Supporting Information: Electronic band structure screening for Dirac points in Heuslers

Meza-Morales, Paul J., Taliaronak, Volha, Fumarola, Alessandro, Shirsekar, Afrid, kidner, Jonathan, Ali, Zaheer, and Ali. Mazhar

Material Mind, Fremont, CA 94555, United States

January 26, 2023

Section S1. Feature Engineering

The table S1 lists properties used as primary features (PFs) and their definition. Tables S3 and S2 list the selected features for the development of Heusler and Cubic compounds datasets ML models respectively.

Table S1. Primary features with their definitions

Electronic Property name/ matminer abbreviation	**Description/Matminer abbreviation**
maximum ionic charge/ max ionic char	Element with the highest oxidation states in the crystal structure
average ionic charge/ avg ionic char	Element with the highest oxidation states in the crystal structure
ewald energy per atom/ ewald_energy_per_atom	Computed energy from Coulombic interactions, using charges already defined for the crystal structure
HOMO element/ HOMO_element	Element with the highest occupied molecular orbital (HOMO) estimated from the atomic orbital energies of the structure chemical composition
LUMO element/ LUMO_element	Element with the lowest unoccupied molecular orbital (LUMO) estimated from the atomic orbital energies of the structure chemical composition
HOMO character/ HOMO_character	Atomic orbital character for the HOMO element ('s', 'p', 'd', or 'f')
LUMO character/ LUMO_character	Atomic orbital character for the LUMO element ('s', 'p', 'd', or 'f')
HOMO_energy	Atomic orbital energy for the HOMO character
LUMO_energy	Atomic orbital energy for the LUMO character
HOMO-LUMO atomic orbital gap/ gap_AO	bandgap from HOMO and LUMO energies
fraction of s valence electrons/ frac s valence electrons	Weighted fraction of 's' valence electrons estimated from the structure chemical composition
fraction of p valence electrons/ frac p valence electrons	Weighted fraction of 'p' valence electrons estimated from the structure chemical composition
fraction of d valence electrons/ frac d valence electrons	Weighted fraction of 'd' valence electrons estimated from the structure chemical composition
fraction of f valence electrons/ frac f valence electrons	Weighted fraction of 'f' valence electrons estimated from the structure chemical composition
Crystal	**Description**
Density	Crystal density
Volume per atom	Average volume taken up by an atom in the crystal structure
structural complexity per cell	Shannon information entropy of a structure. This descriptor treat a structure as a message to evaluate structural complexity (bits/cell)
structural complexity per atom	Shannon information entropy of a structure. This descriptor treat a structure as a message to evaluate structural complexity (bits/atom)
max packing efficiency	Maximum possible packing efficiency of this structure
mean neighbor distance variation	Statistic (e.g., mean) of the neighbor distance variation
mean absolute deviation in relative cell size	Voronoi cell volume across all sites in the structure, divided by the mean Voronoi cell volume
mean absolute deviation in relative bond length	The average bond lengths for all sites, divided by the average bond length

Table S2. Heusler dataset ML model features

Electronic
frac s valence electrons
frac d valence electrons
max ionic char
HOMO_element
avg ionic char
gap_AO
LUMO_element
LUMO_energy
HOMO_character
LUMO_character
frac f valence electrons
frac p valence electrons
HOMO_energy
ewald_energy_per_atom
Crystal
structural complexity per atom
mean absolute deviation in relative cell size
structural complexity per cell
mean neighbor distance variation
mean absolute deviation in relative bond length
volume per atom
packing fraction
max packing efficiency
density

Table S3. Cubic dataset ML model features

Electronic
HOMO_element
LUMO_character
LUMO_element
frac d valence electrons
max ionic char
LUMO_energy
HOMO_energy
frac s valence electrons
frac p valence electrons
packing fraction
frac f valence electrons
ewald_energy_per_atom
avg ionic char
HOMO_character
gap_AO
Crystal
structural complexity per cell
structural complexity per atom
max packing efficiency
mean neighbor distance variation
volume per atom (vpa)
mean absolute deviation in relative cell size
mean absolute deviation in relative bond length
density

Section S2. Band Structure Features Detection

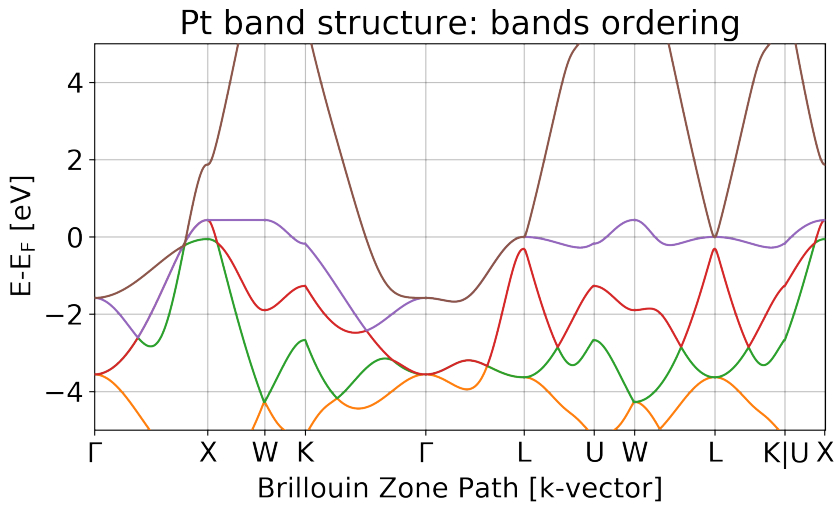

Figure S1. pymatgen band structure of cubic elemental Platinum (Pt, space group 225, mp-126 on Materials Project). Each individual band is plotted in a different color. The relative order of the bands does not change across the BZ path. The band plotted in brown is the highest in the y-axis range, while the orange one is the lowest.

Automatic detection and identification of band structure features can dramatically accelerate large-scale discovery of materials property. The number and type of band structure features are detected using an algorithm developed in *python* with the numerical library *numpy* as the only functional dependency (other libraries are needed for visualization, data management and debug).

The input data is a `pymatgen.electronic_structure.bandstructure.BandStructure` object with the bands represented as a $M \cdot N$ `numpy.array` of eigenvalues. M is the number of bands calculated for the given material while N is the number of reciprocal space locations sampled for each band. A separate array `BandStructure.kpoints` of shape $N \cdot 3$ is also provided to locate the k-points within the Brillouin Zone. The sequence of k-points follows the Setyawan-Curtarolo [1] convention.

Band structure features usually arise from the interaction of multiple bands. The first step of the feature detection algorithm consists in producing the band indexes $i, j \in [1, M]$ of the bands that need to be compared. All three types of features we are focusing on (i.e. degeneracy lines, degeneracy points and gaps) are characteristic of the interaction between two bands, but there is no theoretical limitation to the number of bands that contributes to the calculation. Bands in *pymatgen* keep their relative order which means that if a band has a largest energy value at the Γ point with respect to another, it will keep a higher or equal value for the whole Brillouin Zone path. For this reason, we never compare bands whose indexes are too far from each other as they will most likely have a large energy difference and small interaction. Each band is compared with the 2 closest higher bands and the 2 lower.

```
def find_degeneracy_lines(
    eigenvalues,        # M x N matrix with band structure
    max_delta_energy,   # maximum energy difference
                        # for degenerate bands
    delta_k,            # minimum number of consecutive
                        # degenerate points to form a degenerate line
                        )
    # Calculate indexes of the bands that need to be
    # combined
    index_a, index_b = make_combinations(
                            eigenvalues.shape[0],
                            n_bands_in_combination = 2
                            )

    band_a = eigenvalues[index_a,:] # Shape 1 x N
    band_b = eigenvalues[index_b,:] # Shape 1 x N
    diffs = band_b - band_a         # Shape 1 x N

    # Sequence of Boolean operations to verify
    # existence of a band structure feature
    condition = (diffs<=max_delta_energy)   # Shape 1 x N

    # If a sequence of true values is shorter
    # than delta_k, set all the sequence to False
    condition = filter_short_sequences(
                    condition,
                    delta_k
                    )

    k_indexes = where(condition == True)

    return k_indexes
```

The band pairs are sequentially inputted to the function `find_feature()` together with auxiliary parameters which determine the type of feature the user wants to detect. For each run of the `find_feature()` function, that is the detection of a given feature, we assign a truth value to each

eigenvalue of the band pairs. If at the end of the function this value is `True`, it means that the desired feature is found at the assigned array index $k < N$ (each index corresponds to a reciprocal space location). A series of binary checks are performed on the band values and their combination to determine whether the conditions apply to detect a feature. A pseudo-code to detect degeneracy lines is reported above.

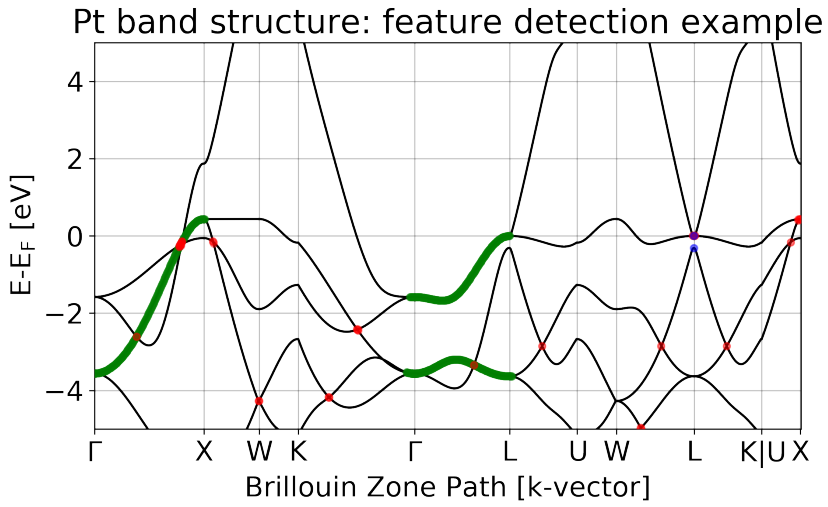

Figure S2. Band structure of cubic elemental Platinum (Pt, space group 225, mp-126 on Materials Project). The eigenvalues are plotted in black, the degeneracy lines are highlighted in green. The degeneracy points and inter-band gaps are highlighted in red and blue respectively

References

[1] W. Setyawan, S. Curtarolo, High-throughput electronic band structure calculations: Challenges and tools, Computational materials science 49 (2) (2010) 299–312.