Supplementary Information for “New Opportunities for Materials Informatics: Resources and Techniques for Uncovering Hidden Relationships”

# Details of Valence and Conduction Band Character Analysis

## Data preparation and number of compounds in study

All data was queried from the Materials Project database1 (April 27, 2015) using the Materials API2 and pymatgen code base3. The following constraints were applied:

* Only elements with Z<=54; we purposely avoided heavy elements due to (i) difficulty in treating *f* electrons in DFT and (ii) because Materials Project data is run without spin-orbit coupling which becomes more relevant for high Z systems.
* Band gap > 0.2, to avoid metallic systems that do not have a clearly defined VBM/CBM.
* Use of BVAnalyzer class in the pymatgen to determine valences; only compounds with a single valence per element were used.
* The material contained one or more of the ionic orbitals of interested, listed in Table 1.

A total of 2558 materials were used in the study.

## Computation Methodology

Details of the Materials Project database computational methodology are listed at its Wiki (www.materialsproject.org/wiki). Here, we mention that *d* states were treated with the GGA+*U* method using a fitting procedure from Wang et al.4 tailored to plain oxides. We note that while the MP methodology should give qualitatively correct results for the majority of transition metals and compounds, the *+U* value in principle depends on the system. See Table 2 for a list of *+U* values and fitting details.

## Analysis of DOS (“scoring” the VBM and CBM character)

The DOS were smeared with a width of 0.25 eV before scoring to smooth out numerical artifacts. In order to determine the “score” of a particular ionic orbital to the VBM, CBM, the projected DOS for that ionic orbital was integrated from the band edge out to 1 eV with a **Fermi weight** that favors states near the band edge. The Fermi weight takes the form , where T was set to 300K, *k* is the Boltzman constant, and *E-Eb­* is the distance from the band edge.

## Bradley-Terry analysis

The band edge data was transformed to symmetric pairwise competitions and fed into the BradleyTerry2 package5 in R to determine win/loss probabilities. The Bradley-Terry model produces scores for each ionic orbital along with an associated standard error. Each score, represented by λ, can be used to calculated win/loss probabilities between pairs of ionic orbitals using the equation ln(O) = λi - λj, where O is the odds of ionic orbital *i* having a greater contribution to a band edge than orbital *j*.

Ionic orbitals with very large standard error (greater than 1) were filtered out. The raw scores from the Bradley-Terry analysis are presented in Table 3 and Table 4.

## References

(1) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. *Commentary: The Materials Project: A materials genome approach to accelerating materials innovation*, APL Mater., 2013, *1*, 011002, doi:10.1063/1.4812323.

(2) Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. a. *The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles*, Comput. Mater. Sci., 2015, *97*, 209–215, doi:10.1016/j.commatsci.2014.10.037.

(3) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. *Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis*, Comput. Mater. Sci., 2013, *68*, 314–319, doi:10.1016/j.commatsci.2012.10.028.

(4) Wang, L.; Maxisch, T.; Ceder, G. *Oxidation energies of transition metal oxides within the GGA+U framework*, Phys. Rev. B, 2006, *73*, 195107, doi:10.1103/PhysRevB.73.195107.

(5) Turner, H.; Firth, D. *Bradley-Terry Models in R: The BradleyTerry2 Package*, J. Stat. Softw., 2012, *48*, 1–21.

# Data Tables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Li(1+):s | Mn(3+):s | Zn(4+):d | Pd(3+):p | Al(3+):p | S(2-):p |
| Be(1+):s | Mn(3+):p | Zr(2+):s | Pd(3+):d | Ga(3+):s | Te(2-):s |
| Na(1+):s | Mn(3+):d | Zr(2+):p | Pd(4+):s | Ga(3+):p | Te(2-):p |
| Na(1+):p | Mn(4+):s | Zr(2+):d | Pd(4+):p | Ga(3+):d | Te(2-):d |
| K(1+):s | Mn(4+):p | Zr(3+):s | Pd(4+):d | In(3+):s | Se(2-):s |
| K(1+):p | Mn(4+):d | Zr(3+):p | Ag(2+):s | In(3+):p | Se(2-):p |
| Rb(1+):s | Fe(2+):s | Zr(3+):d | Ag(2+):p | In(3+):d | Se(2-):d |
| Rb(1+):p | Fe(2+):p | Zr(4+):s | Ag(2+):d | C(4+):s | F(1-):s |
| Mg(2+):s | Fe(2+):d | Zr(4+):p | Ag(3+):s | C(4+):p | F(1-):p |
| Mg(2+):p | Fe(3+):s | Zr(4+):d | Ag(3+):p | C(4-):s | Cl(1-):s |
| Ca(2+):s | Fe(3+):p | Nb(2+):s | Ag(3+):d | C(4-):p | Cl(1-):p |
| Ca(2+):p | Fe(3+):d | Nb(2+):p | Ag(4+):s | Si(4+):s | Br(1-):s |
| Sr(2+):s | Fe(4+):s | Nb(2+):d | Ag(4+):p | Si(4+):p | Br(1-):p |
| Sr(2+):p | Fe(4+):p | Nb(3+):s | Ag(4+):d | Si(4-):s | Br(1-):d |
| Sc(3+):s | Fe(4+):d | Nb(3+):p | Cd(2+):s | Si(4-):p | I(1-):s |
| Sc(3+):p | Co(2+):s | Nb(3+):d | Cd(2+):p | Ge(4+):s | I(1-):p |
| Sc(3+):d | Co(2+):p | Nb(4+):s | Cd(2+):d | Ge(4+):p | I(1-):d |
| Y(3+):s | Co(2+):d | Nb(4+):p | Cd(3+):s | Ge(4+):d |  |
| Y(3+):p | Co(3+):s | Nb(4+):d | Cd(3+):p | Ge(4-):s |  |
| Y(3+):d | Co(3+):p | Mo(2+):s | Cd(3+):d | Ge(4-):p |  |
| Ti(2+):s | Co(3+):d | Mo(2+):p | Cd(4+):s | Ge(4-):d |  |
| Ti(2+):p | Co(4+):s | Mo(2+):d | Cd(4+):p | Sn(4+):s |  |
| Ti(2+):d | Co(4+):p | Mo(3+):s | Cd(4+):d | Sn(4+):p |  |
| Ti(3+):s | Co(4+):d | Mo(3+):p | Cu(1+):s | Sn(4+):d |  |
| Ti(3+):p | Ni(2+):s | Mo(3+):d | Cu(1+):p | Sn(4-):s |  |
| Ti(3+):d | Ni(2+):p | Mo(4+):s | Cu(1+):d | Sn(4-):p |  |
| Ti(4+):s | Ni(2+):d | Mo(4+):p | Ag(1+):s | Sn(4-):d |  |
| Ti(4+):p | Ni(3+):s | Mo(4+):d | Ag(1+):p | N(5+):s |  |
| Ti(4+):d | Ni(3+):p | Ru(2+):s | Ag(1+):d | N(5+):p |  |
| V(2+):s | Ni(3+):d | Ru(2+):p | Nb(5+):s | N(3-):s |  |
| V(2+):p | Ni(4+):s | Ru(2+):d | Nb(5+):p | N(3-):p |  |
| V(2+):d | Ni(4+):p | Ru(3+):s | Nb(5+):d | P(5+):s |  |
| V(3+):s | Ni(4+):d | Ru(3+):p | Mo(5+):s | P(5+):p |  |
| V(3+):p | Cu(2+):s | Ru(3+):d | Mo(5+):p | P(3-):s |  |
| V(3+):d | Cu(2+):p | Ru(4+):s | Mo(5+):d | P(3-):p |  |
| V(4+):s | Cu(2+):d | Ru(4+):p | V(5+):s | As(5+):s |  |
| V(4+):p | Cu(3+):s | Ru(4+):d | V(5+):p | As(5+):p |  |
| V(4+):d | Cu(3+):p | Rh(2+):s | V(5+):d | As(5+):d |  |
| Cr(2+):s | Cu(3+):d | Rh(2+):p | Cr(5+):s | As(3-):s |  |
| Cr(2+):p | Cu(4+):s | Rh(2+):d | Cr(5+):p | As(3-):p |  |
| Cr(2+):d | Cu(4+):p | Rh(3+):s | Cr(5+):d | As(3-):d |  |
| Cr(3+):s | Cu(4+):d | Rh(3+):p | Mo(6+):s | Sb(5+):s |  |
| Cr(3+):p | Zn(2+):s | Rh(3+):d | Mo(6+):p | Sb(5+):p |  |
| Cr(3+):d | Zn(2+):p | Rh(4+):s | Mo(6+):d | Sb(5+):d |  |
| Cr(4+):s | Zn(2+):d | Rh(4+):p | Cr(6+):s | Sb(3-):s |  |
| Cr(4+):p | Zn(3+):s | Rh(4+):d | Cr(6+):p | Sb(3-):p |  |
| Cr(4+):d | Zn(3+):p | Pd(2+):s | Cr(6+):d | Sb(3-):d |  |
| Mn(2+):s | Zn(3+):d | Pd(2+):p | B(3+):s | O(2-):s |  |
| Mn(2+):p | Zn(4+):s | Pd(2+):d | B(3+):p | O(2-):p |  |
| Mn(2+):d | Zn(4+):p | Pd(3+):s | Al(3+):s | S(2-):s |  |

Table 1 All ionic orbitals considered in this work

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Element** | **System** | **Fitting Reaction** | **Redox Couple** | **Calibrated U (eV)** |
| Co | Oxides | 6 CoO + O2 → 2 Co3O4 | Co2+ → Co2.67+ | 3.32 |
| Cr | Oxides | 2/3 Cr2O3 + O2 → 4/3 CrO3 | Cr3+ → Cr6+ | 3.7 |
| Fe | Oxides | 6 FeO + O2 → 2 Fe3O4  4 Fe3O4 + O2 → 6 Fe2O3 | Fe2+ → Fe2.67+  Fe2.67+ → Fe3+ | 5.3 |
| Mn | Oxides | 6 MnO + O2 → 2 Mn3O4  Mn3O4 + O2 → 3 MnO2 | Mn2+ → Mn2.67+  Mn2.67+ → Mn4+ | 3.9 |
| Mo | Oxides | 2 MoO2 + O2 → 2 MoO3 | Mo4+ → Mo6+ | 4.38 |
| Ni | Oxides | Li2O + NiO + O2 → LiNiO2 | Ni2+ → Ni3+ | 6.45 |
| V | Oxides | 2 V2O3 + O2 → 4 VO2  4 VO2 + O2 → 2 V2O5 | V3+ → V4+  V4+ → V5+ | 3.25 |
| W | Oxides | 2 WO2 + O2 → 2 WO3 | W4+ → W6+ | 6.2 |

Table 2 GGA+U values used in this work

|  |  |  |
| --- | --- | --- |
|  | ability | s.e. |
| Cu^1:d | 3.860 | 0.449 |
| Pd^2:d | 2.312 | 0.538 |
| V^3:d | 2.199 | 0.329 |
| Zr^2:d | 2.118 | 0.726 |
| As^-3:p | 2.092 | 0.578 |
| Fe^2:d | 1.495 | 0.313 |
| V^2:d | 1.484 | 0.404 |
| Nb^4:d | 1.454 | 0.584 |
| Co^2:d | 1.320 | 0.269 |
| Mo^2:d | 1.319 | 0.453 |
| Sb^-3:p | 1.223 | 0.525 |
| C^-4:p | 1.207 | 0.858 |
| P^-3:p | 1.074 | 0.556 |
| Te^-2:p | 0.916 | 0.347 |
| Se^-2:p | 0.695 | 0.285 |
| Mo^4:d | 0.650 | 0.444 |
| N^-3:p | 0.612 | 0.192 |
| I^-1:p | 0.583 | 0.215 |
| Mn^2:d | 0.106 | 0.205 |
| V^4:d | 0.046 | 0.194 |
| Br^-1:p | 0.033 | 0.215 |
| Si^-4:p | 0.003 | 0.992 |
| Ag^1:d | 0.000 | 0.000 |
| Cr^3:d | -0.186 | 0.209 |
| O^-2:p | -0.541 | 0.143 |
| Cl^-1:p | -0.602 | 0.190 |
| S^-2:p | -0.849 | 0.190 |
| Co^3:d | -2.248 | 0.294 |
| F^-1:p | -2.322 | 0.163 |
| Cr^5:d | -2.651 | 0.441 |
| Cr^4:d | -2.704 | 0.388 |
| Ni^2:d | -2.762 | 0.276 |
| Mn^3:d | -2.941 | 0.279 |
| Cu^3:d | -3.138 | 0.811 |
| Fe^3:d | -3.158 | 0.186 |

Table Full list of scores (abilities) and standard errors (s.e.) from the valence band analysis

|  |  |  |
| --- | --- | --- |
|  | ability | s.e. |
| V^4:d | 5.898 | 0.434 |
| V^3:d | 5.643 | 0.448 |
| Pd^2:d | 5.495 | 0.783 |
| Fe^3:d | 5.202 | 0.260 |
| Mn^3:d | 5.186 | 0.447 |
| Nb^5:d | 5.040 | 0.253 |
| Cr^4:d | 4.983 | 0.406 |
| V^5:d | 4.631 | 0.234 |
| Co^4:d | 4.617 | 0.366 |
| Cr^3:d | 4.463 | 0.319 |
| Cu^2:d | 4.257 | 0.493 |
| V^2:d | 4.203 | 0.491 |
| Cr^2:d | 3.868 | 0.453 |
| Ni^3:d | 3.610 | 0.726 |
| Mo^3:d | 3.584 | 0.597 |
| Zr^4:d | 3.530 | 0.187 |
| Ti^4:d | 3.231 | 0.145 |
| Sc^3:d | 2.971 | 0.193 |
| Pd^4:d | 2.768 | 0.536 |
| Mn^4:d | 2.740 | 0.159 |
| Co^2:d | 2.697 | 0.187 |
| Fe^2:d | 2.388 | 0.217 |
| Ni^2:d | 2.368 | 0.193 |
| Ni^4:d | 2.025 | 0.272 |
| In^3:s | 2.025 | 0.152 |
| Cu^3:d | 1.912 | 0.445 |
| Cd^2:s | 1.795 | 0.120 |
| Se^-2:p | 1.697 | 0.154 |
| Y^3:d | 1.696 | 0.117 |
| Te^-2:p | 1.537 | 0.191 |
| Mn^2:d | 1.459 | 0.129 |
| Ga^3:s | 1.226 | 0.127 |
| Cl^-1:p | 1.056 | 0.107 |
| Ru^2:d | 0.986 | 0.872 |
| Zn^2:s | 0.945 | 0.113 |
| Sn^4:s | 0.835 | 0.123 |
| S^-2:p | 0.822 | 0.096 |
| O^-2:p | 0.764 | 0.079 |
| Sb^5:s | 0.715 | 0.136 |
| F^-1:p | 0.600 | 0.091 |
| I^-1:p | 0.595 | 0.124 |
| Br^-1:p | 0.510 | 0.122 |
| Ge^4:s | 0.407 | 0.122 |
| Ag^1:s | 0.384 | 0.089 |
| Cu^1:d | 0.229 | 0.123 |
| As^-3:p | 0.209 | 0.203 |
| C^4:p | 0.198 | 0.120 |
| Ag^1:d | 0.000 | 0.000 |
| Mn^2:s | -0.080 | 0.131 |
| Si^4:s | -0.100 | 0.091 |
| O^-2:s | -0.108 | 0.079 |
| P^-3:p | -0.297 | 0.226 |
| N^-3:p | -0.380 | 0.101 |
| Sb^-3:p | -0.386 | 0.276 |
| B^3:p | -0.527 | 0.117 |
| Br^-1:s | -0.560 | 0.125 |
| Na^1:s | -0.595 | 0.093 |
| Al^3:s | -0.644 | 0.099 |
| Cl^-1:s | -0.683 | 0.107 |
| P^5:s | -0.830 | 0.092 |
| S^-2:s | -0.856 | 0.095 |
| Rb^1:s | -0.863 | 0.100 |
| N^-3:s | -0.988 | 0.104 |
| I^-1:s | -1.052 | 0.131 |
| Mg^2:s | -1.124 | 0.115 |
| F^-1:s | -1.193 | 0.096 |
| K^1:s | -1.212 | 0.101 |
| Cu^1:s | -1.301 | 0.133 |
| P^5:p | -1.370 | 0.094 |
| Cu^1:p | -1.936 | 0.143 |
| Li^1:s | -2.059 | 0.105 |
| Si^4:p | -2.184 | 0.103 |
| Ag^1:p | -2.433 | 0.129 |
| Al^3:p | -3.340 | 0.143 |

Table Full list of scores (abilities) and standard errors (s.e.) from the conduction band analysis