

ARTICLE OPEN



Identifying candidate hosts for quantum defects via data mining

Austin M. Ferrenti¹✉, Nathalie P. de Leon², Jeff D. Thompson² and Robert J. Cava¹

Atom-like defects in solid-state hosts are promising candidates for the development of quantum information systems, but despite their importance, the host substrate/defect combinations currently under study have almost exclusively been found serendipitously. Here we systematically evaluate the suitability of host materials by applying a combined four-stage data mining and manual screening process to all entries in the Materials Project database, with literature-based experimental confirmation of band gap values. We identify a total of 541 viable hosts (16 unary and 74 binary) for quantum defect introduction and potential use in quantum information systems. This represents a significant (99.57%) reduction from the total number of known inorganic phases, and the application of additional selection criteria for specific applications will reduce their number even further. The screening principles outlined may easily be applied to previously unrealized phases and other technologically important materials systems.

npj Computational Materials (2020)6:126; <https://doi.org/10.1038/s41524-020-00391-7>

INTRODUCTION

In recent years, significant effort has been devoted to the realization of functional systems for quantum information science (QIS). QIS devices employ the manipulation of quantum states to store, process, and transmit information, potentially enabling the rapid solution of problems long thought impossible or impractical to address through classical methods. One of the most promising platforms for the development of efficient quantum information systems, particularly for application in quantum networks¹, is atomic or atom-like defects in solid-state hosts². The quantum coherence characteristics of the atomic defect, low defect concentrations, and quality of the host when combined should allow for long spin coherence and efficient optical transitions, properties also ideal for nanoscale sensing under ambient conditions³.

Several such atomic defect systems, including vacancy centers in diamond^{4–9} and silicon carbide¹⁰ have been extensively studied for several applications, in particular quantum networks^{11,12}, magnetometry and nanoscale sensors of magnetic fields^{3,13,14}, electric fields¹⁵, temperature^{16,17} or chemical composition using NMR^{18,19}, often under ambient conditions at room temperature²⁰. Another well-known class of defects is transition-metal and rare-earth ion impurities. These defects have been extensively explored in the context of solid-state laser development, in materials where extremely high doping concentrations are possible. However, exploring single defects as qubits is a more recent development^{21,22}, partially facilitated by integration with nano-photonics circuits to modify and enhance their luminescence^{23,24}. Although significant attention has thus far been paid to a handful of materials and defects, they represent only a small fraction of the greater body of potential defect-host systems. As most known quantum information materials systems have been discovered indirectly, a rational search of inorganic materials can yield candidates with superior properties.

Recent work has focused on *ab initio* predictions of host–defect systems that can be carried out with electronic structure calculations, as recently demonstrated with vacancy centers in

silicon carbide²⁵, and such calculations have been carried out for several individual host–defect combinations. An alternative approach is to systematically screen for potential host materials by first postulating which properties would be ideal for host material candidates, and then filtering materials based on those properties. This is done most efficiently through a computational search that substantially narrows the field of potential candidates.

In order to achieve the long spin coherence and high efficiency optical transitions necessary for a usable quantum information system, host substrates must be highly pure (i.e., as free from defects as possible), intrinsically diamagnetic (thereby reducing magnetic noise and spin-based relaxation of the defect quantum state), and possess a band gap large enough to accommodate the ground and excited state energy levels of the defect (separated by an optical frequency). The ideal, pure substrate would be free of paramagnetic impurities or unwanted defects that may influence the band gap character and/or magnetic or electric field environment surrounding the implanted defect, reducing efficiency. Those materials known to be dopable, to possess controllable surfaces, and to possess a known method of epitaxial thin-film production would further facilitate the production of integrated devices.

Reducing magnetic noise and relaxation in the environment of the defect within the host requires both the minimization of paramagnetic centers in the host and the absence of host nuclei with nonzero magnetic moments. Consideration of the natural abundance of zero nuclear spin isotopes of each element, shown in Fig. 1, allows for the exclusion of a significant portion of the periodic table, for which no stable spin-zero isotopes exist. While some elements, such as O and Ca, are known to exist almost exclusively as nuclear spin-free isotopes, elemental species with at least 50% nuclear spin-free isotopes could likely be isotopically enriched to higher concentrations, as has been achieved in diamond^{26,27} and silicon²⁸. Although several of the lanthanide elements appear to have relatively high natural abundances of spin-free isotopes, the difficulty in obtaining pure, diamagnetic starting material of these (free from magnetic lanthanide

¹Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. ²Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA. ✉email: afferenti97@gmail.com

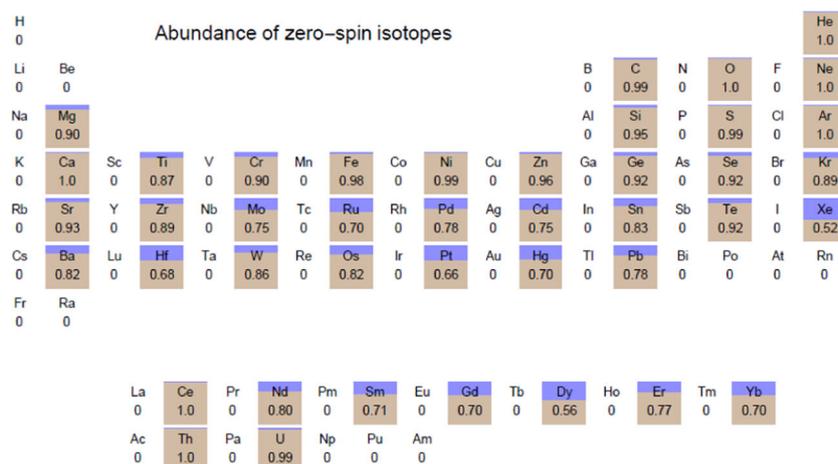


Fig. 1 Color-coded periodic table depicting the natural abundance of stable zero nuclear spin isotopes. Represented by the portion shaded gold ($l=0$) relative to the portion shaded blue ($l \neq 0$). Those elements for which there exist no stable spin-zero isotopes are shaded white.

impurities) excludes them from consideration here. All transition-metal elements with unpaired electrons are eliminated due to their paramagnetism. The optical coherence of defects may be sensitive to the presence of permanent electric dipole moments about the defect site, and thus also to the symmetry of the crystalline host, so in order to isolate only those phases most likely to possess highly symmetric ground and excited state defect sites, phases crystallizing in polar space groups were also not considered.

In the current work, a systematic application of each of the criteria outlined above was performed, beginning with the totality of known inorganic phases listed in the Materials Project database. As the practice of database mining is often highly automated, the sensitivity of the desired properties to the exact phase and crystal structure reported for each potential host species necessitated manual checks after each restriction of the data set. While the true number of viable quantum host-defect pairings may be effectively infinite due to the fact that different quantum defects will be suitable for different applications, of the 100,000+ inorganic materials in the inorganic crystal structure database (ICSD), a maximum of 541 phases were found to be potentially viable hosts for atom-like defect-based quantum information systems. Among these, a small number are simple single elements or binary compounds, making them relatively straightforward to prepare and study as substrates for QIS.

RESULTS

Of the 125,223 inorganic compounds listed in the Materials Project database, a maximum of 541 phases (0.43%) are found to fit the criteria outlined above as suitable hosts for quantum defects (Fig. 2a). While there likely exist significantly more potentially viable phases that have yet to be recognized, this total spans all currently reported experimentally synthesized and theoretically predicted inorganic materials. As the ease of growth of suitably sized single crystals, stability under standard or near-standard atmospheric conditions, and chemical and structural simplicity are major factors in further determining the viability of potential host phases, it is likely that those identified here represent the majority of suitable candidates.

Of the 541 identified phases, the band gap character of at least 521 could be either confirmed or reasonably assumed to be comparable or larger than that of silicon (≥ 1.1 eV), based on calculated and/or reported experimental values. The band gaps of the remaining 20 candidates, grayed in Supplementary Tables 1–4,

have either been reported to be slightly smaller than that of Si, or could only be tentatively extrapolated to a similar value.

Only two admitted elements, Ni and Fe, were not found to be present in any of the listed phases, as all Ni- and Fe- containing materials considered were found to be intrinsically paramagnetic. All of the Os-containing species in the tabulations, as well as the majority of those containing Ru, are transition-metal-based carbonyl cluster compounds (Fig. 2b).

DISCUSSION

From the total number of known, experimentally synthesized inorganic material phases, we have derived a catalog of all potentially viable host substrates for functional, atom-like defect-based quantum information systems. A combination of automatic database and subsequent manual screening was performed to ensure that each reported phase possessed the necessary electronic, magnetic, and optical properties, as well as stability under standard conditions, necessary for application in QIS. This screening was conducted across several materials databases (including the Materials Project, JARVIS-DFT, and the AFLOW repository), the AFLOW-ML prediction API, and the existing body of literature in an attempt to compensate for known computational and experimental biases and errors that may have influenced our results. Of the 541 total viable candidates found, 16 are unary, 74 are binary, 322 are ternary, and 129 are quaternary or higher. Many are oxides and chalcogenides, suggesting that they will be relatively simple to fabricate and study, while others in the tabulation, such as the osmates and higher-order elemental clusters, will likely prove impractical for currently considered device applications. Although the total number of viable candidates remains too large for extensive experimental studies, 90 of them are relatively basic (unary and binary) systems, which are typically easier to synthesize than higher-order species. This list may also be further reduced by additional computational studies, including higher-quality band structure calculations and more application-specific property predictions, such as for surface stability and excited state lifetimes of otherwise promising materials. Future studies concerning inorganic material systems would likely benefit from the implementation of similar systematic screening processes, enabling the rapid and conclusive determination of all viable phases for any number of applications.

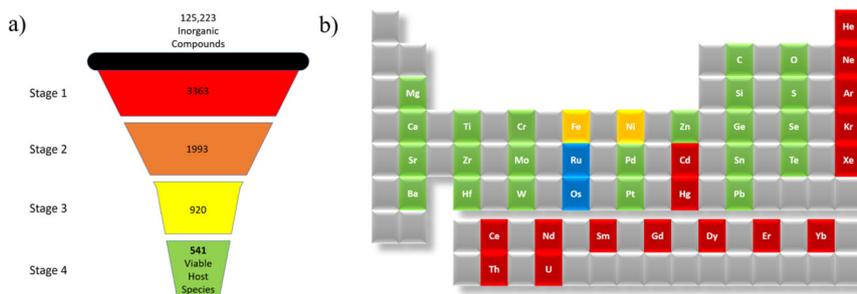


Fig. 2 Tiered screening system for the identification of potentially viable host species for quantum defect implantation. **a** A schematic view of the process. A maximum of 0.43% of all phases in the International Crystal Structure Database were found to be potentially viable host substrates, with the most substantial reduction in candidate phases (97.3%) occurring in Stage 1 of the screening process. **b** Color-coded periodic table of constituent elements in viable host species depicting elements: disregarded due to the absence of spin-zero isotopes (gray); deemed too hazardous, too radioactive, too difficult to purify, and too unreactive for inclusion (red); not present in any identified phases (yellow); present largely only in cluster complex phases (blue); and present in numerous identified phases (green).

METHODS

Screening procedure

To reduce the possibility of accidental exclusion of viable host materials, the screening of the Materials Project's 125,223 inorganic compound entries was conducted in four distinct stages, as shown in Fig. 3.

Stage 1. The first screen was applied utilizing the Materials Project's built-in query API²⁹, and admitted all entries corresponding to phases crystallizing in nonpolar space groups, derived from an experimental ICSD entry, containing only those elements with a >50% natural abundance of zero nuclear spin isotopes, and calculated to be nonmagnetic systems.

Stage 2. The second screen was applied by hand, and included the removal of all phases containing uranium, thorium, cadmium, and mercury, due to the radioactive and/or toxic nature of the majority of stable phases that each may form, as well as those containing noble gases, as none exist as stable solids under standard conditions. All rare-earth-element-containing phases were also removed, due to the difficulty in obtaining sufficiently pure starting materials free of nuclear spin.

Stage 3. The third and most important filter was applied utilizing calculated band gaps for all remaining phases reported by three publicly available DFT computation repositories; the Materials Project, the Joint Automated Repository for Various Integrated Simulations (JARVIS), and the Automatic Flow of Materials Discovery (AFLOW) library, each of which employs a different band gap calculation and/or prediction method. In addition, where available, the machine-learning-predicted band gap reported by Isayev et al. in their predictive modeling studies was also taken into consideration.

At this stage, those phases with a Materials Project band gap of 2.0 eV were deemed likely to possess a large enough band gap to accommodate the necessary optical transitions of useful quantum defects, and, so as not to miss any potentially viable candidates, only those with calculated band gaps <0.5 eV were removed. For all other phases, any reports of experimental or otherwise calculated band gaps in the literature were taken into consideration. Various higher-order compounds for which the calculated band gaps were insufficient, such as in the Ba-Hf-S family, were approved based on general band gap trends. Those phases with calculated and/or experimentally derived band gaps consistently in the range of that of Si, which would represent the lower bound of viability for potential host materials, are grayed in Supplementary Tables 1–4.

The Materials Project³⁰ utilizes the DFT-based Vienna ab initio simulation package (VASP) software³¹ to calculate a wide variety of structural, energetic, and electronic properties for all reported inorganic structures in the ICSD. An initial relaxation of cell and lattice parameters is performed using a 1000 *k*-point mesh to ensure that all properties calculated are representative of the idealized unit cell for each material in its respective crystal structure. To calculate band structures for these materials, the generalized gradient approximation (GGA) functional is applied to the relaxed structures. For structures containing one of several transition-metal elements such as Cr, Fe, Mo, and W, the +U correction is also applied to correct for correlation effects in occupied *d*- and *f*-orbital systems that are not addressed by pure GGA calculations³². The authors caution that due to the high computational cost of more sophisticated calculation methods,

those employed often produce severely underestimated band gaps relative to the experimentally derived values. The energy above the hull (E Above Hull), which indicates the thermodynamic stability for each phase, with respect to decomposition, is also computed. An E Above Hull value of zero is defined as the most stable phase at a given composition, while large positive values indicate increased instability with respect to the stable phase(s).

JARVIS-DFT³³, originally compiled as a database for functional materials, with a focus on the discovery of novel two-dimensional systems, also employs the DFT-based VASP software to perform a variety of material property calculations. As opposed to the Materials Project, JARVIS-DFT employs the OptB88vdW (OPT) functional, which was initially designed to better approximate the properties of two-dimensional van der Waals materials, and has since also been shown to be effective for bulk systems^{34,35}. Structures are first sourced from the Materials Project database, and then re-optimized using the OPT functional. A representative band gap is then calculated through a combination of the OPT and modified Becke–Johnson (mBJ) functionals. The mBJ and combined OPT-mBJ functionals have both been shown to predict band gap sizes with more accuracy than other DFT-based calculation methods³⁶.

The AFLOW³⁷ repository relies on a highly sophisticated and automatic framework for the calculation of a wide array of inorganic material properties³⁸. The GGA-based Perdew–Burke–Ernzerhof functional with projector-augmented wavefunction (PAW) potentials is first used within the VASP software to twice relax and optimize the ICSD-sourced structure using a 3000–6000 *k*-point mesh. The increased *k*-point mesh density, compared to that employed by the Materials Project, is indicative of a more computationally expensive calculation. The band structure is then calculated with an even higher-density *k*-point mesh, as well as with the +U correction term for most occupied *d*- and *f*-orbital systems, and the standard band gap (E_{gap}) is extracted³⁹. A “fitted” band gap (E_{gfit}) is then calculated by applying a standard fit, derived from a selective study of DFT-computed vs. experimentally measured band gap widths, to the initially calculated value⁴⁰.

AFLOW-ML⁴¹, a machine-learning API designed to predict thermo-mechanical and electronic properties based on chemical composition alone, further builds upon the entries present in the ICSD and calculated through the AFLOW framework. Using only provided atomic compositional and positional information, so-called “fragment descriptors”, the system first applies a binary metal/insulator classification model. For materials predicted to be insulators, an additional regression model is applied to predict the band gap width. Each model was subjected to a fivefold cross validation process, in which it was trained to more accurately predict properties in an independent data set. The initial binary classification model is shown to have a 93% prediction success rate, with the majority of misclassified materials being narrow-gap semiconductors. While the accuracy of the predicted band gap sizes relative to experimental values is not mentioned by the authors, roughly 93% of the machine-learning-derived values are found to be within 25% of the DFT + U-calculated gap width. Only those phases identified in the authors' initial cross-validated test set were used for comparison.

Stage 4. The final stage. The criteria by which potential host species were excluded in the first three stages were largely derived computationally,

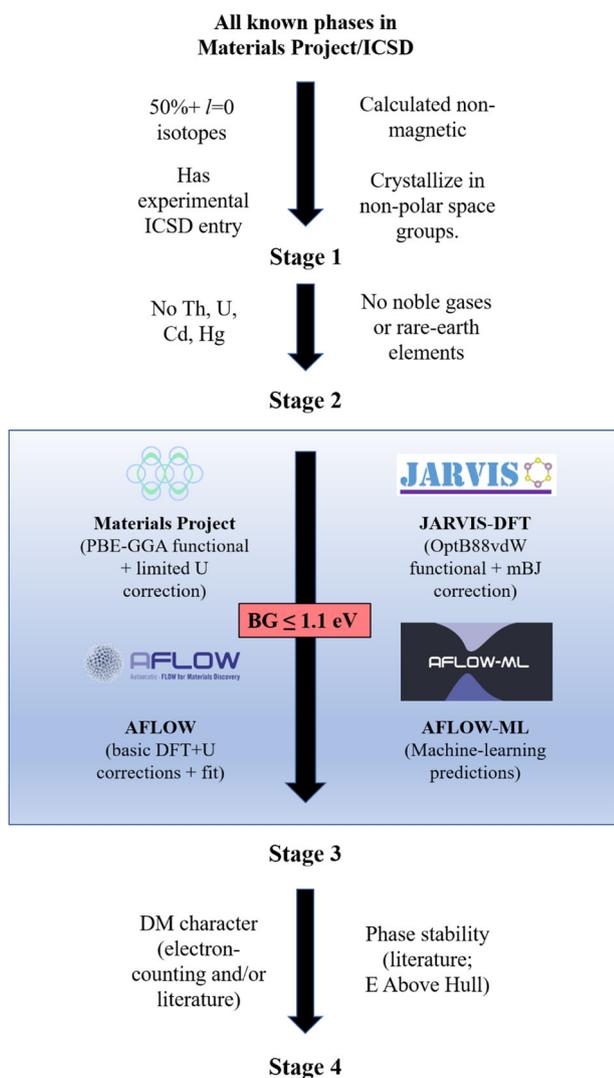


Fig. 3 Four-tier screening process to evaluate the potential viability of materials as hosts for quantum defects. Stage 1: Beginning with all known experimental inorganic phases, those with a calculated nonzero net magnetic moment, crystallization in a polar space group and/or containing atomic species with $<50\%$ $I=0$ isotopes were removed. Stage 2: Phases containing radioactive Th and U, toxic Cd and Hg, and the often magnetically impure rare-earth elements were removed. The one phase containing a noble gas element that passed this sieve, orthorhombic XeO_3 , was also removed due to its known instability under standard conditions. Stage 3: The calculated, predicted, and measured experimental band gaps of the remaining phases were recorded from the Materials Project, the JARVIS-DFT database, the AFLOW repository, the AFLOW-ML API, and the existing literature. Those materials that could not be reasonably predicted to have a band gap larger than 1.1 eV were dismissed. Stage 4: The intrinsic diamagnetic (DM) character and phase stability under standard conditions of the remaining candidates were confirmed using a combination of standard electron-counting principles and literature reports. Remaining phases with calculated E Above Hull values >0.2 eV/atom (from the Materials Project database) were dismissed. All database searches were conducted in February and March of 2020.

with subsequent manual checks. In contrast, the final stage of screening involved the confirmation of various fundamental properties of the materials found in the existing literature.

First, for those remaining stoichiometries for which multiple phases appeared to be viable, the relative stability of each polymorph at STP was

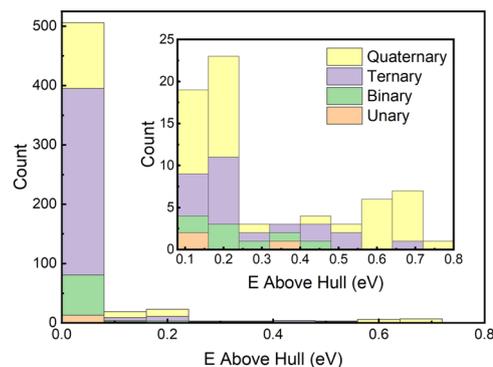


Fig. 4 Distribution of Materials Project-calculated E Above Hull values for otherwise viable host phases. The distribution of calculated E Above Hull values found for unary (orange), binary (green), ternary (purple), and quaternary or higher-order (yellow) phases which appeared to be otherwise viable at the beginning of the Stage 4 screening. (Inset) Zoomed view of the distribution from 0.08 – 0.80 eV/atom above the hull.

recorded. All recognized high-pressure and high-temperature phases that were not reported to be quenchable to a stable state under standard conditions were removed. The relative thermodynamic stability of each identified phase was also considered. As larger calculated E Above Hull values are indicative of a thermodynamically unstable material, phases for which the Materials Project predicted an E Above Hull >0.2 eV/atom were excluded from consideration, since they would likely be too prone to decomposition to be viable (Fig. 4). A total of 39 (largely ternary and quaternary) entries were deemed nonviable.

The intrinsic magnetic character of each pure phase was then confirmed through a combination of standard electron-counting rules and literature reports. As many of the potentially viable materials contained oxygen, particular care was taken to consider whether reported paramagnetic character could be due to oxygen defects, especially in the various molybdate, platinate, and palladate phases. Any pure phase that was reported to deviate from diamagnetic character was removed.

Where the true size of a material's band gap could not be reasonably assumed suitable based on calculations alone, other reported computationally and experimentally derived band gap values were also taken into consideration when available. While experimental band gaps were either not available or not recorded for the majority of phases considered, those that were available were compared to the calculated values from each database, in order to better judge the viability of the selected materials with small calculated gaps (Fig. 5). The standard band gap calculations performed by the Materials Project, JARVIS-DFT, and AFLOW databases were found to be underestimated by roughly 40–50%, relative to measured values, consistent with the long-known inaccuracies of DFT estimates, often due to the presence of significant spin-orbit coupling effects⁴². AFLOW-ML's machine-learning-based predictions were underestimated to a similar degree. On average, the OPT-mBJ hybrid functional employed by the JARVIS-DFT database for some phases was found to reduce this underestimation to 18.4%, while the "fitted" band gap calculated by AFLOW was actually observed to be overestimated by about 1.8% relative to experimental values. However, the latter figure is heavily influenced by several outlying underestimations in the AFLOW "fitted" gap data set, suggesting that on average, band gaps calculated in this manner will be overestimated to a greater degree. The determination of band gap suitability was thus made with these findings in mind, as well as with the understanding that the database-calculated band gap widths for heavier element-containing phases are often significantly underestimated, sometimes due to the failure of standard DFT calculation methods to account for spin-orbit coupling effects⁴².

Additional considerations

It should be noted that despite appearing in the literature, the most stable phases of several, long-recognized materials were found to lack an experimental crystal structure in the ICSD, and subsequently were also absent from the Materials Project database. Several of these, such as the STP-stable phases of BaGeO_3 , BaGe_2O_5 , and C_{70} would be potentially viable host

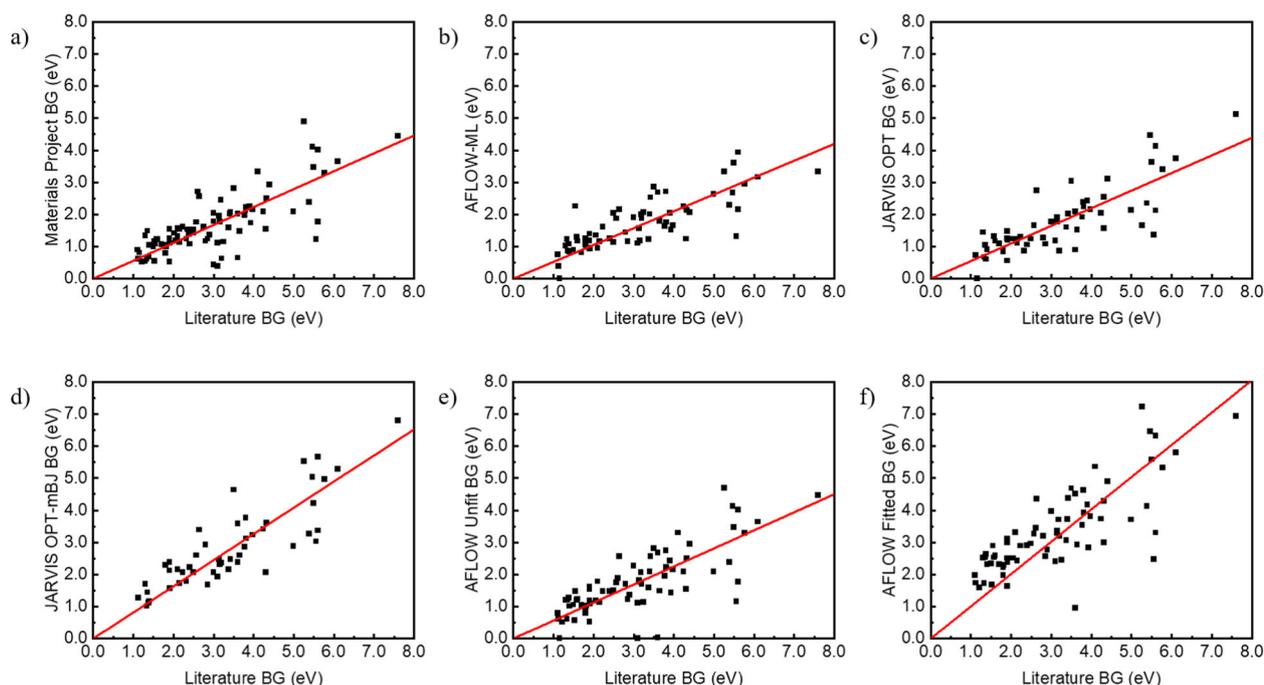


Fig. 5 Distribution of database-calculated and experimentally-measured band gap widths for a representative sample of phases. Comparison of reported experimental band gaps to those calculated by **a** the Materials Project, **b** AFLOW-ML, **c** JARVIS-DFT (OPT only), **d** JARVIS-DFT (OPT-mBJ), **e** AFLOW (unfit), and **f** AFLOW (fit). On average, the standard DFT calculations **a**, **c**, **e** and machine-learning predictions **b** were found to underestimate the experimental band gaps by roughly 40–50%. JARVIS-DFT OPT-mBJ calculations **d** were found to underestimate measured values by about 18% on average, while AFLOW fitted band gap calculations **f** were found to overestimate by about 2% on average, with several significant outlying underestimations. The red lines show the linear least squares fits to the data.

species, but due to their absence from any of the databases studied are not included. However, while comparing computed band gaps with reported literature values, three additional phases were identified that lacked a corresponding Materials Project entry, but did appear in at least one of the other databases. fcc-C_{60} is the only included phase with a corresponding ICSD entry that did not appear in any of the databases considered.

While suitably sized single crystals are necessary for the fabrication of functional devices for quantum information systems, few materials have been reported to be easily grown as large, defect-free, and optically clear single crystals. As such, the existence of a published single-crystal synthesis, regardless of product size, is denoted in the “SC” column in the tables of results by an asterisk. Noted air and moisture instability in the literature was also considered, but was not exclusionary, as additional post-synthesis processing may allow otherwise viable host materials to be utilized.

DATA AVAILABILITY

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files.

Received: 7 May 2020; Accepted: 27 July 2020;

Published online: 18 August 2020

REFERENCES

- Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
- Awschalom, D. D., Hanson, R., Wrachtrup, J. & Zhou, B. B. Quantum technologies with optically interfaced solid-state spins. *Nat. Photonics* **12**, 516–527 (2018).
- Maze, J. R. et al. Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* **455**, 644–647 (2008).
- Awschalom, D. D., Bassett, L. C., Dzurak, A. S., Hu, E. L. & Petta, J. R. Quantum spintronics: engineering and manipulating atom-like spins in semiconductors. *Science* **339**, 1174–1179 (2013).
- Rogers, L. J. et al. All-optical initialization, readout, and coherent preparation of single silicon-vacancy spins in diamond. *Phys. Rev. Lett.* **113**, 263602 (2014).
- Rose, B. C. et al. Observation of an environmentally insensitive solid-state spin defect in diamond. *Science* **361**, 60–63 (2018).
- Bhaskar, M. K. et al. Quantum nonlinear optics with a germanium-vacancy color center in a nanoscale diamond waveguide. *Phys. Rev. Lett.* **118**, 223603 (2017).
- Iwasaki, T. et al. Tin-vacancy quantum emitters in diamond. *Phys. Rev. Lett.* **119**, 253601 (2017).
- Trusheim, M. E. et al. Lead-related quantum emitters in diamond. *Phys. Rev. B* **99**, 075430 (2019).
- Koehl, W. F., Buckley, B. B., Heremans, F. J., Calusine, G. & Awschalom, D. D. Room temperature coherent control of defect spin qubits in silicon carbide. *Nature* **479**, 84–87 (2011).
- Togan, E. et al. Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
- Hensen, B. et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
- Taylor, J. M. et al. High-sensitivity diamond magnetometer with nanoscale resolution. *Nat. Phys.* **4**, 810–816 (2008).
- Lovchinsky, I. et al. Nuclear magnetic resonance detection and spectroscopy of single proteins using quantum logic. *Science* **351**, 836–841 (2016).
- Dolde, F. et al. Electric-field sensing using single diamond spins. *Nat. Phys.* **7**, 459–463 (2011).
- Kucsko, G. et al. Nanometre-scale thermometry in a living cell. *Nature* **500**, 54–58 (2013).
- Fukami, M. et al. All-optical cryogenic thermometry based on NV centers in nanodiamonds. *Phys. Rev. Appl.* **12**, 1–8 (2019).
- Glenn, D. R. et al. High-resolution magnetic resonance spectroscopy using a Solid-State spin sensor. *Nature* **555**, 351–354 (2018).
- Aslam, N. et al. Nanoscale nuclear magnetic resonance with chemical resolution. *Sci. (80–)*. **357**, 67–71 (2017).
- Jelezko, F. & Wrachtrup, J. Single defect centres in diamond: a review. *Phys. status solidi* **203**, 3207–3225 (2006).
- Kolesov, R. et al. Optical detection of a single rare-earth ion in a crystal. *Nat. Commun.* **3**, 1–7 (2012).
- Wolfowicz, G. et al. Vanadium spin qubits as telecom quantum emitters in silicon carbide. *Sci. Adv.* **6**, 1–8 (2020).
- Dibos, A., Raha, M., Phenicie, C. & Thompson, J. Atomic Source of Single Photons in the Telecom Band. *Phys. Rev. Lett.* **120**, 1–6 (2018).

24. Zhong, T. et al. Optically addressing single rare-earth ions in a nanophotonic cavity. *Phys. Rev. Lett.* **121**, 183603 (2018).
25. Weber, J. R. et al. Quantum computing with defects. *Proc. Natl Acad. Sci. USA* **107**, 8513–8518 (2010).
26. Markham, M. L. et al. CVD diamond for spintronics. *Diam. Relat. Mater.* **20**, 134–139 (2011).
27. Balasubramanian, G. et al. Ultralong spin coherence time in isotopically engineered diamond. *Nat. Mater.* **8**, 383–387 (2009).
28. Tyryshkin, A. M. et al. Electron spin coherence exceeding seconds in high-purity silicon. *Nat. Mater.* **11**, 143–147 (2012).
29. Ong, S. P. et al. 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**, 209–215 (2015).
30. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
31. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B Condens. Matter Mater. Phys.* **54**, 11169–11186 (1996).
32. Wang, L., Maxisch, T. & Ceder, G. Oxidation energies of transition metal oxides within the GGA+U framework. *Phys. Rev. B Condens. Matter Mater. Phys.* **73**, 195107 (2006).
33. Choudhary, K., Kalish, I., Beams, R. & Tavazza, F. High-throughput identification and characterization of two-dimensional materials using density functional theory. *Sci. Rep.* **7**, 1–16 (2017).
34. Thonhauser, T. et al. Van der Waals density functional: self-consistent potential and the nature of the van der Waals bond. *Phys. Rev. B Condens. Matter Mater. Phys.* **76**, 125112 (2007).
35. Klime, J., Bowler, D. R. & Michaelides, A. Van der Waals density functionals applied to solids. *Phys. Rev. B Condens. Matter Mater. Phys.* **83**, 195131 (2011).
36. Choudhary, K. et al. Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms. *Sci. Data* **5**, 1–12 (2018).
37. Calderon, C. E. et al. The AFLOW standard for high-throughput materials science calculations. *Comput. Mater. Sci.* **108**, 233–238 (2015).
38. Curtarolo, S. et al. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
39. Setyawan, W. & Curtarolo, S. High-throughput electronic band structure calculations: challenges and tools. *Comput. Mater. Sci.* **49**, 299–312 (2010).
40. Setyawan, W., Gaume, R. M., Lam, S., Feigelson, R. S. & Curtarolo, S. High-throughput combinatorial database of electronic band structures for inorganic scintillator materials. *ACS Comb. Sci.* **13**, 382–390 (2011).
41. Isayev, O. et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 1–12 (2017).
42. Perdew, J. P. Density functional theory and the band gap problem. *Int. J. Quantum Chem.* **28**, 497–523 (2009).

ACKNOWLEDGEMENTS

Funding for this research was provided by the AFOSR (contract FA9550-18-1-0334), the Eric and Wendy Schmidt Transformative Technology Fund at Princeton University and the Princeton Catalysis Initiative. The authors acknowledge helpful conversations with Chris Phenicie, Paul Stevenson, and Sacha Welinski.

AUTHOR CONTRIBUTIONS

J.D.T., N.P.dL., and R.J.C. devised the main conceptual ideas and guiding principles for the project. A.M.F. performed the data screening and analysis. R.J.C. supervised the research. A.M.F. wrote the paper with input from all authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41524-020-00391-7>.

Correspondence and requests for materials should be addressed to A.M.F.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020