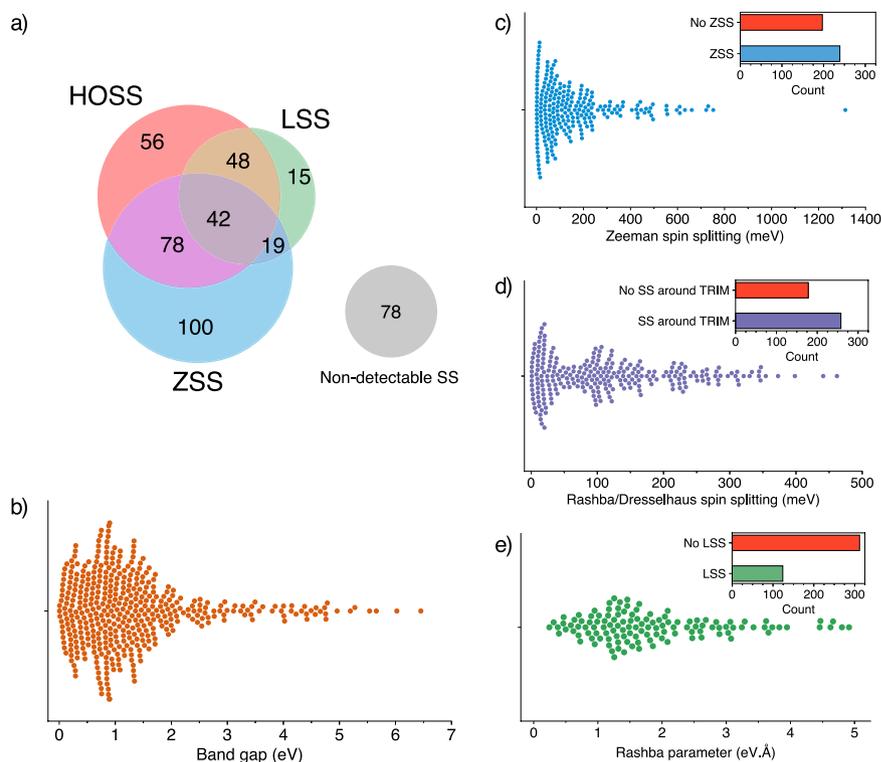


Fig. 4 (a) Distribution of the 436 materials according to different SS prototypes presented at their valence and/or conduction bands. The gray set represents the materials in which the SS Identification Algorithm (SSIA) could not identify any spin splitting; (b) Swarm plot of band gap values for the entire set of materials; (c) Swarm plot of the Zeeman spin splitting (ZSS) values. Inset shows the proportion of ZSS materials (239/436); (d) Swarm plot of spin splittings around TRIM k-points for the LSS and HOSS materials. Inset shows the proportion in this set of materials (216/436); (e) Swarm plot of the Rashba parameter presented by LSS materials. Inset shows the proportion of materials presenting LSS (124/436); As a material can have multiple instances of SS in their valence or conduction bands, only the SS with the highest value is plotted on the swarm plots.

where the three right-side terms are:

- $P(A)$: Probability of a material having the A property;
- $P(H)$: The prior probability. This is the probability of H without knowing anything regarding the material. It is simply given by the ratio of materials belonging to h ;
- $P(A|H)$: The likelihood function. The probability of a material with property A given that it belongs to h .

Equation 2 shows how to update the probability of given new information regarding A . Observing how the knowledge of a material's property A updates the prior probability allows to infer about correlations between A and h . Therefore, an increasing probability update shows a strong correlation between A and h .

The likelihood term $P(A|H)$ is a calculated probability given by feature/property A and its distribution in materials within h . Properties can be continuous, discrete, or categorical, which implies that different probability distributions should be used accordingly to A 's nature.

For the case of A as a categorical feature, such as for structural cluster, we are interested in the posterior probabilities given that the material belongs to a structural cluster t from the n possible structural clusters, i.e. we want to calculate $P(H|A = t)$. One should use the Categorical distribution for the likelihood term $P(A = t|H)$, given by:

$$P(A = t|H) = \frac{N_{th} + \alpha}{N_h + n\alpha}, \quad (3)$$

where N_{th} is the number of times the materials from structural cluster appears t in the class h , N_h is the total number of materials belonging to the class h , and α is a regularization term. The regularization term prevents $P(A = t|H)$ from having absolute values (0 or 100%) when the category t has all materials outside or within the class h .

For the case of A as the presence of an element in the material's composition, A can assume False or True values. We wish to calculate $P(H|X \text{ element in composition} = \text{True})$ due to the boolean and non-exclusive nature of this feature (a material can have more than one element in its chemical composition). One then should use the Bernoulli distribution:

$$P(A_X|H) = \begin{cases} p_X & \text{if } A_X = \textit{True} \\ 1 - p_X & \text{if } A_X = \textit{False} \end{cases}, \quad (4)$$

where A_X is the event of having element X in its composition and p_X is similar to the probability given by Eq. 3:

$$p_X = \frac{N_{X=\textit{True},h} + \alpha}{N_h + \alpha}, \quad (5)$$

where $N_{X=\textit{True},h}$ is the number of materials with X in the composition that belongs to class h , and N_h the total number of materials belonging to class h . Once again, α acts as a regularization term.

The Categorical and Bernoulli distributions were used to infer the effect of crystal structure and composition on the calculated spin-splitting properties, respectively, as presented in the Technical Validation Section. We used the scikit-learn⁵⁶ implementations, with $\alpha = 1$ as regularization.

Data Records

As each material may present multiple SSs in its band structure (in the valence and/or conduction band) classified into different SS types, the generated data shows a highly unstructured nature. For this reason, we opted to provide the complete data in multiple formats, suited for the different types of data structure and use cases.

Firstly, an overview of this work's main findings is contained in tables available in the Supplementary files 2–5 and in an Excel-compatible.csv file. There, the reader may find the full list of SSs identified in this work, separated by the SS prototypes (Rashba, Dresselhaus, Zeeman, and High-order). (See *Design principles for spin splitting prototypes* section), as well as the materials general information (e.g. id, symmetry, band gap, energy above convex hull) and SS related information (e.g. SS magnitude and localization in the band structure). As each line in the table represents a single SS, one compound may be repeated several times in each SS category. For a visual representation of the data, the Supplementary file 1 contains the rendered image of the structure and band structure with spin polarization resolution for all materials calculated in this work. The reader is then able to correspond the SS reported in the tables with its localization in the plotted band structure.

These images are also available in the Materials Cloud⁵⁷ repository for this work⁵⁸, subdivided into folders for each compound. There, the reader also finds a .cif file containing the material's crystal structure information, that determines the choice of the unit cell and origin of the coordinate system of the structure representations employed in the calculations.

Regarding the raw results of the calculations, these are available in two sources. For data provenance, we store the main input and output files of the VASP DFT band structure calculations in a NOMAD repository⁵⁹. In this manner, the reader can find the necessary raw files that would reproduce the calculations in this work. For accessing the band structure results, on the other hand, we alternatively provide a DFT-code-agnostic format which mainly relies on Pymatgen⁴⁴ and ASE⁵⁴ python objects. These are available in a pickle file inside each compound's folder in the Materials Cloud repository⁵⁸. There, the reader may get the full list of k-point coordinates, eigenvalues, orbital projections, spin polarization, and other relevant data of all the band structure calculations in this work. The README.md file available in the repository contains detailed instructions to open the data using Python.

Regarding the main results of this work, which represent the full post-processed data with the SS identification and description generated by the developed algorithm for all screened materials, these are available in a single dataset as a JSON file in the Materials Cloud repository. Alternatively, we also made it available as a binary export dump file that can be imported directly to a MongoDB database (www.mongodb.com). These two files contain the same information and may suit different preferences and use cases. Detailed instructions for accessing both data files are also available in the README.md file.

In this dataset, each entry corresponds to one material and provides three classes of information: the material-specific data, the band structure description data, and the spin-splitting description data.

The former provides the material's general description regarding its composition, crystal structure, band gap, and other properties. Box 1 shows a complete overview of this data. The band structure data (Box 2 gives information about the band structure calculations, such as the number of calculated bands, the number of k-points, the Brillouin zone, and all energy eigenvalues. Here, the key `NOMAD_files` has the URLs to the material-specific files for the band structure calculation at the NOMAD repository.

The spin splitting data, as the name suggests, provides information of all spin splittings occurring in the material's valence and conduction band, which compiles all the data resulting from SSIA into SS types (LSS, ZSS, or HOSS) and their specific parameters. Box 3 details all keys describing the SSs.

The user can easily query the data by importing it to MongoDB (check the MongoDB website on setting up a database in your specific system) or any other NoSQL database option. Alternatively, one can also convert the JSON file to a simple table, with special care to normalize the `vb/cb` keys and their subkeys. By using the MongoDB approach, one can query the unstructured spin splitting data for further analysis by using the many programming languages which MongoDB provides support, or by using the mongosh (shell environment) or even the Compass program (MongoDB's graphical user interface environment). In the README.md file from the Materials Cloud repository we also provide instructions for doing so. As an example, the simple query shown in Box 4 retrieves the cations, anions, and the structural cluster of all compounds with a Zeeman spin splitting greater than 0.1 eV in the valence or the conduction band. The resulting data will be used in the next section of Technical Validation.

Box 1 Material-specific description keys used in the MongoDB database provided in the Materials Cloud⁵⁸. In gray are the key labels followed by the data type (in pink), an example of the data value, and a brief description of their meaning.

<code>_id</code>	string	'AsISe-5d829e480507'	C2DB unique identifier.
<code>stoichiometry</code>	string	'ABC'	Anonymous formula describing the material stoichiometry.
<code>elements</code>	array[string]	['As', 'I', 'Se']	List of elements in composition.
<code>cations</code>	array[string]	['As']	List of cations defined by the calculated Bader charges.
<code>anions</code>	array[string]	['I', 'Se']	List of anions defined by the calculated Bader charges.
<code>unit_cell</code>	Object		Unit cell basis vectors (Å).
	.a1 array[float]	[3.9342, 0.0000, 0.0000]	
	.a2 array[float]	[-1.9671, 3.4071, 0.0000]	
	.a3 array[float]	[0.0000, 0.0000, 19.5000]	
<code>sites</code>	array[Object]		List of site objects. Each object provides the element and the coordinates of each occupied site in the unit cell.
	[
	element string	'As'	
	coords array[float]	[0.0000, 0.0000, 0.5104]	
	...]		
<code>bravais_lattice</code>	string	'primitive hexagonal'	Two-dimensional Bravais lattice (primitive hexagonal, centred rectangular, primitive oblique, primitive square, primitive rectangular).
<code>structural_cluster</code>	string	'ABC-c3'	Structural cluster label (see Section 1 of Supplementary file 6).
<code>space_group</code>	integer	156	Space group number.
<code>polar</code>	boolean	True	Crystal structure is polar.
<code>e_hull</code>	float	0.0	Energy above hull (eV).
<code>band_gap</code>	float	1.1641	Band gap (eV).

Technical Validation

The data generated in this work opens the way for two possible direct applications: i) the selection of optimized compound, i.e., *the best of a class* for a given application and ii) the understanding of the interplay between the physical mechanisms behind a given target property based on data trends that can be analyzed with the Bayesian optimization of materials properties. We then highlight in this technical validation section how the proposed workflow for selection and optimization leads to the *best of a class* for a 2D semiconductor with large SS, illustrating the use of the Bayesian optimization applied for the specific case of the Zeeman SS.

Bayesian inference. As symmetry and structure are major components of bands behavior, it is fundamental that we convey these properties into data in an appropriate way. The C2DB provides a label called crystal prototype, with a *stoichiometry - space group - occupied Wyckoff positions* format. The H-phase of monolayer MoS₂ is described as AB₂-187-ai, for example. We have observed that grouping the materials using this label leads to some erroneous clusters. Some materials that should be clustered together are separated, resulting in a sparse space of crystal prototypes. Figure 1 in Supplementary file 6 illustrates this and some other limitations of using space group or crystal prototype for grouping materials. To bypass these limitations and also to get to more intuitive conclusions regarding the structure degree of freedom, we used the methodology described in ref. 60 to generate crystal fingerprints (CF) for each of the materials in the C2DB. As the resulting CFs have large dimensionality, we then used them as input for the UMAP⁶¹ embedding technique. The result is a 2d embedding that characterizes the structural differences and similarities of C2DB's materials. We used DBSCAN⁶² to define the clusters in the embedding (Fig. 2 in Supplementary file 6) and then further subdivided them accordingly to the materials'

Box 2 Band structure description keys used in the MongoDB database provided in the Materials Cloud⁵⁸. In gray are the key labels followed by the data type (in pink), an example of the data value, and a brief description of their meaning.

n_electrons	integer	18	Total number of valence electrons used in the calculations.
n_bands	integer	36	Number of calculated bands.
BZ_cell	Object		Brillouin zone basis vectors (\AA^{-1}).
	.b1	array[float]	[0.2542, 0.1468, 0.0000]
	.b2	array[float]	[0.0000, 0.2935, 0.0000]
	.b3	array[float]	[0.0000, 0.0000, 0.0000]
n_kpoints	integer	202	Number of k-points along the entire band structure.
kpoints	array[array]		List of k-points' fractional coordinates. It is ordered by the band structure k-point sequence.
k_path	array[Object]	[[0.0000, 0.0000, 0.0000], ...]	List of high-symmetry k-points objects ordered by the band structure's k-path sequence. Each object provides the label and the coordinates of the respective k-point.
	[
	label	string	'G'
	coords	array[float]	[0.0000, 0.0000, 0.0000]
	...]		
bands	array[array]	[band0, band1, ...]	Array containing n_bands arrays, each representing an individual band with a total of n_kpoints energy eigenvalues.
NOMAD_files	array[string]	['https://nomad-lab.eu/...', ...]	List of URLs to input and output files of the band structure calculation (Vasp). These include the INCAR, POSCAR, KPOINTS, OUTCAR, and vasprun.xml.

stoichiometry. As an example, MoS₂ is labeled as AB₂-c22 because of its binary stoichiometry and its given cluster number (22). A total of 23 structural clusters were found for the 436 2D materials. Further details of this process are presented in the Section 1 of Supplementary file 6. Due to the categorical nature of the structural cluster (i.e., the material can only be at one specific cluster), we used the categorical distribution implementation of Naive Bayes for evaluating how being a member of such clusters weights for a given material regarding a target property.

As an illustrative example, we used the Zeeman spin-splitting (ZSS) as a target property to be analyzed. Setting a bottom limit of 100 meV for materials with “large ZSS”, we then proceeded to use Bayesian inference to generate probabilities for each structural cluster. These probabilities are the posterior probabilities, i.e., the probability that a material has a large ZSS given that it is in a given structural cluster. They were calculated using the categorical distribution given by Eq. 3. The results are presented in Fig. 5. Only structural clusters with at least 5 materials are shown and analyzed.

The posterior probabilities shows that the only structural clusters associated with the target are AB-c25 (Buckled Hexagonal AB), ABC-c3 (T-TMDC MXY Janus), ABC-c4 (H-TMDC MXY Janus), and AB₂-c4 (H-TMDC). All of them are hexagonal and present ZSS at the *K* points of the Brillouin Zone. The Zeeman-type spin splitting in these structures, except for ABC-c3, has been extensively studied in the context of valleytronics⁶³. In AB₂-C4 (H-TMDCs), the non-degeneracy of spin states at $\pm K$ -points combined with time-reversal symmetry requires that spin splittings at $+K$ and $-K$ to be opposite^{64,65}. This condition gives rise to the called valley Zeeman effect⁶⁶⁻⁶⁹, which happens when the degeneracy between valleys $+K$ and $-K$ is broken by applying a perpendicular magnetic field. The same characteristic can be found in the literature for ABC-c4 (H-TMDC MXY Janus) systems⁷⁰⁻⁷² and AB-c25 (Buckled Honeycomb AB) structures⁷³, with the latter presenting ferroelectricity as a co-functionality. For ABC-c3 (T-TMDC MXY Janus), the spin-splitting arises uniquely from the intrinsic out-of-plane dipole originated from the electronegativity difference between X and Y anions. Materials with this structure have been studied primarily because of their large Rashba-type spin splittings located around Γ ⁷⁴⁻⁷⁶. The ZSS localized at *K* shows to be inaccessible as it is usually distant from the VBM/CBM around Γ .

To evaluate the composition influence on spin splitting properties, we first separated the composition of each material into cations and anions. This separation is performed according to the calculated Bader charges⁷⁷ of each ion in the unit cell, which is currently available in C2DB. If the cation/anion X is (not) present at the material composition, the feature cation/anion X is set to (False) True. The distribution of cations and anions in the composition of all 436 materials is shown in Fig. 5 from Supplementary file 6. We used a Naive Bayes implemented with the Bernoulli distribution (Eq. 4) as these features are not exclusive, i.e., one entry can have more than one kind of cation/anion. The results are presented in Fig. 6. Only the cations/anions present in at least five

Box 3 Spin-splitting description keys used in the MongoDB database provided in the Materials Cloud⁵⁸. In gray are the key labels followed by the data type (in pink), an example of the data value, and a brief description of their meaning.

<code>vb</code>	Object	
		The object containing the valence band spin-splitting data description.
<code>.index</code>	integer	17
		Valence band index.
<code>.LSS</code>	array[Object]	
		List of linear spin splittings objects.
		[
<code>.rashba_param</code>	float	1.6258
		Rashba parameter (eV.Å). Calculated with the Eq. 1.
<code>.delta_e</code>	float	0.2679
		Delta energy (eV).
<code>.delta_k</code>	float	0.3296
		Delta k (Å ⁻¹).
<code>.accessibility</code>	float	0.5163
		Distance in energy to the VBM (eV).
<code>.kpoint</code>	array[float]	[0.3968, 0.2063, 0.0000]
		K-point fractional coordinates of the LSS.
<code>.k_segment</code>	string	'M-K'
		K-path segment which the LSS happens.
<code>.anti_crossing</code>	boolean	True
		Spin splitting has anti-crossing.
		...]
<code>.HOSS</code>	array[Object]	
		List of high-order spin-splittings objects. Each object has the same keys as the LSS object, except <code>rashba_param</code> as it is not defined for HOSS.
<code>.ZSS</code>	array[Object]	
		List of Zeeman spin splittings objects. Each object has three keys: <code>spin_splitting</code> and <code>accessibility</code> , with the same meaning as in the LSS object; and <code>kpoint</code> , which is a high-symmetry k-point object indicating the location of the ZSS.
<code>cb</code>	Object	
		The object containing the conduction band spin-splitting data description. It is analogous to the <code>vb</code> structure described above, with the exception that the SS values are calculated concerning the conduction band minimum instead.
<code>has_ZSS</code>	boolean	True
		Valence or conduction band has ZSS.
<code>has_LSS</code>	boolean	True
		Valence or conduction band has LSS.
<code>has_HOSS</code>	boolean	True
		Valence or conduction band has HOSS.
<code>vb_SS</code>	boolean	True
		Valence band has SS.
<code>cb_SS</code>	boolean	True
		Conduction band has SS.

materials are shown. We can see that the cations have a larger range of posterior probabilities, indicating that they are more important than anions regarding the Zeeman spin splitting in those systems. One can also notice the top-down trend in the periodic table groups: heavier atoms are associated with higher ZSS due to their higher spin-orbit coupling. The most significant cations were Bi, Sb, As, Ir and Hf, while the most important anions were the heaviest ones: Te and I.

This same methodology is then applied to Rashba/Dresselhaus spin-splitting, Rashba parameter, and band gap as target properties. The results of the Bayesian Inference analysis are provided by Table 3. The analogous heatmaps are presented in the Supplementary file 6 (Section 3 - Target properties structural and compositional heatmaps).

Case study for optimizing Zeeman spin splitting. Combining the optimal compositions and structures found in the previous section might lead to the exploration of novel spin-based materials by simple ion substitution. We used the results in Table 3 for ZSS and combined the optimal structural clusters, cations, and anions to generate a set of materials that are likely to show large ZSS (defined above as larger than 100 meV). The result is a set of 30 materials, from which nine were already in the screened dataset of 436 non-metal materials (Duplicated

Box 4 Example of querying the unstructured spin splitting data using MongoDB. The filter object specifies that only the materials presenting a ZSS greater than 0.1 eV (in the valence OR in the conduction band) should be selected. The project object sets the data keys that should be returned. The result is a list of 120 entries that will be used in the Technical Validation section.

```

1 //FILTER:
2 {$or: [
3   {"vb.zss.spin_splitting": {$gte: 0.1}},
4   {"cb.zss.spin_splitting": {$gte: 0.1}}
5 ]}
6 }

7 //PROJECT:
8 {"cations": 1,
9  "anions": 1,
10 "structural_cluster": 1}

11 //RESULT: 120 entries
12 [
13 {
14   "_id": "HgSe-619ed885f677",
15   "cations": ["Hg"],
16   "anions": ["Se"],
17   "structural_cluster": "AB-c2"
18 },
19 {
20   "_id": "HgO-a8678fa85c38",
21   "cations": ["Hg"],
22   "anions": ["O"],
23   "structural_cluster": "AB-c2"
24 },
25 ...
26 ]

```

NM), seven were in the general C2DB database as metals (Duplicated M), and 14 are new combinations (New). Figure 7a illustrates the combinatorial aspect of this ZSS optimization.

The new ion-substituted combinations were then structurally relaxed until the Hellmann-Feynman forces on each ion were less than 1.0 meV/Å, and then their respective band structures were calculated. All of them maintained their structural characteristics. Also, all of them presented a ZSS larger than 100 meV, the target of the optimization process. However, except for two materials (IrTeI on the two Janus phases), almost all of them resulted in metallic band structures. The set Duplicated M consisting of 7 metals in C2DB that fits the criteria also validates our optimization process, except for only one material with ZSS below the threshold. Figure 7b shows the ZSS distribution over these different sets of materials.

Best of a class: the case of Zeeman SS Materials. To present a list of “best of a class” materials from the resulting screened C2DB database, one first needs to specify what “best” means with well-defined objectives. For such, we delineated three parameters:

- **SS intensity:** Given by the energy delta of spin splitting. It is defined as Zeeman spin splitting for non-TRIM points and as Rashba/Dresselhaus spin splitting for TRIM points;
- **SS accessibility:** The energy delta of the spin splitting to the VBM or CBM. It measures how accessible the SS is for exploitation and experimental verification;
- **Energy above hull (ehull):** It measures the material stability to its decomposition to other structures with the same composition. An ehull of 0 eV means that the material is thermodynamically stable in comparison to other phases.

We described the dataset in terms of these three parameters so one can obtain materials with a large and accessible SS but, at the same time, that are thermodynamically stable in comparison to other competing phases. By filtering out the materials with ehull >30 meV, with accessibility <100 meV and Zeeman spin splitting >30 meV, we obtained the materials presented at Table 4.

Effect of anti-crossing bands. Acosta *et al.* were the first to propose a causal relationship between the anti-crossing orbital character of energy bands and a large Rashba coefficient for 3D-bulk compounds²³. In this work, we investigate the extension of this concept in the context of two-dimensional materials.

An anti-crossing analysis procedure is implemented in the SSIA and consists of two steps: i) identifying aligned pairs of SS between valence and conduction bands and ii) measuring the change in the orbital character of the selected bands in the k-interval between the high-symmetry k-point and the inflection point where the SS is maximum. If a monotonic and inverse change on the orbital contribution is present in the pair of SSs at the valence and conduction bands, the anti-crossing of SS bands (AC-SS) is verified.

Composition	Crystal structure	Zeeman spin splitting (meV)	Accessibility (meV)	Ehull (meV)
WTe ₂	H-TMDC	485.2	0.0	26.3
WSe ₂	H-TMDC	466.1	0.0	0.0
WSSe	H-TMDC MXY Janus	445.1	0.0	10.2
WS ₂	H-TMDC	430.0	0.0	0.0
MoTe ₂	H-TMDC	215.1	0.0	0.0
MoSeTe	H-TMDC MXY Janus	199.8	0.0	25.0
MoSe ₂	H-TMDC	184.6	0.0	0.0
MoSSe	H-TMDC MXY Janus	168.5	0.0	9.3
MoS ₂	H-TMDC	148.0	0.0	0.0
ZrI ₂	H-TMDC	121.8	0.0	27.0
Al ₂ Te ₂	InSe	118.0	55.4	0.0
CrSe ₂	H-TMDC	90.3	0.0	0.0
CrSSe	H-TMDC MXY Janus	81.5	0.0	10.4
ZrBr ₂	H-TMDC	75.8	0.0	0.0
CrS ₂	H-TMDC	68.3	0.0	0.0
ZrClBr	H-TMDC MXY Janus	59.4	0.0	9.8
Al ₂ Se ₂	InSe	56.3	62.8	0.0
TiBr ₂	H-TMDC	55.9	0.0	0.0
TiClBr	H-TMDC MXY Janus	45.4	0.0	15.3
ZrCl ₂	H-TMDC	40.8	0.0	0.0
TiCl ₂	H-TMDC	31.7	0.0	0.8

Table 4. Best-of-a-class ZSS materials according to Zeeman spin splitting, accessibility (how far from the band edge the splitting occurs) and ehull (energy above hull) values.

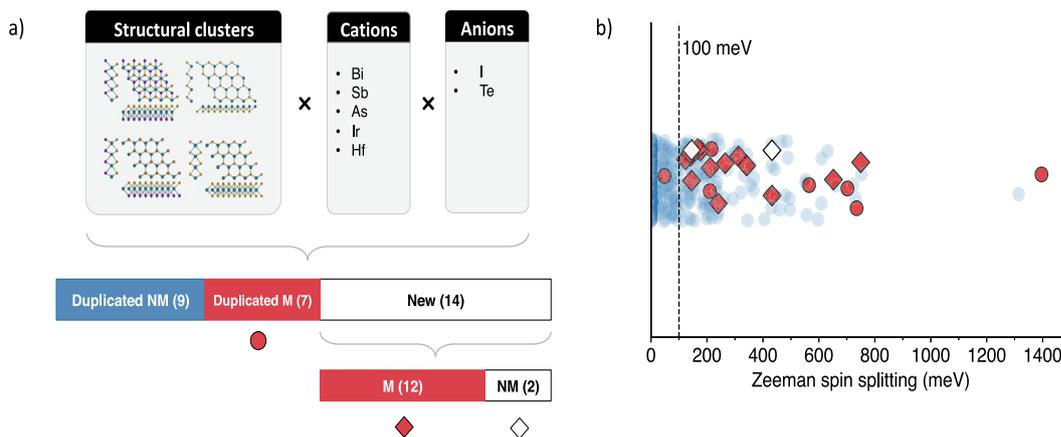


Fig. 7 (a) Illustration of the combination of optimized structural clusters and optimized compositions, resulting in 30 combinations. 'M' (red) and 'NM' indicate materials that are metals and non-metals, respectively. Duplicated NM: combinations already presented in the screened dataset. Duplicated M (red circle): combinations presented in the non-screened dataset. New (diamonds): new materials generated by optimizing criteria; (b) Strip plot of the Zeeman spin splitting. Blue circles correspond to all Zeeman spin splittings for all materials in the dataset (materials that do not have ZSS are represented by a Zeeman spin splitting of zero). White diamonds: new materials with band gap > 0.

Figure 8a illustrates this analysis for the 2D compounds with SS identified in this work. In such context, the presence of anti-crossing bands may indicate a sufficient condition, as all AC-SS present Rashba coefficients larger than $0.862 \text{ eV/\AA}^{-1}$. But they are not necessary, as far as other large Rashba compounds do not display a measurable anti-crossing between valence and conduction bands. While we understand that other symmetry-driven physical mechanisms can play an important role in SS in 2D compounds, an indirect effect of anti-crossing bands is also noted, as it is illustrated for the AsIS (MXY Janus) compound in Fig. 8b. For its respective band structure, anti-crossing is verified among the valence and other deeper bands, leading to a large Rashba coefficient for such SS that may have a consequential effect on the SS observed at the valence band.

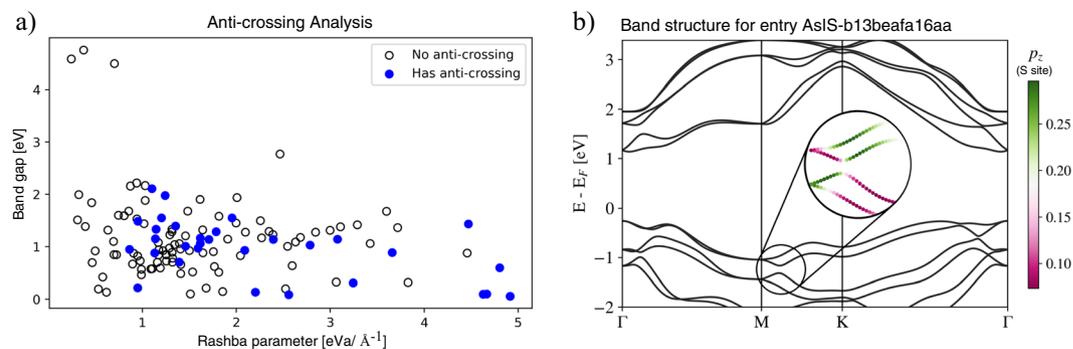


Fig. 8 Presence of anti-crossing bands. **(a)** Distribution of the 124 compounds with Rashba or Dresselhaus SS prototypes identified in this work, according to its Rashba parameter α_R and its band gap. Filled blue (unfilled black) dots correspond to compounds whose SS were identified to have (do not have) the presence of anti-crossing among valence and conduction bands. **(b)** Representation of a compound where anti-crossing among deeper bands may have significant consequences for the SS observed at the band edges, for the case of entry with C2DB uid AsIS-b13beafa16aa. The inset represents the p_z orbital character from the S site.

Usage Notes

As noted in the Data Records section, the complete data generated for all materials throughout this work are available in various formats, that may suit the reader's different needs and use cases. A complete guide to accessing and working with the data using python is available in a README markdown file in the Materials Cloud repository⁵⁸ for this work.

The computational code, i.e. a python class, used for all the SS analysis from the output of the DFT calculations is also provided (see Code Availability section) and can be used to identify and measure SS effects of other 2D materials calculations. Due to the context of this work, the algorithm currently supports non-collinear band structure calculations performed with VASP, whose k-paths over the reciprocal space have been generated with ASE's Brillouin zone sampling automatic scheme. The code initialization requires only the specification of the folder path where the band structure calculation was performed, and presents different methods to identify SS effects in materials valence/conduction bands and the presence of anti-crossing bands, and also offers tools for plotting band structures with spin texture resolution. A more detailed description of the current functionalities of the code are available in its GitHub repository.

Code availability

The entire computational code employed in the SS analysis within this work is openly available at the GitHub repository github.com/simcomat/SS_2D_Materials. It is intensely built upon tools and methods from Pymatgen⁴⁴ and ASE⁵⁴, and provide functions to identify, measure and classify SS effects that appear valence/conduction bands of 2D materials band structure calculations.

Received: 6 October 2021; Accepted: 3 March 2022;

Published online: 29 April 2022

References

- Zunger, A. Inverse design in search of materials with target functionalities. *Nat. Rev. Chem.* **2** (2018).
- Smidt, T. E., Mack, S. A., Reyes-Lillo, S. E., Jain, A. & Neaton, J. B. An automatically curated first-principles database of ferroelectrics. *Sci. Data* **7** (2020).
- Garrity, K. F. High-throughput first-principles search for new ferroelectrics. *Phys. Rev. B* **97**, 024115 (2018).
- Acharya, M. *et al.* Searching for new ferroelectric materials using high-throughput databases: An experimental perspective on BiAlO₃ and BiInO₃. *Chem. Mater.* **32**, 7274–7283 (2020).
- Mounet, N. *et al.* Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* **13**, 246–252 (2018).
- Zhou, J. *et al.* 2dmatpedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches. *Sci. Data* **6** (2019).
- Gjerding, M. N. *et al.* Recent progress of the computational 2d materials database (c2db). *2D Mater.* **8**, 044002 (2021).
- Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M. & Fazzio, A. From DFT to machine learning: recent approaches to materials science—a review. *J. Physics: Mater.* **2**, 032001 (2019).
- Rodrigues, J. F., Florea, L., de Oliveira, M. C. F., Diamond, D. & Oliveira, O. N. Big data and machine learning for materials science. *Discov. Mater.* **1** (2021).
- Schleder, G. R., Acosta, C. M. & Fazzio, A. Exploring two-dimensional materials thermodynamic stability via machine learning. *ACS Appl. Mater. & Interfaces* **12**, 20149–20157 (2019).
- Acosta, C. M. *et al.* Analysis of topological transitions in two-dimensional materials by compressed sensing. ArXiv 1805.10950 (2018).
- Schleder, G. R., Padilha, A. C. M., Rocha, A. R., Dalpian, G. M. & Fazzio, A. Ab initio simulations and materials chemistry in the age of big data. *J. Chem. Inf. Model.* **60**, 452–459 (2019).
- Laks, D. B., Ferreira, L. G., Froyen, S. & Zunger, A. Efficient cluster expansion for substitutional systems. *Phys. Rev. B* **46**, 12587–12605 (1992).
- Zutic, I., Fabian, J. & Das Sarma, S. Spintronics: Fundamentals and applications. *Rev. Mod. Phys.* **76**, 323–410 (2004).
- Yang, S.-H., Naaman, R., Paltiel, Y. & Parkin, S. S. P. Chiral spintronics. *Nat. Rev. Phys.* **3**, 328–343 (2021).

16. Acosta, C. A. M. *Transistor spintronico: descoberta e caracterizagao de isolantes topologicos*. Ph.D. thesis, Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA) (2018).
17. Sierra, J. F., Fabian, J., Kawakami, R. K., Roche, S. & Valenzuela, S. O. Van der waals heterostructures for spintronics and opto-spintronics. *Nat. Nanotechnol.* **16**, 856–868 (2021).
18. Xia, W. *et al.* Recent progress in van der waals heterojunctions. *Nanoscale* **9**, 4324–4365 (2017).
19. Kamalakar, M. V., Dankert, A. & Dash, S. P. Spintronics with graphene and van der waals heterostructures. In *Contemporary Topics in Semiconductor Spintronics*, 241–258, (WORLD SCIENTIFIC, 2017).
20. Acosta, C. M. & Fazzio, A. Spin-polarization control driven by a rashba-type effect breaking the mirror symmetry in two-dimensional dual topological insulators. *Phys. Rev. Lett.* **122**, 036401 (2019).
21. Mera Acosta, C., Babilonia, O., Abdalla, L. & Fazzio, A. Unconventional spin texture in a noncentrosymmetric quantum spin hall insulator. *Phys. Rev. B* **94**, 041302 (2016).
22. Pan, H., Wu, M., Liu, Y. & Yang, S. A. Electric control of topological phase transitions in dirac semimetal thin films. *Sci. Reports* **5** (2015).
23. Acosta, C. M., Ogoshi, E., Fazzio, A., Dalpian, G. M. & Zunger, A. The rashba scale: Emergence of band anti-crossing as a design principle for materials with large rashba coefficient. *Matter* **3**, 145–165 (2020).
24. Bychkov, Y. A. & Rashba, E. I. Properties of a 2D electron gas with lifted spectral degeneracy. *Sov. J. Exp. Theor. Phys. Lett.* **39**, 78 (1984).
25. Manchon, A., Koo, H. C., Nitta, J., Frolov, S. M. & Duine, R. A. New perspectives for rashba spin-orbit coupling. *Nat. Mater.* **14**, 871–882 (2015).
26. Dresselhaus, G. Spin-orbit coupling effects in zinc blende structures. *Phys. Rev.* **100**, 580–586 (1955).
27. Acosta, C. M., Fazzio, A. & Dalpian, G. M. Zeeman-type spin splitting in nonmagnetic three-dimensional compounds. *npj Quantum Mater.* **4** (2019).
28. Yuan, H. *et al.* Zeeman-type spin splitting controlled by an electric field. *Nat. Phys.* **9**, 563–569 (2013).
29. Acosta, C. M., Yuan, L., Dalpian, G. M. & Zunger, A. Different shapes of spin textures as a journey through the brillouin zone. *Phys. Rev. B* **104** (2021).
30. Vajna, S. *et al.* Higher-order contributions to the rashba-bychkov effect with application to the bi/ag(111) surface alloy. *Phys. Rev. B* **85**, 075404 (2012).
31. Cartoixa, X., Wang, L., Ting, D. & Chang, Y. Higher-order contributions to rashba and dresselhaus effects. *Phys. Rev. B* **73**, 205341 (2006).
32. Acosta, C. M., Fazzio, A., Dalpian, G. M. & Zunger, A. Inverse design of compounds that have simultaneously ferroelectric and rashba cofunctionality. *Phys. Rev. B* **102**, 144106 (2020).
33. Zunger, A. & Malyi, O. I. Understanding doping of quantum materials. *Chem. Rev.* **121**, 3031–3060 (2021).
34. Malyi, O. I. & Zunger, A. False metals, real insulators, and degenerate gapped metals. *Appl. Phys. Rev.* **7**, 041310 (2020).
35. Zhang, X., Liu, Q., Luo, J.-W., Freeman, A. J. & Zunger, A. Hidden spin polarization in inversion-symmetric bulk crystals. *Nat. Phys.* **10**, 387–393 (2014).
36. Liu, Q., Guo, Y. & Freeman, A. J. Tunable rashba effect in two-dimensional LaOBiS₂ films: Ultrathin candidates for spin field effect transistors. *Nano Lett.* **13**, 5264–5270 (2013).
37. Zhang, R., Marrazzo, A., Verstraete, M. J., Marzari, N. & Sohler, T. D. P. Gate control of spin-layer-locking FETs and application to monolayer LuIO. *Nano Lett.* **21**, 7631–7636 (2021).
38. MaaB, H. *et al.* Spin-texture inversion in the giant rashba semiconductor BiTeI. *Nat. Commun.* **7** (2016).
39. Ishizaka, K. *et al.* Giant rashba-type spin splitting in bulk BiTeI. *Nat. Mater.* **10**, 521–526 (2011).
40. Feng, Y. *et al.* Rashba-like spin splitting along three momentum directions in trigonal layered PtBi₂. *Nat. Commun.* **10** (2019).
41. Zeeman, P. On the influence of magnetism on the nature of the light emitted by a substance. *The London, Edinburgh, Dublin Philos. Mag. J. Sci.* **43**, 226–239 (1897).
42. Hastrup, S. *et al.* The computational 2d materials database: high-throughput modeling and discovery of atomically thin crystals. *2D Mater.* **5**, 042002 (2018).
43. Larsen, P. M., Pandey, M., Strange, M. & Jacobsen, K. W. Definition of a scoring parameter to identify low-dimensional materials components. *Phys. Rev. Mater.* **3**, 034003 (2019).
44. Ong, S. P. *et al.* Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
45. Tao, L. & Tsybal, E. Y. Insulator-to-conductor transition driven by the rashba-zeeman effect. *npj Comput. Mater.* **6** (2020).
46. Pekar, S. & Rashba, E. Combined resonance in crystals in inhomogeneous magnetic fields. *Sov. Phys.-JETP* **20**, 1295 (1965).
47. Naka, M. *et al.* Spin current generation in organic antiferromagnets. *Nat. Commun.* **10** (2019).
48. Kabiraj, A., Kumar, M. & Mahapatra, S. High-throughput discovery of high curie point two-dimensional ferromagnetic materials. *npj Comput. Mater.* **6** (2020).
49. Acosta, C. M., Ogoshi, E., Souza, J. A. & Dalpian, G. M. Machine learning study of the magnetic ordering in 2d materials. *ACS Appl. Mater. & Interfaces.* **14**, 9418–9432 (2022).
50. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
51. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
52. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
53. Bahn, S. R. & Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **4**, 56–66 (2002).
54. Larsen, A. H. *et al.* The atomic simulation environment—a python library for working with atoms. *J. Physics: Condens. Matter* **29**, 273002 (2017).
55. Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
56. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
57. Talirz, L. *et al.* Materials cloud, a platform for open computational science. *Sci. Data* **7** (2020).
58. Nascimento, G. M., Ogoshi, E., Fazzio, A., Acosta, C. M. & Dalpian, G. M. High throughput inverse design and bayesian optimization of functionalities: spin splitting in two-dimensional compounds. *Mater. Cloud Arch.* **2021.224**, <https://doi.org/10.24435/materialscld:kr-7s> (2021).
59. Nascimento, G. M., Ogoshi, E., Fazzio, A., Acosta, C. M. & Dalpian, G. M. 2d ss materials. *NOMAD* <https://doi.org/10.17172/NOMAD/2021.09.20-2> (2021).
60. Zimmermann, N. E. & Jain, A. Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity. *RSC Adv.* **10**, 6063–6081 (2020).
61. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXivpreprint arXiv:1802.03426* (2018).
62. Ester, M. *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd* **96**, 226–231 (1996).
63. Schaibley, J. R. *et al.* Valleytronics in 2d materials. *Nat. Rev. Mater.* **1**, 1–15 (2016).

64. Xiao, D., Liu, G.-B., Feng, W., Xu, X. & Yao, W. Coupled spin and valley physics in monolayers of mos 2 and other group-vi dichalcogenides. *Phys. review letters* **108**, 196802 (2012).
65. Jones, A. M. *et al.* Optical generation of excitonic valley coherence in monolayer wse 2. *Nat. nanotechnology* **8**, 634–638 (2013).
66. Aivazian, G. *et al.* Magnetic control of valley pseudospin in monolayer wse 2. *Nat. Phys.* **11**, 148–152 (2015).
67. Srivastava, A. *et al.* Valley zeeman effect in elementary optical excitations of monolayer wse 2. *Nat. Phys.* **11**, 141–147 (2015).
68. MacNeill, D. *et al.* Breaking of valley degeneracy by magnetic field in monolayer mose 2. *Phys. review letters* **114**, 037401 (2015).
69. Li, Y. *et al.* Valley splitting and polarization by the zeeman effect in monolayer mose2. *Phys. review letters* **113**, 266804 (2014).
70. Peng, R., Ma, Y., Zhang, S., Huang, B. & Dai, Y. Valley polarization in janus single-layer mosse via magnetic doping. *The journal physical chemistry letters* **9**, 3612–3617 (2018).
71. Hu, T. *et al.* Intrinsic and anisotropic rashba spin splitting in janus transition-metal dichalcogenide monolayers. *Phys. Rev. B* **97**, 235404 (2018).
72. Cheng, Y., Zhu, Z., Tahir, M. & Schwingenschlogl, U. Spin-orbit-induced spin splittings in polar transition metal dichalcogenide monolayers. *EPL (Europhysics Lett.)* **102**, 57001 (2013).
73. Di Sante, D., Stroppa, A., Barone, P., Whangbo, M.-H. & Picozzi, S. Emergence of ferroelectricity and spin-valley properties in two-dimensional honeycomb binary compounds. *Phys. Rev. B* **91**, 161401 (2015).
74. Ma, Y., Dai, Y., Wei, W., Li, X. & Huang, B. Emergence of electric polarity in bitex (x= br and i) monolayers and the giant rashba spin splitting. *Phys. Chem. Chem. Phys.* **16**, 17603–17609 (2014).
75. Zhuang, H. L. *et al.* Rashba effect in single-layer antimony telluroiodide sbte. *Phys. Rev. B* **92**, 115302 (2015).
76. Riis-Jensen, A. C., Deilmann, T., Olsen, T. & Thygesen, K. S. Classifying the electronic and optical properties of janus monolayers. *ACS nano* **13**, 13354–13364 (2019).
77. Tang, W., Sanville, E. & Henkelman, G. A grid-based bader analysis algorithm without lattice bias. *J. Physics: Condens. Matter* **21**, 084204 (2009).

Acknowledgements

The authors thank Brazilian agencies FAPESP (2017/02317-2, 2018/11641-0, 2018/11856-7 and 2019/04176-2) and CNPq for financial support. High throughput calculations were performed using the Santos Dumont supercomputer at LNCC.

Author contributions

G.M.N. performed the ab initio calculations, wrote the code that calculated the spin splittings and generated band structures and datasets. E.O. performed the inverse optimization process, with the Bayes analysis and organized the MongoDB database. C.M.A. and G.M.D. conceived the idea. A.F., C.M.A. and G.M.D. supervised the work. All authors helped writing and reviewing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01292-8>.

Correspondence and requests for materials should be addressed to C.M.A. or G.M.D.

Reprints and permissions information is available at www.nature.com/reprints.

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) 2022