Quantifying uncertainty in high-throughput density functional theory: A comparison of AFLOW, Materials Project, and OQMD

Vinay I. Hegde, Christopher K. H. Borg, Zachary del Rosario, Yoolhee Kim, Maxwell Hutchinson, Erin Antono, Julia Ling, Paul Saxe, James E. Saal, and Bryce Meredig

Phys. Rev. Materials **7**, 053805 — Published 30 May 2023

DOI: 10.1103/PhysRevMaterials.7.053805

Quantifying uncertainty in high-throughput density functional theory: a comparison of AFLOW, Materials Project, and OQMD

Vinay I. Hegde,[1,*] Christopher K. H. Borg,[1,*] Zachary del Rosario,[1,2] Yoolhee Kim,[1] Maxwell Hutchinson,[1] Erin Antono,[1] Julia Ling,[1] Paul Saxe,[3] James E. Saal,[1] and Bryce Meredig[1,†]

[1] *Citrine Informatics, 2629 Broadway, Redwood City, CA 94063*
[2] *Olin College of Engineering, 1000 Olin Way, Needham, MA 02492*
[3] *Molecular Sciences Software Institute, Virginia Tech, Blacksburg, VA 24061*

(Dated: May 3, 2023)

A central challenge in high throughput density functional theory (HT-DFT) calculations is selecting a combination of input parameters and post-processing techniques that can be used across all materials classes, while also managing accuracy-cost tradeoffs. To investigate the effects of these parameter choices, we consolidate three large HT-DFT databases: Automatic-FLOW (AFLOW), the Materials Project (MP), and the Open Quantum Materials Database (OQMD), and compare reported properties across each pair of databases for materials calculated using the same initial crystal structure. We find that HT-DFT formation energies and volumes are generally more reproducible than band gaps and total magnetizations; for instance, a notable fraction of records disagree on whether a material is metallic (up to 7%) or magnetic (up to 15%). The variance between calculated properties is as high as 0.105 eV/atom (median relative absolute difference, or MRAD, of 6%) for formation energy, 0.65 Å^3/atom (MRAD of 4%) for volume, 0.21 eV (MRAD of 9%) for band gap, and 0.15 μ_B/formula unit (MRAD of 8%) for total magnetization, comparable to the differences between DFT and experiment. We trace some of the larger discrepancies to choices involving pseudopotentials, the DFT+U formalism, and elemental reference states, and argue that further standardization of HT-DFT would be beneficial to reproducibility.

Keywords: high-throughput DFT, uncertainty quantification, reproducibility, materials databases

I. INTRODUCTION

Over the past decade, high-throughput (HT) density functional theory (DFT) has emerged as a widely-used tool for materials discovery and design [1–3]. In a standard HT-DFT workflow, software tools automate the process of calculating materials properties of interest within DFT, including submitting jobs to high-performance computing infrastructure, on-the-fly error handling, post-processing and dissemination of results, and so on, enabling researchers to evaluate typically 10^3–10^6 materials with minimal human intervention. The resulting database can then be screened for candidate materials exhibiting promising combinations of calculated properties or to search for trends amongst materials behavior to gain new chemical insights or develop surrogate models.

The increasingly widespread usage of HT-DFT in materials research can be attributed to a combination of three key factors. First, a large number of specialized codes implement fully automated calculations of specific materials properties within DFT, ranging from phonon dispersions to dielectric tensors. For example, VASP 5.1 [4, 5] introduced a feature enabling users to calculate elastic tensors by simply setting a parameter in the input file. Second, the ongoing growth of computing power has ensured that HT-DFT is now well within reach of a single university research group. Third, sophisticated, free, often open-source, software is readily available for managing large numbers of DFT calculations, post-processing output, and storing the resulting data systematically in databases. Thus, a number of HT-DFT databases with various focus areas have emerged [3, 6–17]; a list of exemplars, including any supporting workflow automation software [18–33], is given in Section S-I of the Supplemental Material (SM) [34].

However, the entirely-automated nature of HT-DFT introduces a few key challenges. First, by definition, the volume of data from HT-DFT is too high for each individual calculation to undergo manual review or analysis [1]. How, then, are the quality and integrity of calculations monitored in high-throughput? Second, HT-DFT requires choosing, often at the outset, settings that are consistent across all calculations, encompassing all materials classes and properties being calculated. For example, it may not be known *a priori* whether the material being calculated is a metal or an insulator. As a result, the calculation parameters that affect, e.g., how electronic occupancies are smeared near the Fermi level must be chosen so that they are applicable to both metals and insulators. Third, practical HT-DFT calculations involve balancing accuracy and computational cost; best-practice recommendations [35] involve steps such as explicit convergence tests, which become computationally infeasible in the HT context. Of these challenges, only the first, related to monitoring the quality and integrity of calculations in high-throughput has been addressed. Software frameworks, such as Custodian [36], qmpy [23], and AiiDa [37], can store provenance information to en-

* These authors contributed equally to this work
† bryce@citrine.io

sure the integrity of calculations, and gracefully handle errors associated with catastrophic failures, e.g., those related to file read/write operations or memory issues during runtime, insufficient walltimes on high-performance computing resources, and misconfiguration of the underlying numerical libraries.

Since HT-DFT has become increasingly central to materials informatics efforts across the spectrum, from high-throughput screening to machine learning [38, 39] it is crucial to resolve the following concerns: (a) There is no one "correct" solution to some of the challenges of HT-DFT mentioned above, and different databases have tackled them slightly differently. How sensitive are the calculated materials properties to the different HT-DFT parameter choices? (b) The focus areas of many prominent HT-DFT databases in terms of the materials and properties calculated are often quite different. As a result, materials data from the various HT-DFT databases are often mixed with one another for thermochemical or other analysis. How interoperable are these various calculated materials properties across HT-DFT databases? We emphasize that such a comparison across HT-DFT databases is different from analyzing the reproducibility of DFT across software implementations and potentials, e.g. focusing on equations of state of elemental crystals: [40] the challenges of HT-DFT lie in choosing parameters that are applicable across a wide variety of materials and properties, targeting both reasonable accuracy *and* computational cost—very distinct from performing highly-accurate DFT calculations of a small set of materials.

Here, we analyze the reproducibility and interoperability of HT-DFT calculations. We critically compare the agreement between three databases for four properties: formation energy (ΔE_f), volume (V), band gap (E_g), and total magnetization (M). We find certain properties (formation energies and volumes) to be more consistent across databases than others (band gap and magnetization). We then quantify the variability in each of the properties across databases and find that the typical differences between two HT-DFT databases are similar to those between DFT and experiment. Finally, we compare properties across different materials classes to identify characteristics of materials and/or properties that are harder than others to reproduce. In all cases, we identify trends, surface outliers, and investigate potential causes for an observed systematic differences between the databases.

II. METHODS

We focus on three prominent HT-DFT databases in this work: Automatic FLOW (AFLOW) [6], the Materials Project (MP) [15], and the Open Quantum Materials Database (OQMD) [3, 23]. All three databases contain calculations of a large number of mostly-experimentally reported, ordered compounds from the Inorganic Crystal Structure Database (ICSD) [41]. In addition, they contain calculations of many thousands of hypothetical compounds generated from common structural prototypes or other informatics approaches. As noted earlier, there are many other large HT-DFT databases, e.g., JARVIS-DFT [13], Materials Cloud [14], and others listed in Table S-I of the SM [34]. Here, we limit our focus to AFLOW, Materials Project, and OQMD as the latter (a) are among the longest-running, mature, widely-used, and general-purpose, and (b) use the VASP software package [4, 5] and projector augmented wave (PAW) potentials [42, 43] with the Perdew-Burke-Ernzerhof (PBE) parameterization [44] of a generalized-gradient approximation (GGA) to the DFT exchange-correlation functional. The variance in HT-DFT-calculated properties studied in the present work is, therefore, almost entirely due to differences in various choices involved in HT-DFT (e.g., those involving calculation parameters such as k-point density, the DFT+U approach, post-calculation processing techniques, different versions of VASP and any associated software bugs, different versions of PBE pseudopotentials used) and *not* due to different implementations of DFT or approximations to the underlying exchange-correlation functional itself.

AFLOW has standardized band structure calculations [18, 45], binary alloy cluster expansions [46], finite-temperature thermodynamic properties [47], elastic and thermomechanical properties [48] calculated for many materials, and has an application programming interface (API) based on the REpresentational State Transfer (REST) standard (commonly referred to as "RESTful API") for accessing data [6, 49]. The Materials Project includes a variety of properties calculated for specific subsets of materials in the database, including elastic [50], thermoelectric [51], piezoelectric [52], dielectric [53], vibrational [54] properties, and X-ray adsorption spectra [55]. It also includes a collection of apps such as a Pourbaix diagram calculator [56], and the underlying data are accessible via a RESTful API [57, 58]. Finally, the Open Quantum Materials Database (OQMD) contains calculations of a large number of hypothetical compounds based on structural prototypes, [59–61] and provides tools for the construction of DFT ground state phase diagrams at ambient and high-pressures [62–64]. The OQMD provides the entirety of the underlying database to download all at once, and a RESTful API for programmatic access [65]. License and access information for the three databases is included in Section S-II of the SM [34].

We query all three databases (AFLOW: queried June 2021; MP: v2019.05; OQMD: v1.2) for the calculated properties of materials whose crystal structures were sourced from the ICSD and aggregate them into a single dataset, after converting records from all sources into a unified, consistent data format, the Physical Information File (PIF) [66, 67]. We then generate a set of comparable records for each pairwise combination of the databases—all calculations using the same initial crystal structure,

by matching their ICSD Collection Codes (hereafter referred to as "ICSD ID"). In instances where more than one calculation within a single database was labeled with the same ICSD ID, we use the lowest energy calculation for all analysis. In addition, we discard records with obviously unphysical property values (those with formation energy outside the [−5 eV/atom, +5 eV/atom] window and volumes above 150 Å^3/atom), and normalize properties to the same units, where required. We then perform statistical analysis on the final curated set of comparable records across the three databases. Definitions of the metrics used in our analysis are given in Appendix A and details of the query and curation steps are provided in Section S-II of the SM [34].

III. RESULTS

The aggregation and processing of the data from the three HT-DFT databases results in a set of ∼70,000 total comparable DFT calculations. For each property of interest, i.e., formation energy per atom, volume per atom, band gap, total magnetization per formula unit (f.u.), the counts of records, and overlapping records for each pair of databases are shown in Table I. Approximately 15,000–25,000 comparisons can be made for each property and database pair, except for comparisons to formation energies from AFLOW, where only ∼2,200 records are reported. As mentioned earlier, overlapping records across databases were determined by using exact ICSD ID matches for the reported calculations.

A. Overall pairwise comparison statistics

Table II shows some overall statistics for comparisons of all properties across comparable records in the three databases: the median absolute difference (MAD), the interquartile range (IQR), the Pearson correlation coefficient (r), and Spearman's rank correlation coefficient (ρ) (definitions of the metrics are in Appendix A). For band gap and total magnetization, the statistics were calculated only on subsets of overlapping records where both databases agreed that a material is non-metallic ($E_g > 0.01$ eV) and is magnetic ($M > 0.01$ μ_B/atom), respectively. The latter threshold on the per-formula unit total magnetization ensures that undesired comparisons of different magnetic configurations for the same crystal structure (i.e., ferromagnetic configuration in one database being compared to antiferromagnetic configuration in another) are avoided as much as possible.

Overall, we find that: (a) The MAD in formation energy across pairs of databases can be up to 0.105 eV/atom, comparable to the ∼0.1 eV/atom difference between DFT and experimental formation energies [23]. (b) The MAD in volume across pairs of databases can be up to 0.65 Å^3/atom (median absolute difference relative to mean (MRAD), of 3.8%), comparable to error between DFT and experiment [68]. (c) The MAD in band gap across pairs of databases can be up to 0.21 eV, even when comparing only records where both databases agree that a material is not metallic. For around 5%–7% of overlapping records, databases disagree whether a material is metallic. (d) The comparison of total magnetization shows high variability across database pairs. While the dispersion of differences for the MP-OQMD comparison is very small (MAD of 0.01 μ_B/f.u. and IQR of 0.05 μ_B/f.u.), the dispersion of differences in comparisons with AFLOW are rather large (up to MAD of 0.15 μ_B/f.u. and IQR of up to 2.0 μ_B/f.u.). In all cases, the correlation between calculated values is lower than for the other three properties, with both Pearson and Spearman correlation coefficients ranging from 0.6–0.8. We further note that the latter poor correlation exists even after excluding overlapping records where the two databases disagree on whether the material is magnetic (10%–15% of the records).

B. Distribution of differences in calculated properties

We first analyze the raw differences in the calculated properties for records overlapping across pairs of databases. Figure 1 shows the distribution of the differences in calculated values for each of formation energy, volume, band gap, and total magnetization, for each pairwise combination of databases.

Formation energy: The distribution of differences in calculated formation energy across AFLOW-MP and MP-OQMD is surprisingly bimodal, with peaks around 0 and ±0.2 eV/atom. We find that the peak near 0.2 eV/atom in both pairwise comparisons corresponds mostly to oxides (see Figure S1), and is a result of different approaches in the two databases toward correcting DFT-calculated formation energies (see Section IV B). While the median difference ($\widetilde{\Delta x}$ in Figure 1) are reasonably small across all three pairwise comparisons (up to ∼0.074 eV/atom), the difference distributions for AFLOW-MP and MP-OQMD are rather wide. The median absolute difference (MAD) and the interquartile range (IQR), both robust measures of the spread of a distribution, are up to ∼0.105 eV/atom and ∼0.173 eV/atom, respectively.

Volume: The distribution of differences in calculated volumes is skewed towards smaller volumes in the OQMD, but such a skew is absent in the AFLOW-MP comparison. Correspondingly, the median difference between AFLOW and MP volumes are ∼0.01 Å^3/atom, whereas the median differences are ∼0.62 Å^3/atom and ∼0.47 Å^3/atom for AFLOW-OQMD and MP-OQMD, respectively. The consistently smaller volumes calculated in the OQMD can be understood to result from the choice of the plane wave energy cutoff used for DFT relaxation calculations. The OQMD chooses a plane wave cutoff that is lower than that used in AFLOW and MP (ENMAX in the POTCAR file, up to 400 eV in OQMD, as opposed

	AFLOW	MP	OQMD	AFLOW-MP	AFLOW-OQMD	MP-OQMD
Formation Energy	2196	34907	22248	2070	1717	19082
Volume	21929	34907	22248	19258	15857	19082
Band Gap	21921	34907	22169	19253	15790	19007
Total Magnetization	21929	34907	22248	19258	15857	19082

TABLE I. The number of records after establishing ICSD ID equivalency for each property of interest in the AFLOW, Materials Project (MP), and OQMD HT-DFT databases, as well as for pairwise comparisons of the three databases.

	AFLOW-MP				AFLOW-OQMD				MP-OQMD			
	MAD	IQR	r	ρ	MAD	IQR	r	ρ	MAD	IQR	r	ρ
Formation Energy (eV/atom)	0.105	0.173	0.99	0.99	0.019	0.036	0.99	0.99	0.087	0.168	0.99	0.99
Volume (Å^3/atom)	0.180	0.389	0.98	0.99	0.647	1.117	0.97	0.97	0.512	0.902	0.98	0.98
Band Gap (eV)*	0.078	0.203	0.94	0.92	0.209	0.364	0.92	0.91	0.178	0.277	0.93	0.92
Total Magnetization (μ_B/f.u.)*	0.015	0.759	0.77	0.75	0.149	2.001	0.60	0.56	0.012	0.052	0.80	0.74

TABLE II. Overall statistics (median absolute difference (MAD), interquartile range (IQR), Pearson's linear correlation coefficient (r), and Spearman's rank correlation coefficient (ρ)) for the comparison of properties across HT-DFT databases. For each property, records overlapping across a pair of databases are compared (* for band gap and magnetization, only non-zero values are compared). Generally, lower MAD, lower IQR, higher r, and higher ρ values indicate better reproducibility of calculated properties.

to 520 eV in MP and up to 560 eV in AFLOW) for full cell relaxations. The lower plane wave cutoff results in Pulay stresses and generally smaller volumes than fully relaxed calculations. The MAD in volumes for comparisons, especially for OQMD with the other two databases, is up to ∼0.65 Å^3/atom. In addition, some differences in reported volumes can result from the different relaxation schemes employed in the three HT-DFT databases: AFLOW and MP perform two sequential relaxations, while the OQMD performs sequential relaxations until the volume change during a relaxation is less than 5%.

Band gap: The distribution of differences in the calculated band gaps is slightly skewed towards larger band gaps in the OQMD, but this skew is absent in the AFLOW-MP comparison. Correspondingly, the median difference in band gaps between AFLOW and MP is ∼0.01 eV, and up to ∼0.14 eV for comparisons with OQMD. The larger band gaps calculated in the OQMD might be due to smaller volumes from the choice of lower plane wave energy cutoffs. An increase in the fundamental band gap due to compressive strains (in the OQMD, due to unresolved Pulay stresses) has been observed in many semiconductor families [69–71]. In addition, the spread in the differences in calculated band gaps is quite large: with an MAD of up to ∼0.21 eV and an IQR of up to ∼0.36 eV for comparisons with OQMD. The spread may be, in addition to the choice of energy cutoff as discussed above, due to the different ways in which the databases calculate the band gap. For example, OQMD calculates band gap from the electronic density of states (DOS), in contrast to AFLOW and MP which calculate it from band dispersions. The energy grid used for the calculation of DOS and/or k-point meshes used for band structure calculations can also have a notable effect on the precision and accuracy of the reported band gap. For instance, while AFLOW and MP both report gaps calculated from band dispersion calculations, the high-symmetry k-path in the Brillouin zone used for such calculations can be different [18, 72].

Total magnetization: The median differences in AFLOW-MP and MP-OQMD are nearly zero, with reasonably small MAD values as well. However, the differences between the magnetization reported in AFLOW and the other two databases skew towards larger values in AFLOW, with long tails and correspondingly large dispersions. The difference between AFLOW and OQMD, in particular, shows an MAD of ∼0.15 μ_B/atom and an IQR of ∼2.0 μ_B/atom. Further, as noted earlier, a significant fraction of 10–15% overlapping records across databases disagree on whether the material has non-zero total magnetization. This disagreement may in part be due to different pseudopotential choices for various elements (and correspondingly different number of valence electrons), and sampling of different magnetic configurations, the choice of unit cell in such magnetic configuration sampling, etc. For instance, AFLOW and MP calculate ferromagnetic configurations for all materials, and ferrimagnetic and antiferromagnetic configurations for a subset of materials [73, 74], while the OQMD only calculates ferromagnetic configurations [23]. For a given material, since we only compare the lowest-energy configurations across databases with one another, it is possible that a material is predicted to be non-magnetic in one database and antiferromagnetic in another database. Alternately, a ferrimagnetic configuration in one database could be compared to a ferromagnetic calculation in another, if both converged to finite magnetic moments.

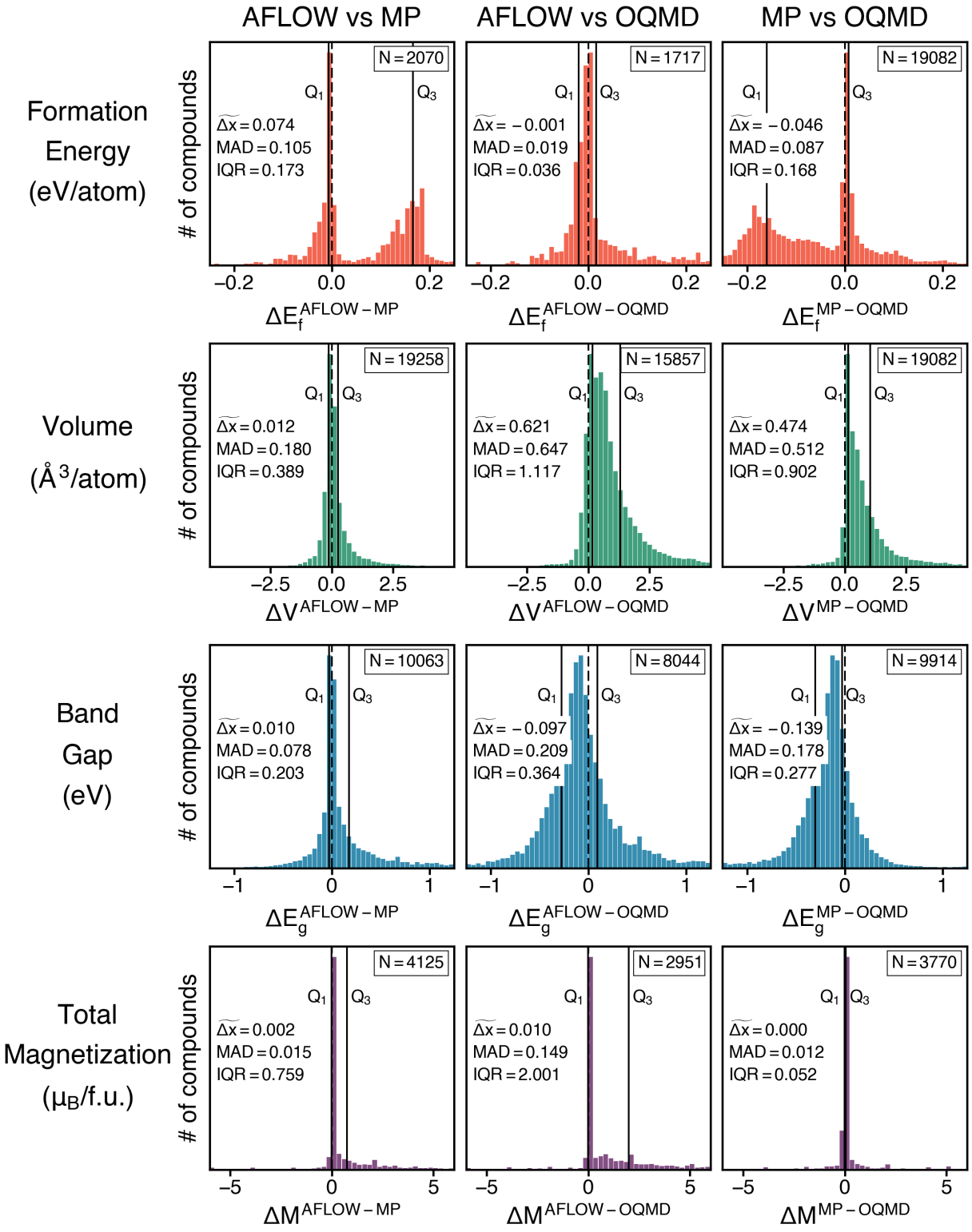

FIG. 1. Distribution of the differences in calculated properties across HT-DFT databases. Each panel corresponds to a property and pair of databases being compared. Solid vertical black lines correspond to the first (Q_1) and third (Q_3) quartiles of the distribution. The number of records overlapping across the two databases is shown in the top right corner of each panel; the median of distribution ($\widetilde{\Delta x}$), the median absolute difference (MAD), and the interquartile range (IQR) are noted on the left.

C. Rank-order comparisons across properties

We next seek to make comparisons *across* properties. Instead of comparing the raw values of the properties directly, we compare overlapping records using the ordinal

rank of the property in each database being compared (hereafter, referred to as "percentile rank"). Comparing the percentile ranks of the properties has a few advantages: (a) It allows for a single consistent metric for comparison across all four properties regardless of the magnitude of the actual value and physical units. (b) It is not affected by many systematic differences, e.g., a constant shift of 0.1 eV in all calculated band gaps in one database. Such constant shifts in calculated properties do not affect the internal consistency of a HT-DFT database, and the percentile ranks which are similarly unaffected capture this property. (c) It is a robust, uniform, identifier of outliers in calculated properties.

Figure 2 consists of percentile rank scatterplots (closely related to the quantile-quantile or Q-Q plots) of each property of interest for each database pair. Note that for band gap (total magnetization), we only include overlapping records where the two databases being compared both report the material to be non-metallic (magnetic), to avoid having to rank near-zero or zero values against one another. A compact line along the diagonal corresponds to perfect correlation between the ranked properties, with more diffuse scattering indicating lower levels of correlation.

Formation energy: Of the four properties, formation energy shows the best correlation between each database pair, consistent with all r and ρ values close to 0.99 in Table II. Nonetheless, there is some off-diagonal scatter for the MP-OQMD comparison for larger (more positive) values of formation energy that is not found in the other database pairs. These calculations correspond to compounds with smaller (positive) formation energies, where the precision necessary to reliably rank the structure approaches the accuracy of the calculation.

Volume: The percentile rank comparison of volume shows higher off-diagonal scatter than that seen in comparisons of formation energy. There is a skew towards higher volumes in AFLOW and MP when compared to OQMD (scatter towards top-left of the diagonal in the AFLOW-OQMD and MP-OQMD comparisons), consistent with the discussion around plane wave energy cutoffs in the previous section.

Band gap: The percentile rank comparison of band gap shows even higher off-diagonal scatter than that observed in comparisons of both formation energy and volume. In particular, there is meaningful scatter *along the axes*, corresponding to cases where one database predicts the material to have a near-zero band gap whereas the other database predicts a (much larger) non-zero band gap.

Total magnetization: The percentile rank comparison of total magnetization per formula unit in all three pairwise comparisons shows a few distinct clusters along the diagonal, corresponding to nominally integer values of magnetic moment per formula unit. There is considerable off-diagonal "bowing" in the comparisons with AFLOW, consistent with the distribution of differences between AFLOW and the other two databases showing a skew towards larger magnetizations in AFLOW *and* long tails (lower panel in Figure 1). In addition, there is considerable off-diagonal scatter (horizontal and vertical bands in the magnetization panel of Figure 2) indicating significant disagreement between the values reported in the two databases.

Overall, a comparison of rank-ordered properties across two databases shows that formation energies and volumes are more easily reproduced than band gaps and total magnetizations, consistent with correlation coefficients decreasing from ∼0.99 for formation energy to ∼0.6 for total magnetization (Table II).

D. Reproducibility across materials classes

Intuitively, we expect the level of agreement among the databases to be a strong function of materials class. Therefore, we compare specific subsets of calculations based on various materials classes to elucidate potential causes of differences. The materials classes are defined based on chemical composition, the number of elemental components, the presence of magnetism, band gap, pseudopotential choices, and space group, as summarized in Table III. For classes defined by the output of a calculation (i.e., those based on magnetization and band gap), comparisons are only made if both databases agree that the property has a non-zero value. Note that according to our definition, the "Magnetic" class of materials may potentially include both ferromagnetic and ferrimagnetic materials, and the "Non-Magnetic" class may potentially include both non-magnetic and antiferromagnetic materials.

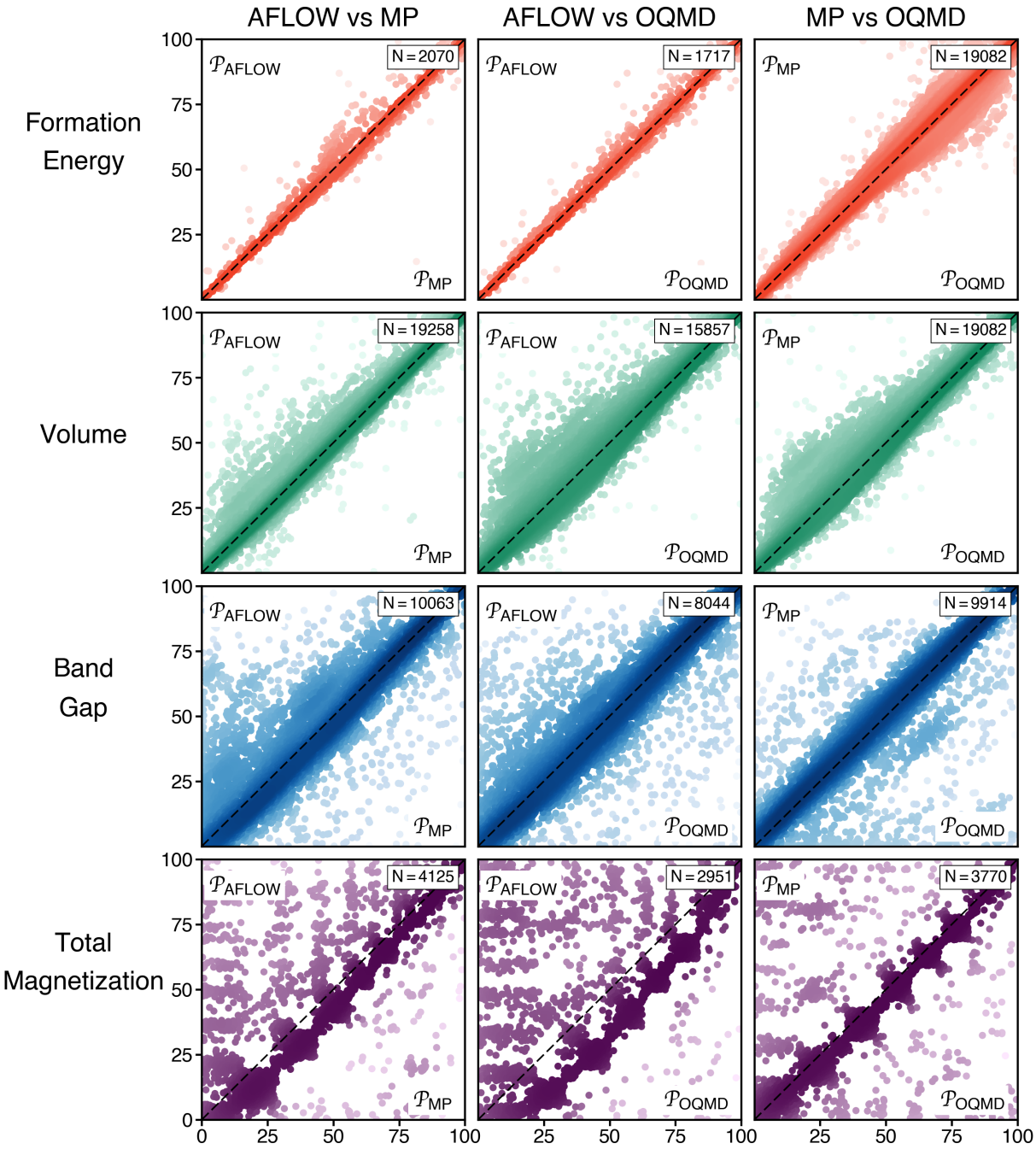

FIG. 2. Comparison of the calculated properties (formation energy, volume, band gap, and total magnetization) over records overlapping across pairwise combinations of HT-DFT databases plotted as a percentile rank (i.e., ordinal rank of the property in each database being compared). A compact line along the diagonal corresponds to perfect correlation between the ranked properties. Overall, formation energies and volumes show better reproducibility than band gaps and magnetizations. The clusters seen in the magnetization comparisons correspond to nominally integer values of magnetic moments.

Class	**Definition**
Oxide	Contains O
Nitride	Contains N
Pnictide	Contains a group 15 element
Chalcogenide	Contains a group 16 element, except O
Halide	Contains a group 17 element
Alkali Metal	Contains a group 1 element, except H
Alkaline Earth Metal	Contains a group 2 element
Transition Metal	Contains a *d*-block element
Metalloid	Contains B, Si, Ge, As, Sb, or Te
Rare-Earth	Contains an element from the lanthanide series
Actinide	Contains an element from the actinide series
Metal-Nonmetal	Contains at least one metal element *and* at least one of C, N, O, F, P, S, Cl, Se, Br, I
Intermetallic	Contains only metallic elements
Magnetic	Both databases report a net magnetic moment $> 10^{-2}$ μ_{B}/f.u.
Non-magnetic	Both databases report no net magnetic moment $> 10^{-2}$ μ_{B}/f.u.
Disagree on Magnetic	The two databases disagree on whether a net magnetic moment $> 10^{-2}$ μ_{B}/f.u. is present
Metallic	Both databases predict a band gap of $< 10^{-2}$ eV
Semiconductor	Both databases predict a band gap between 10^{-2} and 1.5 eV
Insulator	Both databases predict a band gap larger than 1.5 eV
Disagree on Metallic	The two databases disagree on whether a band gap $< 10^{-2}$ eV is present
Pseudopotentials Agree	Both databases use the same set of pseudopotentials for all elements
Pseudopotentials Disagree	The databases use different pseudopotentials for at least one element
Use GGA+U	Both databases use the GGA+U approach
Use GGA	Both databases use plain GGA
Disagree on GGA/GGA+U	One database uses GGA whereas the other uses GGA+U
Elements	Contains only one element
Binaries	Contains two elements
Ternaries	Contains three elements
Quaternaries	Contains four elements
Triclinic	Space group 1–2
Monoclinic	Space group 3–15
Orthorhombic	Space group 16–74
Tetragonal	Space group 75–142
Trigonal	Space group 143–167
Hexagonal	Space group 168–194
Cubic	Space group 195–230

TABLE III. Definitions for the materials classes used in this work.

	AFLOW vs MP				AFLOW vs OQMD				MP vs OQMD			
	ΔE_f	V	E_g	M	ΔE_f	V	E_g	M	ΔE_f	V	E_g	M
All	5.8 (2070)	1.0 (19258)	3.8 (10063)	0.5 (4125)	1.3 (1717)	3.8 (15857)	8.9 (8044)	7.6 (2951)	6.3 (19082)	2.9 (19082)	8.0 (9914)	0.5 (3770)
Oxide	6.3 (989)	0.6 (6468)	3.3 (5289)	0.1 (1694)	1.0 (818)	5.8 (5269)	8.2 (4208)	0.1 (1159)	6.8 (6616)	5.7 (6616)	8.1 (5300)	0.0 (1601)
Nitride	0 (0)	1.2 (1639)	2.1 (1193)	0.3 (218)	0 (0)	9.3 (1257)	8.2 (936)	16.4 (149)	14.5 (1422)	6.9 (1422)	7.0 (1068)	0.6 (199)
Pnictide	6.4 (727)	0.9 (5470)	2.7 (3520)	0.3 (919)	2.3 (611)	4.5 (4330)	9.0 (2735)	2.5 (594)	8.2 (5210)	3.6 (5210)	8.1 (3385)	0.4 (805)
Chalcogenide	4.2 (412)	1.0 (3985)	4.8 (2982)	0.1 (714)	1.5 (340)	3.3 (3154)	8.6 (2328)	0.8 (470)	10.6 (3951)	2.3 (3951)	7.4 (2924)	0.2 (623)
Halide	14.1 (298)	1.0 (3850)	2.8 (3291)	0.0 (891)	10.7 (245)	8.7 (3107)	8.1 (2565)	0.1 (567)	7.8 (3700)	8.1 (3700)	6.7 (3029)	0.0 (726)
Alkali Metal	5.5 (695)	0.7 (4541)	2.9 (3605)	0.0 (897)	1.3 (599)	5.2 (3887)	9.4 (3019)	0.1 (618)	6.6 (4961)	5.0 (4961)	8.4 (3904)	0.0 (860)
Alkaline Earth Metal	4.8 (709)	0.5 (3448)	1.7 (2002)	0.1 (621)	0.8 (580)	3.1 (2929)	8.4 (1614)	0.3 (466)	5.7 (3398)	3.2 (3398)	8.1 (1960)	0.1 (575)
Transition Metal	4.9 (393)	1.1 (13028)	10.1 (5342)	0.8 (3595)	1.0 (313)	3.4 (10669)	10.8 (4176)	7.7 (2773)	5.7 (12500)	2.2 (12500)	8.4 (5025)	0.4 (3447)
Metalloid	5.7 (1063)	0.8 (6004)	2.0 (2965)	3.8 (979)	1.0 (856)	2.5 (4926)	8.6 (2330)	42.8 (637)	4.5 (5933)	1.6 (5933)	8.4 (2879)	1.8 (768)
Rare-Earth	3.0 (63)	1.1 (4630)	9.1 (1236)	3.9 (1229)	0.6 (48)	2.0 (3974)	9.9 (1038)	121.1 (651)	4.8 (5511)	1.3 (5511)	8.5 (1610)	12.0 (977)
Actinide	9.7 (4)	3.2 (670)	36.4 (222)	8.5 (300)	4.7 (4)	5.5 (569)	9.9 (187)	10.7 (272)	6.3 (758)	1.3 (758)	43.0 (233)	1.8 (438)
Metal-Nonmetal	5.8 (1246)	0.8 (12270)	4.5 (8717)	0.1 (3053)	1.2 (1035)	4.9 (10020)	8.9 (7004)	0.1 (2129)	7.4 (12374)	4.4 (12374)	7.9 (8690)	0.1 (2829)
Intermetallic	4.7 (182)	1.4 (3056)	20.4 (50)	41.2 (528)	2.7 (157)	1.6 (2582)	36.3 (35)	81.1 (404)	3.4 (2910)	1.2 (2910)	42.7 (54)	12.4 (446)
Magnetic	5.5 (12)	1.4 (4125)	27.4 (1580)	0.5 (4125)	13.0 (5)	5.6 (2951)	24.4 (1013)	7.6 (2951)	6.4 (3770)	3.4 (3770)	13.5 (1330)	0.5 (3770)
Non-Magnetic	5.8 (1991)	0.8 (13119)	2.3 (8106)	0 (0)	1.3 (1680)	3.2 (10520)	7.9 (6522)	0 (0)	6.2 (12922)	2.8 (12922)	7.3 (8146)	0 (0)
Disagree On Magnetic	5.1 (67)	1.3 (2014)	17.0 (377)	0 (0)	1.8 (32)	2.8 (2386)	18.9 (509)	0 (0)	6.7 (2390)	2.6 (2390)	53.1 (438)	0 (0)
Metallic	2.4 (498)	1.3 (7951)	0 (0)	32.0 (1888)	1.6 (436)	1.7 (6765)	0 (0)	67.3 (1508)	3.7 (8092)	1.0 (8092)	0 (0)	5.1 (2066)
Semiconductor	4.6 (472)	0.7 (2456)	11.9 (2456)	0.1 (394)	2.6 (328)	2.3 (1743)	17.9 (1743)	0.1 (231)	7.7 (2400)	2.3 (2400)	16.5 (2400)	0.0 (445)
Insulator	6.0 (1032)	0.7 (6785)	1.8 (6785)	0.0 (751)	1.0 (855)	6.4 (5577)	6.9 (5577)	0.0 (500)	7.1 (6791)	6.2 (6791)	6.4 (6791)	0.0 (656)
Disagree On Metallic	15.8 (47)	1.5 (1239)	0 (0)	0.3 (656)	39.2 (32)	5.9 (981)	0 (0)	0.4 (419)	7.1 (1001)	3.9 (1001)	0 (0)	0.1 (361)
Pseudopotentials Agree	5.7 (1547)	0.9 (11410)	3.3 (6621)	0.2 (2571)	1.5 (1139)	4.4 (5635)	7.3 (3793)	2.4 (453)	6.6 (10616)	3.3 (10616)	7.5 (6604)	1.6 (980)
Pseudopotentials Disagree	5.8 (523)	1.1 (7848)	4.9 (3442)	4.3 (1554)	1.0 (578)	3.5 (10222)	10.6 (4251)	9.1 (2498)	5.9 (8466)	2.5 (8466)	9.2 (3310)	0.3 (2790)
Use GGA+U	0 (0)	0.5 (1970)	8.1 (1553)	0.0 (1220)	0 (0)	5.7 (1573)	10.3 (1071)	0.1 (890)	5.2 (1419)	5.8 (1419)	8.1 (1142)	0.0 (1101)
Use GGA	5.8 (2045)	0.4 (4918)	0.6 (4031)	0.5 (89)	1.3 (1717)	5.2 (4026)	7.3 (3255)	0.4 (48)	6.5 (16353)	2.2 (16353)	7.7 (8003)	2.9 (2282)
Disagree on GGA/GGA+U	0 (0)	1.3 (12176)	12.8 (4392)	8.7 (2759)	0 (0)	2.9 (10258)	11.8 (3718)	39.1 (2013)	6.0 (1120)	6.9 (1120)	24.8 (677)	0.0 (326)
Element	8.1 (91)	1.5 (159)	2.4 (45)	12.4 (8)	5.3 (76)	2.1 (149)	7.9 (33)	27.6 (5)	9.9 (152)	2.3 (152)	6.4 (42)	4.5 (6)
Binary	5.8 (648)	1.4 (3352)	3.0 (975)	17.7 (492)	1.8 (542)	2.6 (2698)	8.6 (790)	46.5 (377)	6.9 (2934)	1.7 (2934)	7.2 (877)	5.1 (438)
Ternary	5.5 (1003)	1.0 (10319)	4.0 (4526)	4.3 (2229)	1.3 (841)	2.9 (8706)	9.7 (3720)	25.4 (1665)	5.5 (10423)	1.8 (10423)	8.2 (4443)	1.3 (2025)
Quaternary	5.8 (308)	0.7 (4270)	3.9 (3530)	0.1 (1130)	0.6 (250)	5.4 (3438)	8.3 (2786)	0.1 (725)	6.7 (4497)	4.9 (4497)	8.0 (3630)	0.1 (1022)
Triclinic	6.0 (104)	0.9 (1003)	3.9 (902)	0.1 (216)	1.2 (89)	8.6 (805)	7.6 (714)	0.1 (132)	8.3 (1052)	7.5 (1052)	7.4 (930)	0.0 (207)
Monoclinic	6.0 (446)	0.9 (3691)	3.8 (2991)	0.1 (764)	1.7 (373)	6.1 (2991)	8.5 (2340)	0.1 (491)	7.5 (4052)	5.3 (4052)	7.6 (3178)	0.0 (738)
Orthorhombic	5.8 (507)	0.9 (4550)	3.7 (2552)	0.3 (746)	1.3 (405)	3.2 (3625)	8.8 (1953)	2.2 (522)	5.7 (4634)	2.5 (4634)	8.0 (2509)	0.4 (724)
Tetragonal	5.8 (246)	1.0 (2797)	5.0 (1042)	5.9 (641)	1.2 (202)	3.0 (2369)	10.0 (861)	38.2 (496)	5.1 (2762)	1.9 (2762)	8.8 (980)	2.4 (572)
Trigonal	5.4 (241)	0.9 (1746)	3.5 (1144)	0.1 (389)	0.9 (199)	3.9 (1404)	9.7 (896)	0.1 (285)	6.8 (1507)	3.5 (1507)	8.0 (947)	0.1 (326)
Hexagonal	4.7 (196)	1.0 (2108)	3.3 (545)	11.3 (427)	0.8 (158)	1.9 (1626)	8.8 (438)	49.8 (231)	4.2 (1798)	1.2 (1798)	9.0 (440)	3.3 (339)
Cubic	5.7 (295)	1.1 (2988)	2.9 (752)	20.7 (795)	1.5 (217)	2.6 (2203)	10.6 (527)	38.6 (541)	5.4 (2343)	1.4 (2343)	9.2 (600)	2.8 (557)

FIG. 3. Median percent absolute differences between properties (formation energy, volume, band gap, total magnetization) calculated in the three databases (AFLOW, MP, OQMD), compared two at a time, across various classes of materials as defined in Table III. The numbers in parentheses indicate the number of overlapping records belonging to the respective material class for a given pair of databases. Trivial comparisons are left blank (e.g., the difference in total magnetization for non-magnetic compounds).

Figure 3 contains the median absolute difference relative to the mean (MRAD) values for pairwise comparisons between databases, divided into materials classes as defined in Table III. Cells are colored based on the MRAD value listed. Empty cells correspond to trivial comparisons (e.g., values of band gap where both database agree the structure is metallic). We use MRAD as the metric here to reduce the effect of outliers (as compared to calculating means) as well as to enable comparisons across properties using the same metric. Overall, HT-DFT volumes show the best agreement (lowest MRAD values), from 1–4%. Band gaps show the worst overall agreement (highest MRAD values), 4–10% across all pairwise comparisons. Formation energy comparisons with MP show MRAD values up to 6%, but the AFLOW-OQMD MRAD is only 1.3%. MRAD values for total magnetization vary highly from 0.5% for comparisons with MP to 7.6% for AFLOW-OQMD. In all cases, certain materials classes have distinctly higher or lower MRAD when compared to the MRAD averaged over all materials classes.

Formation Energy: In the comparisons with AFLOW, two materials classes, "Halides" and "Disagree on Metallic", show the highest MRAD values of up to 14% and 40%, respectively. The high MRAD in halide formation energies can be understood to result from post hoc corrections to the effective elemental reference energies performed in MP and OQMD, but not in AFLOW, for the halide group of elements (see discussion in Section IV B). The high MRAD of the "Disagree on Metallic" class is likely an artifact of the small formation energies of the few records (∼30–50) in the comparison. As noted earlier, since AFLOW reports notably fewer formation energy values than the other databases, the comparisons are made with a much smaller set of records (∼2,000). Therefore, we ignore here some of the MRAD outliers in cases where the number of records being compared is very small (e.g., the material class "Magnetic" shows an MRAD of 13% between AFLOW and MP but there are only 5 records in the comparison). Further, the formation energies dataset has very few transition metal, rare-earth, and actinide element-containing compounds (Figures S3 and S7). New, different insights are likely to result from a larger dataset. In the MP-OQMD comparison, with a much larger comparable dataset (∼19,000), the "Nitride", "Pnictide", and "Chalcogenide" material classes show the highest MRAD values, 14%, 8%, and 11% respectively. This is partly due to differences in fitted elemental chemical potentials for pnictogen and chalcogen elements in MP and OQMD (Section IV B).

Volume: The best agreement is observed in the AFLOW-MP comparisons, with only the "Actinide" material class showing an MRAD greater than 2%. For comparisons with OQMD, the MRAD in volume is generally higher—due to the choice of lower plane wave energy cutoff used for cell relaxation, as discussed earlier (Section III B). The highest MRAD values in the comparisons with OQMD volumes are for the "Nitride" and "Halide" classes (∼7–9%). The default plane wave energy cutoffs in the VASP PAW potentials (ENMAX parameter) for N and F are among the highest (400 eV) of all elements. Thus, the lower energy cutoff used by OQMD for relaxation impacts the calculated volumes of nitrides and fluorides the most (Figures S8 and S12). Another material class, "Triclinic", shows similarly high MRAD values of ∼8% in comparisons with OQMD. Upon examination, we find that most triclinic materials in the comparisons are oxides, nitrides, and halides, and thus the high MRAD values are due to the chemical composition of these compounds rather than their crystal symmetry.

Band gap: While band gap comparisons show the highest MRAD values across properties, some materials classes in particular show MRAD values much greater than ∼10%. Of these, in the "Intermetallic" and "Semiconductor" material classes, the MRAD values are expectedly high due to small average band gaps relative to which differences are reported, even though the absolute differences themselves are not conspicuously large (Figure S2). In other cases, the high MRAD values are a result of (a) different pseudopotential choices for elements (e.g., Cu/Cu_pv, Ce/Ce_3, Eu/Eu_2 choices in the "Disagree on Magnetic" class for the MP-OQMD comparison with an MRAD of ∼53%; see Figure S13), (b) disagreement on whether to use the GGA or GGA+U approach to calculate properties (e.g., the "Actinide" material class with MRAD of up to 43% in comparisons with MP, the "Disagree on GGA/GGA+U" class in all three comparisons with MRAD of 12–25%), or a combination of both factors (e.g., for the "Magnetic" material class with an MRAD of up to 27% in comparisons with AFLOW), (c) non-overlapping sampling of magnetic configurations across databases. For instance, the "Magnetic" (MRAD of 13–27% across comparisons) and "Disagree on Magnetic" (MRAD of 17–53% across comparisons) classes may respectively include comparing ferromagnetic vs ferrimagnetic and non-magnetic vs antiferromagnetic ground states across two databases (note, however, that both the "Magnetic" and "Disagree On Magnetic" comparisons also include effects from other HT-DFT choices, such as choice of pseudopotential used). Note also that the errors in band gaps for the "Use GGA+U" materials class are larger than those for the "Use GGA" materials class across all three pairwise comparisons, the choice of slightly different effective U values used in the three databases being a likely contributor. Further discussions of some of the above parameter choices are in Section IV.

Total magnetization: While MRAD values in the MP-OQMD comparison are generally small ($< 5\%$), some material classes show much higher MRAD values, especially in comparisons with AFLOW. As in the case of band gap values, we find these comparisons to be influenced by pseudopotential choice (of rare-earth elements in particular, e.g., Nd, Nd_3, Nd_3 in AFLOW, MP, and OQMD, respectively; see Figures S10 and S14), choice of using GGA or GGA+U (e.g., MRAD of up to ∼40%

in AFLOW-OQMD comparisons for the "Disagree on GGA/GGA+U" class), or both (e.g., the "Metalloid" and "Rare-Earth" material classes in the AFLOW-OQMD comparisons, "Intermetallic" and "Metallic" classes in the AFLOW-MP and AFLOW-OQMD comparisons). We note that some other material classes show high MRAD values, e.g., "Element", "Binary", "Ternary", "Tetragonal", "Hexagonal", and "Cubic" (up to MRAD values up to ∼50%) due to, upon further examination, the parameter choices discussed above rather than due to number of components in the compound or crystal symmetry.

Finally, we note that while our scheme of constructing a set of comparable records across pairs of databases (by matching ICSD IDs exactly) ensures comparisons between the same initial crystal structures, it excludes a number of experimentally well-studied materials with multiple ICSD entries associated with them. We investigated whether this "bias away from well-studied materials" affects our results by using a larger comparison set constructed by linking very similar ICSD entries using the crystal structure matching algorithm employed by the Materials Project (see Section S-II in the SM [34]). While some of the quantitative metrics we report varied by a few percent in the expanded comparison, the overall conclusions remain unchanged (see Tables S-XI, S-XII, and Figures S15–S18 in the SM [34]), consistent with recent findings [75].

IV. DISCUSSION

We discuss some of the most important factors affecting the differences across HT-DFT calculations of properties below. Some of the other factors that either have a minor effect (e.g., *post hoc* calculation of band gap from band dispersions or density of states) or are specific to a database/property (e.g., plane wave cutoff energy for full cell relaxations in OQMD) have been discussed in the earlier sections.

A. Effects of pseudopotential choice

For nearly all elements, VASP provides multiple PAW potentials to choose from, with different numbers of electrons in the valence. The choice of pseudopotential varies across the HT-DFT databases due to factors such as changes in VASP recommendations and issues of calculation convergence or reproduction of experimental thermochemical data [76, 77]. Interestingly, the choice of pseudopotential has minimal effect on the calculated formation energies and volumes (up to a difference of 1% in cases where pseudopotentials do or do not match; see rows "Pseudopotentials Agree" and "Pseudopotentials Disagree" in Figure 3). On the other hand, the number of valence electrons and consequently the choice of pseudopotential affects the calculated band gaps and magnetization values severely. Especially egregious differences across those properties in material classes such as "Rare-Earth" and "Magnetic" (Figure 3) can be directly traced to different pseudopotential choices. For rare-earth and actinide elements in particular, with f-electrons that are poorly described by DFT [78], using pseudopotentials that treat f-electrons in core or valence can have a significant impact on the calculated band gap (e.g., "Intermetallic" and "Magnetic" classes in Figure 3) and magnetization (e.g., "Rare-Earth" and "Intermetallic" classes in Figure 3) values.

B. Elemental references and energy corrections

The largest disagreements in HT-DFT formation energies can be understood to result from different elemental reference states and/or post-calculation energy corrections performed in the databases. To our knowledge, the formation energies reported in AFLOW use DFT total energies of the bulk elements as the reference states [79]. MP and OQMD both correct DFT-calculated energies to closely reproduce experimental formation enthalpy data. While MP adds corrections to the compound formation energies [76, 77], OQMD fits the elemental reference energies using a FERE-like approach [16, 23]. Such correction schemes involve some more HT-DFT choices: (a) Should all elemental reference energies and/or compound formation energies be effectively fit to experimental data or only a subset? For instance, MP corrects the compound formation energies of nitrides, fluorides, chlorides, hydrides, sulfides of alkali, alkaline earth, and aluminum containing compounds [22]. The OQMD fits the reference energies of only elements whose DFT ground states are poor representation of the experimental reference states (i.e., elements that are gases or that have a solid-solid phase transition below room temperature) [23]. (b) What experimental thermochemical data should be used such correction schemes, given a lack of a single, widely-accepted set of standard experimental dataset for solids? For instance, MP and OQMD use experimental formation energies from different sources to fit elemental reference energies: MP uses data from Materials Thermochemistry [80], while OQMD uses data from SGTE SUBstance Database (SSUB) [81] in addition to others (see Refs. 23 and 77 for details of the fitting data used in the two databases). Some other standard reference databases are also widely used, such as the NIST-JANAF Thermochemical Tables [82]. Since a given material may have experimental data in one or more such reference databases of experimental properties, the choice of the source of experimental data affects the fitted formation energies in HT-DFT databases, even in cases where other parameters such as pseudopotentials used are held constant. This effect of fitted elemental reference states is shown in the calculated formation energies averaged over compounds containing each element in Figures S3, S7, and S11.

C. GGA vs. GGA+U approach

One of the ways to treat the issue of over-delocalization in DFT is to use the DFT+U approach [83, 84] (or "GGA+U" when used with GGA). Similar to the case of fitting elemental references, using the GGA+U approach requires additional HT-DFT choices. (a) Whether or not to use GGA+U for calculating properties of a given material. All three HT-DFT databases have slightly different sets of compounds for which the GGA+U approach is applied. The OQMD uses GGA+U only for oxides of certain $3d$ transition metals (the V–Cu series) and actinide metals [23]. MP uses GGA+U for oxides, fluorides, and sulfides of a larger set of transition metals, but not actinides [77]. AFLOW applies it to an even larger set of compounds, nearly all those containing d- or f-block elements [85]. (b) What effective U value should be used for each element? The three HT-DFT databases all use different effective U values for each element, obtained either from previous work (OQMD) or in-house parameterization by fitting to experimental data (AFLOW and MP) [18, 86]. Such choices around when to use the GGA+U approach to calculate a compound and what effective U value to use can impact some properties more than others, e.g., discrepancies in total magnetization values in the AFLOW-OQMD comparisons, particularly for "Rare-Earth", "Intermetallic", and "Metallic" classes. For some properties, such as formation energies, *post hoc* corrections are required to maintain consistency between those calculated using the GGA and GGA+U approaches, especially while constructing phase diagrams involving compounds calculated using the two different approaches. Such corrections are obtained by fitting to experimental reaction energies, and can be different between HT-DFT databases based on the source of such reaction energies.

V. CONCLUSION

Recent years have seen a dramatic increase in the application of informatics methods for materials development, using high-throughput DFT data. Several prominent HT-DFT databases exist and each uses different input parameters and post-processing techniques to calculate materials properties. Quantifying the uncertainty in calculated properties due to such parameter choices is therefore crucial to understanding the reproducibility and interoperability of such data. In this work, we centralize data from three of the largest HT-DFT databases, AFLOW, Materials Project, and OQMD, into a common data repository, allowing records to be accurately compared. We then compare four properties—formation energy, volume, band gap, and total magnetization—of materials calculated in each of the HT-DFT databases using the same initial crystal structure.

Our comparisons show that formation energy and volume are more easily reproduced than band gap and total magnetization. Interestingly, we find that the average difference in calculated properties across two HT-DFT databases is comparable to that between DFT and experiment: up to 0.105 eV/atom for formation energy, 4% for volume, 0.21 eV for band gap, and 0.15 μ_B/formula unit for total magnetization. Further, certain input parameter choices disproportionately affect HT-DFT properties of particular classes of materials, e.g. choice of planewave cutoff on formation energies and volumes of oxides and halides, and the choice of pseudopotential on the band gaps and magnetization of rare-earth compounds. Our results inform users of the variability to account for in reported materials properties, especially when using data from multiple HT-DFT databases in their own analyses. In addition, our quantitative uncertainty estimates can directly aid materials informatics efforts, e.g., for separation of model uncertainty and inherent noise in data.

As HT-DFT databases continue to mature, systematic comparisons, interoperability, and standardization of calculations become increasingly crucial. Efforts to improve the interoperability of materials databases, e.g., by the development of a common data schema by the OPTiMaDe consortium [87], are already ongoing. Toward improving the standardization of calculations, HT-DFT choices and reproducibility in particular, we list a few recommendations for next-generation and new iterations of current HT-DFT databases:

(a) *In-depth, versioned documentation* of the various parameter choices made in a high-throughput project, including the data-driven rationale for the choices, if any.

(b) *Visibility for possible uncertainty* in reported properties (in both the web and programmatic interfaces used to interact with HT-DFT data) for which HT-DFT choices are expected to have a significant impact. Further, we recommend providing estimated uncertainties in calculated properties, either determined from literature references (e.g., this work), or from in-house investigations (e.g., by performing a set of HT-DFT calculations with different input parameters as part of a sensitivity analysis).

(c) *Community-led initiative to reach a consensus* on which HT-DFT choices ought to be standardized (e.g., energy cutoffs, fitting sets for empirical corrections, post-processing steps to determine properties such as band gap) and which HT-DFT choices could be a source of greater scientific insight if they were more diverse (e.g., DFT codes, pseudopotentials, DFT exchange-correlation functionals).

CONFLICTS OF INTEREST

ZdR was previously employed by Citrine Informatics. PS has worked as a subcontractor to Citrine Informatics.

ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Small Business Technology Transfer Program under Award Number DE-SC0015106. The authors would like to thank Cormac Toher for advice on using the AFLUX RESTful API, and Anubhav Jain, Shyue Ping Ong, Matthew Horton, Eric B. Isaacs, and Chris Wolverton for their comments on an earlier version of this manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: C.K.H.B., V.H., P.S., M.H., J.E.S., B.M.; Methodology: C.K.H.B., V.H., E.A., Y.K., M.H., P.S., J.E.S.; Software: C.K.H.B., V.H., E.A., Y.K.; Validation: C.K.H.B., V.H., B.M.; Formal analysis: C.K.H.B., V.H., Z.d.R., E.A., Y.K.; Investigation: C.K.H.B., V.H.; Data Curation: C.K.H.B., V.H.; Writing – Original Draft: C.K.H.B., V.I.H., M.H., P.S., J.E.S.; Writing – Review & Editing: all authors; Visualization: E.A., Y.K., C.K.H.B., V.H.; Supervision: J.E.S., B.M., J.L.

DATA AVAILABILITY

All data and Python scripts required to perform the analysis presented in this work are made available via the GitHub repository at https://github.com/CitrineInformatics-ERD-public/htdft-uq.

Appendix A: Definitions of statistical quantities

The definitions of statistical quantities and their symbols used in this work throughout are as follows (x_i and y_i refer to the two sets of data being compared, e.g. from two different databases):

1. Median difference ($\widetilde{\Delta x}$):

$$\widetilde{\Delta x} = \text{median}(x_i - y_i) \tag{A1}$$

2. Median absolute difference (MAD):

$$\text{MAD} = \text{median}\big(\,|x_i - y_i|\,\big) \tag{A2}$$

3. Interquartile range (IQR):

$$\text{IQR} = \text{Q}_3 - \text{Q}_1 \tag{A3}$$

where Q_1 and Q_3 are the first and third quartiles (25th and 75th percentiles), respectively.

4. Median relative absolute difference (MRAD):

$$\text{MRAD} = \text{median}\left(\frac{|x_i - y_i|}{|x_i + y_i|/2} \times 100\right) \tag{A4}$$

5. Pearson correlation coefficient (r):

$$r(x,y) = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (y_i - \bar{y})^2}} \tag{A5}$$

where $\bar{x} = \frac{1}{n}\sum_i^n x_i$ is the sample mean, and n is the sample size.

6. Spearman's rank correlation coefficient (ρ) is defined as the Pearson correlation coefficient between rank variables x_i^{R} and y_i^{R} corresponding to raw data values x_i and y_i, respectively:

$$\rho(x,y) = r(x^{\text{R}}, y^{\text{R}}) \tag{A6}$$

[1] S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, The high-throughput highway to computational materials design, Nat. Mater. **12**, 191 (2013).

[2] A. Jain, Y. Shin, and K. A. Persson, Computational predictions of energy materials using density functional theory, Nat. Rev. Mater. **1**, 15004 (2016).

[3] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD), JOM **65**, 1501 (2013).

[4] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. **6**, 15 (1996).

[5] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab-initio total energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[6] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli, *et al.*, AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations, Comput. Mater. Sci. **58**, 227 (2012).

[7] M. Widom and M. Mihalkovic, Stability of Fe-based alloys with structure type C6Cr23, J. Mater. Res. **20**, 237 (2005).

[8] J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard, and J. K. Nørskov, CatApp: a web application for surface chemistry and heterogeneous catalysis, Angew. Chem. **124**, 278 (2012).

[9] R. D. Johnson III, *Computational Chemistry Comparison and Benchmark Database*, Tech. Rep. (National Institute of Standards and Technology, 1999).

[10] D. D. Landis, J. S. Hummelshøj, S. Nestorov, J. Greeley, M. Dułak, T. Bligaard, J. K. Nørskov, and K. W. Jacobsen, The computational materials repository, Comput. Sci. Eng. **14**, 51 (2012).

[11] R. Tran, Z. Xu, D. W. Balachandran Radhakrishnan, W. Sun, K. A. Persson, and S. P. Ong, Surface energies of elemental crystals, Sci. Data **3** (2016).

[12] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, and A. Aspuru-Guzik, The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid, J. Phys. Chem. Lett. **2**, 2241 (2011).

[13] K. Choudhary, K. F. Garrity, A. C. Reid, B. DeCost, A. J. Biacchi, A. R. H. Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, *et al.*, The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design, npj Comput. Mater. **6**, 1 (2020).

[14] L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos, *et al.*, Materials Cloud, a platform for open computational science, Sci. Data **7**, 1 (2020).

[15] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, *et al.*, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater. **1**, 011002 (2013).

[16] V. Stevanović, S. Lany, X. Zhang, and A. Zunger, Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies, Phys. Rev. B **85**, 115104 (2012).

[17] P. Gorai, D. Gao, B. Ortiz, S. Miller, S. A. Barnett, T. Mason, Q. Lv, V. Stevanović, and E. S. Toberer, TE

Design Lab: A virtual laboratory for thermoelectric material design, Comput. Mater. Sci. **112**, 368 (2016).
[18] W. Setyawan and S. Curtarolo, High-throughput electronic band structure calculations: Challenges and tools, Comput. Mater. Sci. **49**, 299 (2010).
[19] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, AiiDA: automated interactive infrastructure and database for computational science, Comput. Mater. Sci. **111**, 218 (2016).
[20] A. Larsen, J. Mortensen, J. Blomqvist, I. Castelli, R. Christensen, M. Dulak, J. Friis, M. Groves, B. Hammer, C. Hargus, *et al.*, The Atomic Simulation Environment: A Python library for working with atoms, J. Phys. Condens. Matter **29**, 273002 (2017).
[21] K. Mathew, A. K. Singh, J. J. Gabriel, K. Choudhary, S. B. Sinnott, A. V. Davydov, F. Tavazza, and R. G. Hennig, MPInterfaces: A Materials Project based Python tool for high-throughput computational screening of interfacial systems, Comput. Mater. Sci. **122**, 183 (2016).
[22] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. **68**, 314 (2013).
[23] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, npj Comput. Mater. **1**, 15010 (2015).
[24] A. Khorshidi and A. A. Peterson, Amp: a modular approach to machine learning in atomistic simulations, Comput. Phys. Commun. **207**, 310 (2016).
[25] A. Van De Walle, M. Asta, and G. Ceder, The alloy theoretic automated toolkit: A user guide, Calphad **26**, 539 (2002).
[26] K. Mathew, J. H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I.-h. Chu, T. Smidt, B. Bocklund, M. Horton, *et al.*, Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows, Comput. Mater. Sci. **139**, 140 (2017).
[27] Y. Wang, J. Lv, L. Zhu, and Y. Ma, CALYPSO: A method for crystal structure prediction, Comput. Phys. Commun. **183**, 2063 (2012).
[28] T. Mayeshiba, H. Wu, T. Angsten, A. Kaczmarowski, Z. Song, G. Jenness, W. Xie, and D. Morgan, The MAterials Simulation Toolkit (MAST) for atomistic modeling of defects and diffusion, Comput. Mater. Sci. **126**, 90 (2017).
[29] A. Togo and I. Tanaka, First principles phonon calculations in materials science, Scr. Mater. **108**, 1 (2015).
[30] Y. Hinuma, G. Pizzi, Y. Kumagai, F. Oba, and I. Tanaka, Band structure diagram paths based on crystallography, Comput. Mater. Sci. **128**, 140 (2017).
[31] Q.-J. Hong and A. van de Walle, A user guide for SLUSCHI: solid and liquid in ultra small coexistence with hovering interfaces, Calphad **52**, 88 (2016).
[32] C. W. Glass, A. R. Oganov, and N. Hansen, USPEX: evolutionary crystal structure prediction, Comput. Phys. Commun. **175**, 713 (2006).
[33] D. C. Lonie and E. Zurek, XtalOpt: An open-source evolutionary algorithm for crystal structure prediction, Comput. Phys. Commun. **182**, 372 (2011).
[34] See Supplemental Material at [URL to be inserted by publisher] for HT-DFT databases and libraries, data management details, expanded analysis with element-wise HT-DFT differences, and example PIFs.
[35] A. E. Mattsson, P. A. Schultz, M. P. Desjarlais, T. R. Mattsson, and K. Leung, Designing meaningful density functional theory calculations in materials science—a primer, Modell. Simul. Mater. Sci. Eng. **13**, R1 (2004).
[36] S. P. Ong, S. M. Blau, X. Qu, W. Richards, S. Dwaraknath, S. Dacek, J. Montoya, R. Kingsbury, A. Jain, JSX, M. Horton, D. Waroquiers, R. Tran, H. Tang, P. Huck, G. Hautier, G. Petretto, sivonxay, C. Zheng, KeLiu, and A. Rutt, materialsproject/custodian: v2020.4.27 (2020).
[37] S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A. V. Yakutovich, C. W. Andersen, *et al.*, AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance, Scientific data **7**, 1 (2020).
[38] Y. Zhuo, A. M. Tehrani, A. O. Oliynyk, A. C. Duke, and J. Brgoch, Identifying an efficient, thermally robust inorganic phosphor host via machine learning, Nat. Commun. **9**, 1 (2018).
[39] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, Phys. Rev. B **89**, 094104 (2014).
[40] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, *et al.*, Reproducibility in density functional theory calculations of solids, Science **351**, aad3000 (2016).
[41] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design, Acta Crystall. B **58**, 364 (2002).
[42] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1996).
[43] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).
[44] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).
[45] W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson, and S. Curtarolo, High-throughput combinatorial database of electronic band structures for inorganic scintillator materials, ACS Combi. Sci. **13**, 382 (2011).
[46] O. Levy, G. L. Hart, and S. Curtarolo, Uncovering compounds by synergy of cluster expansion and high-throughput methods, J. Am. Chem. Soc. **132**, 4830 (2010).
[47] C. Toher, J. J. Plata, O. Levy, M. de Jong, M. Asta, M. B. Nardelli, and S. Curtarolo, High-throughput computational screening of thermal conductivity, debye temperature, and grüneisen parameter using a quasiharmonic debye model, Phys. Rev. B **90**, 174107 (2014).
[48] C. Toher, C. Oses, J. J. Plata, D. Hicks, F. Rose, O. Levy, M. de Jong, M. Asta, M. Fornari, M. B. Nardelli, *et al.*, Combining the AFLOW GIBBS and elastic libraries to efficiently and robustly screen thermomechanical properties of solids, Phys. Rev. Mater. **1**, 015401 (2017).
[49] R. H. Taylor, F. Rose, C. Toher, O. Levy, K. Yang, M. B. Nardelli, and S. Curtarolo, A RESTful API for exchanging materials data in the AFLOWLIB.org consortium,

Comput. Mater. Sci. **93**, 178 (2014).
[50] M. De Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. K. Ande, S. Van Der Zwaag, J. J. Plata, *et al.*, Charting the complete elastic properties of inorganic crystalline compounds, Sci. Data **2**, 150009 (2015).
[51] W. Chen, J.-H. Pöhls, G. Hautier, D. Broberg, S. Bajaj, U. Aydemir, Z. M. Gibbs, H. Zhu, M. Asta, G. J. Snyder, *et al.*, Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment, J. Mater. Chem. C **4**, 4414 (2016).
[52] M. De Jong, W. Chen, H. Geerlings, M. Asta, and K. A. Persson, A database to enable discovery and design of piezoelectric materials, Sci. Data **2**, 150053 (2015).
[53] I. Petousis, D. Mrdjenovich, E. Ballouz, M. Liu, D. Winston, W. Chen, T. Graf, T. D. Schladt, K. A. Persson, and F. B. Prinz, High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials, Sci. Data **4**, 160134 (2017).
[54] G. Petretto, S. Dwaraknath, H. P. Miranda, D. Winston, M. Giantomassi, M. J. Van Setten, X. Gonze, K. A. Persson, G. Hautier, and G.-M. Rignanese, High-throughput density-functional perturbation theory phonons for inorganic materials, Sci. Data **5**, 1 (2018).
[55] C. Zheng, K. Mathew, C. Chen, Y. Chen, H. Tang, A. Dozier, J. J. Kas, F. D. Vila, J. J. Rehr, L. F. Piper, *et al.*, Automated generation and ensemble-learned matching of X-ray absorption spectra, npj Comput. Mater. **4**, 1 (2018).
[56] K. A. Persson, B. Waldwick, P. Lazic, and G. Ceder, Prediction of solid-aqueous equilibria: scheme to combine first-principles calculations of solids with experimental aqueous states, Phys. Rev. B **85**, 235438 (2012).
[57] S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder, and K. A. Persson, 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. **97**, 209 (2015).
[58] The Materials Project: The Materials API, https://materialsproject.org/docs/api (accessed: December 2019).
[59] S. Kirklin, J. E. Saal, V. I. Hegde, and C. Wolverton, High-throughput computational search for strengthening precipitates in alloys, Acta Mater. **102**, 125 (2016).
[60] A. A. Emery, J. E. Saal, S. Kirklin, V. I. Hegde, and C. Wolverton, High-throughput computational screening of perovskites for thermochemical water splitting applications, Chem. Mater. **28**, 5621 (2016).
[61] D. Wang, M. Amsler, V. I. Hegde, J. E. Saal, A. Issa, B.-C. Zhou, X. Zeng, and C. Wolverton, Crystal structure, energetics, and phase stability of strengthening precipitates in Mg alloys: A first-principles study, Acta Mater. **158**, 65 (2018).
[62] A. R Akbarzadeh, V. Ozoliņš, and C. Wolverton, First-principles determination of multicomponent hydride phase diagrams: Application to the Li-Mg-N-H system, Adv. Mater. **19**, 3233 (2007).
[63] V. I. Hegde, M. Aykol, S. Kirklin, and C. Wolverton, The phase stability network of all inorganic materials, Sci. Adv. **6**, eaay5606 (2020).
[64] M. Amsler, V. I. Hegde, S. D. Jacobsen, and C. Wolverton, Exploring the high-pressure materials genome, Phys. Rev. X **8**, 041021 (2018).
[65] OQMD RESTful API, http://oqmd.org/static/docs/restful.html (2019), accessed: May 2020.
[66] K. Michel and B. Meredig, Beyond bulk single crystals: a data format for all materials structure–property–processing relationships, MRS Bull. **41**, 617 (2016).
[67] pypif: Python toolkit for working with PIFs, https://github.com/CitrineInformatics/pypif (2018), accessed: May 2020.
[68] P. Haas, F. Tran, and P. Blaha, Calculation of the lattice constant of solids with semilocal functionals, Phys. Rev. B **79**, 085104 (2009).
[69] G. Olsen, C. Nuese, and R. Smith, The effect of elastic strain on energy band gap and lattice parameter in III-V compounds, J. Appl. Phys. **49**, 5523 (1978).
[70] C. Kuo, S. Vong, R. Cohen, and G. Stringfellow, Effect of mismatch strain on band gap in III-V semiconductors, J. Appl. Phys. **57**, 5428 (1985).
[71] S.-H. Wei and A. Zunger, Predicted band-gap pressure coefficients of all diamond and zinc-blende semiconductors: Chemical trends, Phys. Rev. B **60**, 5404 (1999).
[72] J. M. Munro, K. Latimer, M. K. Horton, S. Dwaraknath, and K. A. Persson, An improved symmetry-based approach to reciprocal space path selection in band structure calculations, npj Comput. Mater. **6**, 1 (2020).
[73] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozman, M. Venkatesan, M. Coey, and S. Curtarolo, Accelerated discovery of new magnets in the Heusler alloy family, Sci. Adv. **3**, e1602241 (2017).
[74] M. K. Horton, J. H. Montoya, M. Liu, and K. A. Persson, High-throughput prediction of the ground-state collinear magnetic order of inorganic materials using density functional theory, npj Comput. Mater. **5**, 1 (2019).
[75] J. Marquez Chavez and B. Kiefer, Matcor, a program for the cross-validation of material properties between databases, Comput. Mater. Sci. **187**, 110103 (2021).
[76] Materials Project: Calculations Guide, https://materialsproject.org/docs/calculations (accessed: December 2019).
[77] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, A high-throughput infrastructure for density functional theory calculations, Comput. Mater. Sci. **50**, 2295 (2011).
[78] L. Eyring, K. A. Gschneidner, and G. H. Lander, *Handbook on the physics and chemistry of rare earths*, Vol. 32 (Elsevier, 2002).
[79] S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, *et al.*, Aflow: An automatic framework for high-throughput materials discovery, Comput. Mater. Sci. **58**, 218 (2012).
[80] O. Kubaschewski, C. B. Alcock, and P. Spencer, *Materials Thermochemistry* (Pergamon Press, 1993).
[81] SGTE, *Thermodynamic Properties of Inorganic Materials*, Vol. 19 (Springer-Verlag, 1999).
[82] M. W. Chase, NIST-JANAF Thermochemical Tables 4th Ed., J. Phys. Chem. Ref. Data **Monograph No. 9**, 1529 (1998).
[83] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Band theory and mott insulators: Hubbard U instead of Stoner I, Phys. Rev. B **44**, 943 (1991).
[84] H. J. Kulik, Perspective: Treating electron overdelocalization with the DFT+U method, J. Chem. Phys. **142**, 240901 (2015).

[85] C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. Hart, M. B. Nardelli, and S. Curtarolo, The AFLOW standard for high-throughput materials science calculations, Comput. Mater. Sci. **108**, 233 (2015).

[86] A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson, and G. Ceder, Formation enthalpies by mixing GGA and GGA+U calculations, Phys. Rev. B **84**, 045115 (2011).

[87] C. Andersen, R. Armiento, E. Blokhin, G. Conduit, S. Dwaraknath, M. L. Evans, A. Fekete, A. Gopakumar, S. Gražulis, V. Hegde, M. Horton, S. Kumbhar, N. Marzari, A. Merkys, F. Mohamed, A. Morris, C. Oses, G. Pizzi, T. Purcell, G.-M. Rignanese, M. Scheffler, M. Scheidgen, L. Talirz, C. Toher, M. Uhrin, D. Winston, and C. Wolverton, The OPTIMADE Specification (2020).