



OPEN

DATA DESCRIPTOR


# A band-gap database for semiconducting inorganic materials calculated with hybrid functional

Sangtae Kim<sup>1,3</sup>, Miso Lee<sup>1,3</sup>, Changho Hong<sup>1</sup>, Youngchae Yoon<sup>1</sup>, Hyungmin An<sup>1</sup>, Dongheon Lee<sup>1</sup>, Wonseok Jeong<sup>1</sup>, Dongsun Yoo<sup>1</sup>, Youngho Kang<sup>2</sup>, Yong Youn<sup>1,4</sup>  & Seungwu Han <sup>1,4</sup> 

Semiconducting inorganic materials with band gaps ranging between 0 and 5 eV constitute major components in electronic, optoelectronic and photovoltaic devices. Since the band gap is a primary material property that affects the device performance, large band-gap databases are useful in selecting optimal materials in each application. While there exist several band-gap databases that are theoretically compiled by density-functional-theory calculations, they suffer from computational limitations such as band-gap underestimation and metastable magnetism. In this data descriptor, we present a computational database of band gaps for 10,481 materials compiled by applying a hybrid functional and considering the stable magnetic ordering. For benchmark materials, the root-mean-square error in reference to experimental data is 0.36 eV, significantly smaller than 0.75–1.05 eV in the existing databases. Furthermore, we identify many small-gap materials that are misclassified as metals in other databases. By providing accurate band gaps, the present database will be useful in screening materials in diverse applications.

## Background & Summary

The band gap ( $E_g$ ) is a fundamental quantity that directly relates to usability of materials in optical, electronic, and energy applications. For instance, in photovoltaic devices, materials with a direct  $E_g$  of  $\sim 1.3$  eV<sup>1,2</sup>, corresponding to the Shockley-Queisser limit, are favored as photo-absorbers that maximize the solar-cell efficiency. In power electronics, semiconductors with  $E_g \geq 3$  eV are employed to sustain high electric fields<sup>3</sup>. To increase the figure of merit in thermoelectric devices, materials with  $E_g$  of  $10 k_B T_{op}$  where  $k_B$  and  $T_{op}$  are the Boltzmann constant and operating temperature, respectively, are selected<sup>4</sup>. Given the central role of  $E_g$ , a database of  $E_g$  over a wide range of materials can expedite the material selection in specific applications by factoring out suboptimal candidates rapidly. Currently, popular material databases such as the Materials Project<sup>5</sup>, the Automatic Flow of Materials Discovery Library (AFLOW)<sup>6</sup>, the Open Quantum Materials Database (OQMD)<sup>7</sup>, and the Joint Automated Repository for Various Integrated Simulations (JARVIS)<sup>8</sup> provide theoretical  $E_g$  for up to one million inorganic materials. However, most of them were obtained by semilocal functionals with a generalized gradient approximation (GGA), which underestimates  $E_g$  by typically 30–40%<sup>9</sup>. (MatDB<sup>10</sup> provides accurate quasi-particle band gaps, but the number of data is limited to hundreds.) To compensate for this, AFLOW provides adjusted  $E_g$  using a linear fit to experimental data<sup>11</sup>. However, such a universal correction may not address element-dependent error fluctuations. We note that JARVIS provides  $E_g$  based on meta-GGA<sup>12</sup>, which significantly improves the accuracy. As a related issue, many small-gap semiconductors such as Ge, InAs, PdO, Zn<sub>3</sub>As<sub>2</sub>, and Ag<sub>2</sub>O are misclassified as metals, which can affect selection of narrow-gap semiconductors in IR sensors<sup>13</sup>, for instance. (In JARVIS, some of these errors are resolved by meta-GGA.) Besides the underestimation of  $E_g$ , all the databases consider only the ferromagnetic ordering for spin-polarized systems due to computational convenience, which can cause significant errors in  $E_g$  of antiferromagnetic materials. For instance, the antiferromagnetic NiO has an experimental  $E_g$  of 4.3 eV<sup>14</sup>, but the computational  $E_g$  ranges over 2.2–2.6 eV in the ferromagnetic ordering and GGA functional<sup>15–7</sup> while the correct antiferromagnetic ordering produces 4.5 eV within the hybrid functional.

<sup>1</sup>Department of Materials Science and Engineering and Research Institute of Advanced Materials, Seoul National University, Seoul, 08826, Korea. <sup>2</sup>Department of Materials Science and Engineering, Incheon National University, Incheon, 22012, Korea. <sup>3</sup>These authors contributed equally: Sangtae Kim, Miso Lee. <sup>4</sup>These authors jointly supervised this work: Yong Youn, Seungwu Han.  e-mail: [yybyb1@gmail.com](mailto:yybyb1@gmail.com); [hansw@snu.ac.kr](mailto:hansw@snu.ac.kr)

Addressing limitations in the existing material databases, we herein report a theoretical dataset of fundamental and optical  $E_g$  computed by employing a hybrid functional and identifying the stable magnetic ordering, thus providing more accurate  $E_g$  than the existing databases. For the high-throughput computational workflow, we employ ‘Automated *Ab initio* Modeling of Materials Property Package’ (AMP<sup>2</sup>)<sup>15</sup>, which is a fully automated package for density functional theory (DFT) calculations of crystalline properties. AMP<sup>2</sup> addresses the band-gap underestimation in semilocal functionals with the help of a hybrid functional, thereby producing a more accurate  $E_g$ , even if the material is incorrectly metallic within the semilocal functional. Furthermore, the package finds the antiferromagnetic ground state based on an effective Ising model. The present database focuses on materials with  $0\text{ eV} < E_g < 5\text{ eV}$ , which covers most semiconducting materials. The target materials are selected from Inorganic Crystal Structure Database (ICSD)<sup>16</sup> and partly filtered by information from the Materials Project database. In total, the database collects  $E_g$  for 10,481 materials that encompass most inorganic solids with  $E_g$  ranging between 0 and 5 eV. For 116 benchmark materials, the root-mean-square error (RMSE) with respect to experimental data is 0.36 eV, significantly smaller than 0.75–1.05 eV in the existing databases. The resulting data are available online at figshare<sup>17</sup> or SNUMAT<sup>18</sup>.

## Methods

**High-throughput methodology: AMP<sup>2</sup>.** The present database is constructed by employing AMP<sup>2</sup> which is an automation script operating VASP<sup>19–21</sup>. Starting only with the initial crystalline structure, AMP<sup>2</sup> provides band structure,  $E_g$ , effective mass, density of states (DOS) and dielectric constants of the crystal by automatically pipelining computational procedures. To summarize computational settings relevant in the present work, we employ GGA developed by Perdew-Burke-Ernzerhof (PBE)<sup>22</sup> for the exchange–correlation functional for structural relaxation and identifying band edges. The  $E_g$  is obtained by ‘one-shot’ hybrid functional (specifically, HSE06<sup>23</sup> (simply HSE hereafter)) calculations in which the package estimates  $E_g$  from HSE eigenvalues at  $\mathbf{k}$  points of band edges found with PBE (crystal structures are also fixed to those relaxed by PBE). In the previous study<sup>24</sup>, it was demonstrated that band edges from PBE and HSE lie at the same  $\mathbf{k}$  points, which is confirmed again in the present work with Si, SrS, BAs, BeS, AlAs, AgI, AgGaTe<sub>2</sub>, ZnSiAs<sub>2</sub>, and ZnIn<sub>2</sub>Se<sub>4</sub>. In addition, the small structural differences between PBE and HSE<sup>25</sup> would not affect the band gap significantly, except for small-gap semiconductors (see below). (This is also the case for systems that go from metallic in PBE to insulating in HSE.) This supports that the one-shot scheme can produce  $E_g$  close to the full hybrid calculations. If the material is identified as a metal within PBE, AMP<sup>2</sup> inspects DOS, and if DOS at the Fermi level normalized by the valence band ( $D_F/D_{VB}$ ) is less than a threshold, the package further tests a possible gap-opening by the one-shot hybrid calculation. The PBE+ $U$  method is applied on 3d orbitals<sup>26</sup> only when the material has a finite  $E_g$ . About pseudopotentials, we mostly employ those without any tags in the VASP database, which tends to reduce the number of valence electrons. For further details, we refer to the original publication<sup>15</sup>.

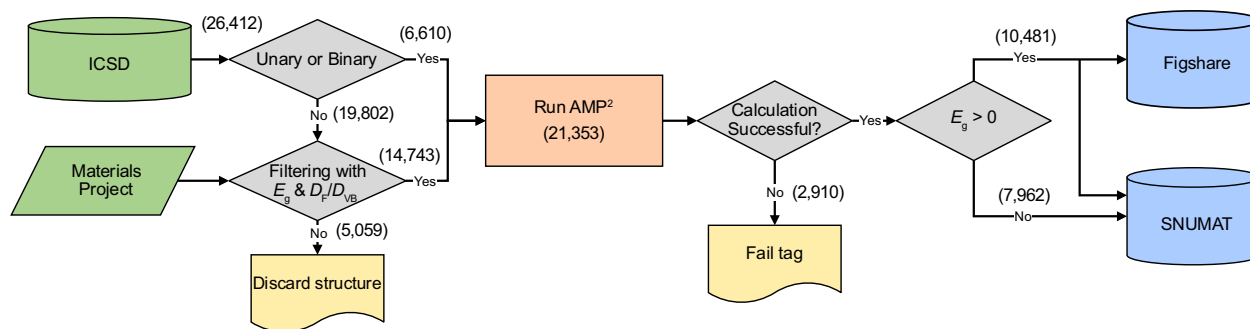
Computational parameters used in the present work follow the default setting of AMP<sup>2</sup> except that the package applies the PBE+ $U$  method on Ce 4f levels with the  $U$  value of 4 eV<sup>27</sup>. (The pseudopotentials for La and Ce treat f levels as valence.) Furthermore, for compounds including Tl, Pb, and Bi,  $E_g$  is recalculated by including the spin-orbit coupling (SOC) when the default  $E_g$  without SOC is smaller than 1 eV. (The band edges are also resought with SOC.) This is because typical SOC corrections of  $\sim 0.5\text{ eV}$  would be critical in these cases.

In identifying the stable collinear magnetic ordering, AMP<sup>2</sup> applies a genetic algorithm to the Ising model<sup>28</sup>. This approach finds the stable magnetic ordering correctly for many compounds. However, the original formulation requires a large supercell to isolate exchange interactions from periodic images, which costs significant computational resources and also suffers from ill-convergence in electronic iterations. To resolve this, we here develop an alternative method in obtaining exchange parameters. First, we choose a minimal supercell under the following two conditions: (i) A magnetic site  $\alpha$  and its periodic images in other supercells are apart more than 5 Å (cutoff range for magnetic interactions). (ii) If two magnetic sites  $\alpha$  and  $\beta$  (not necessarily belong to the same supercell) are within 5 Å, then the distance between  $\alpha$  and  $\beta'$ , a periodic image of  $\beta$  ( $\beta' \neq \beta$ ) is longer than 5 Å except when  $\alpha$ - $\beta$  and  $\alpha$ - $\beta'$  are symmetrically equivalent. Within the Ising model, the total energy of the supercell ( $E$ ) can be expressed as follows:

$$E = E_0 + \sum_I^m J_I(N_{I,P} - N_{I,A}), \quad (1)$$

where  $E_0$  is the base energy excluding the magnetic interaction, and  $I$  is the index for independent exchange interactions (total  $m$  interactions) with the maximum range of 5 Å and the exchange parameter of  $J_I$ . In Eq. (1),  $N_{I,P}$  and  $N_{I,A}$  are the numbers of parallel and antiparallel spin pairs within the supercell corresponding to the interaction  $I$ , respectively. Then, based on the ferromagnetic configuration (all spin-up), diverse spin configurations are sampled by spin-flipping a magnetic pair (both atoms) or a certain magnetic site. The number of resulting equations is larger than  $m$  and an optimal  $\{E_0, J_I\}$  can be obtained by the pseudoinverse method. We find that this approach produces essentially the same parameters as the original scheme but is more reliable and efficient.

**Selection of materials.** Figure 1 schematizes the workflow of constructing the database. Starting from the ICSD, we only consider compounds consisting of elements with atomic number ( $Z$ ) < 84. Among the lanthanides, we limit the elements to La and Ce. We remove structural duplicates and structures with partially occupied sites, and also omit large primitive cells that contain more than 40 atoms. For unary and binary compounds, all the structures are calculated with AMP<sup>2</sup>. For ternary and higher compounds, we utilize information on  $E_g$  and DOS in the Materials Project database (calculated by PBE) to filter out materials that are likely to be metallic or large-gap insulators. To be specific, we exclude materials with  $E_g^{\text{GGA}}$  bigger than 3 eV since they are likely to have  $E_g^{\text{HSE}}$  larger than 5 eV. (Compiling data of 4,421 compounds from the previous screening studies<sup>24,29–31</sup>, we find



**Fig. 1** The computational workflow for collecting the dataset. Numbers in the parentheses indicate material counts.

Key	Type	Description
SNUMAT_id	string	ID in the SNUMAT
ICSD_number	int	ICSD collection code
Band_gap_GGA	float	Calculated fundamental band gap in GGA (eV)
Band_gap_GGA_optical	float	Calculated direct band gap in GGA (eV)
Band_gap_HSE	float	Calculated fundamental band gap in HSE (eV)
Band_gap_HSE_optical	float	Calculated direct band gap in HSE (eV)
Direct_or_indirect	string	Type of band gap in GGA (direct or indirect)
Direct_or_indirect_HSE	string	Type of band gap in HSE (direct or indirect)
Structure_rlx	string	Relaxed structure information (VASP POSCAR format)
Space_group_rlx	int	Space group number of relaxed structure
Magnetic_ordering	string	Magnetic ordering of final structure
SOC	boolean	Spin-orbit coupling (True or False)

**Table 1.** Description of metadata keys in JSON file.

that 99.7% of materials with  $E_g^{\text{HSE}} < 5$  eV have  $E_g^{\text{GGA}} < 3$  eV.) We also include metallic materials with  $D_F/D_{VB} < 0.8$  for possible gap opening (see above; a larger threshold is used because of low-resolution DOS in the Materials Project). If a Materials Project data has incomplete entries for  $E_g$  or DOS, the material is included in the computation list. In this way, we could factor out 5,059 materials from the list of ternary and higher compounds. Finally, we calculate 21,353 materials with AMP<sup>2</sup>. After computation, we collect 10,481 materials with finite  $E_g$  (unary: 63, binary: 1,919, ternary: 5,074, quaternary: 2,804, quinary: 573, and higher: 48).

### Data Records

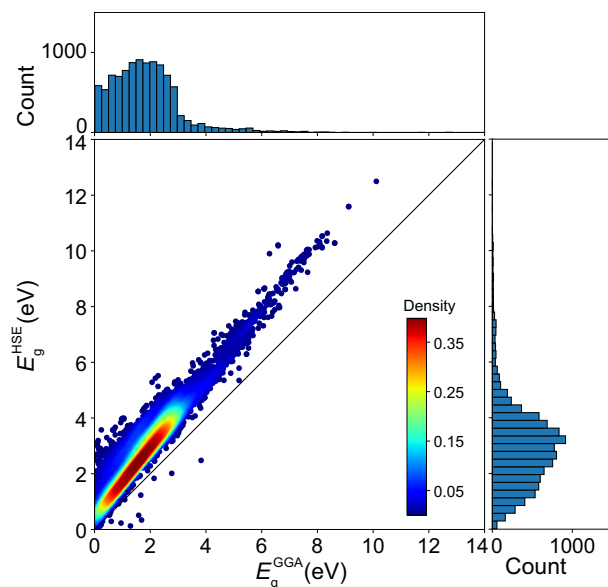
All the calculated properties for 10,481 compounds can be downloaded from the Figshare Repository<sup>17</sup>. The whole data including metals are also uploaded to SNUMAT ([www.snumat.com](http://www.snumat.com)), which provides easy search and visualization of materials through its own interactive interface. SNUMAT also supports REST API<sup>32</sup> for users to search the materials with authorization. The authorization token expires 24 hours after they are issued.

**File format.** The data are stored in the JSON format. The name of the file is  $X\_ICSD\#.json$ , where  $X$  is chemical formula and  $ICSD\#$  is the ICSD collection code of the initial structure used for calculation. Each JSON file includes final relaxation structure information,  $E_g^{\text{GGA}}$ ,  $E_g^{\text{HSE}}$ , and DOS. Table 1 summarizes keys for metadata.

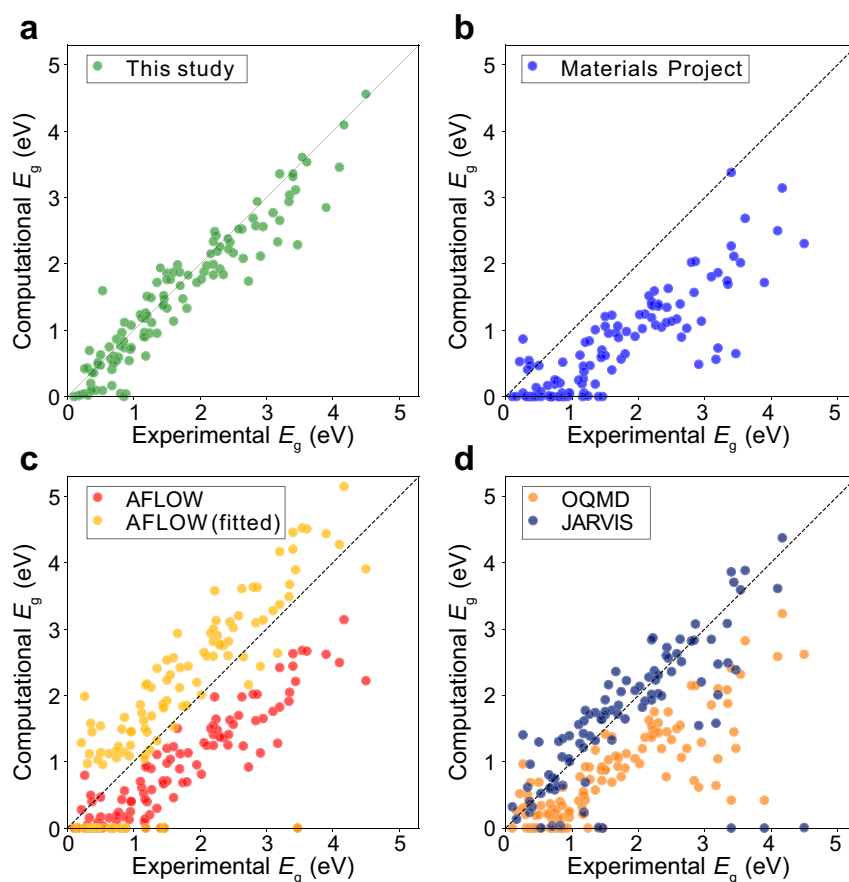
**Graphical representation of the data.** In Fig. 2, we present the distribution of  $E_g^{\text{GGA}}$  and  $E_g^{\text{HSE}}$  for 10,481 materials. Most materials with  $E_g^{\text{HSE}} > 5$  eV (663 cases) are unary or binary compounds for which AMP<sup>2</sup> is applied to the whole structure dataset from ICSD.

### Technical Validation

**Comparison to experimental measurements and other databases.** In Fig. 3a, we compare experimental and theoretical values for 116 benchmark materials with experimental  $E_g$  between 0 and 5 eV. The list of compounds is shown in Online-only Table 1. For comparison, theoretical results from other databases are also shown in Fig. 3b–d. The RMSE values are 0.36 eV (present work), 1.05 eV (Materials Project), 0.93 eV (AFLOW), 0.75 eV (AFLOW fitted), 1.02 eV (OQMD), and 0.85 eV (JARVIS). (The meta-GGA values of 19 materials, mostly with small  $E_g$ , are missing in JARVIS.) This confirms that the present database provides more accurate  $E_g$  than the existing databases on average. In particular, we correctly identify the semiconducting nature for small-gap semiconductors such as AgSbTe<sub>2</sub>, CdO, CoP<sub>3</sub>, Cu<sub>3</sub>AsSe<sub>4</sub>, Cu<sub>3</sub>SbS<sub>4</sub>, Cu<sub>3</sub>SbSe<sub>4</sub>, CuFeS<sub>2</sub>, Ge, Mg<sub>2</sub>Sn, RhSb<sub>3</sub>, and ZnSnSb<sub>2</sub>, which are mostly misreported as metals in other databases. In addition, other databases exhibit pronounced errors for every antiferromagnetic material (CuFeS<sub>2</sub>, CuO, FeF<sub>2</sub>, MnO, MnTe, and NiO) because these



**Fig. 2** Distribution of  $E_g^{\text{GGA}}$  and  $E_g^{\text{HSE}}$ . Top and right are occurrence histograms of  $E_g^{\text{GGA}}$  and  $E_g^{\text{HSE}}$ , respectively.



**Fig. 3** Comparison of  $E_g$  for benchmark materials between experimental and theoretical data from (a) this work, (b) Materials Project, (c) AFLOW and (d) OQMD and JARVIS. AFLOW-fitted values are obtained from  $E_g^{\text{fit}} = 1.34 E_g^{\text{GGA}} + 0.913 \text{ eV}$ .

materials are considered as ferromagnetic or non-magnetic. (For non-magnetic materials in Online-only Table 1, the  $E_g$  calculated with pure PBE (without  $+U$  and SOC) by AMP<sup>2</sup> agrees well with those from Materials Project (the mean absolute error is 0.034 eV).)

In most cases, the present database provides  $E_g$  that agrees well with experiment. However, there are some materials with large errors of  $\geq 0.5$  eV such as  $\text{AgAlTe}_2$ ,  $\text{Cu}_3\text{AsSe}_4$ ,  $\text{CuAlSe}_2$ ,  $\text{CuBr}$ ,  $\text{CuCl}$ ,  $\text{CuFeS}_2$ ,  $\text{CuO}$ ,  $\text{Ge}$ ,  $\text{IrSb}_3$ ,  $\text{La}_2\text{S}_3$ ,  $\text{MnO}$ ,  $\text{RhAs}_3$ ,  $\text{RhSb}_3$ ,  $\text{SnO}_2$ ,  $\text{SrS}$ , and  $\text{ZnO}$ . For small-gap materials such as  $\text{Cu}_3\text{AsSe}_4$ ,  $\text{Ge}$ , and  $\text{IrSb}_3$ ,  $E_g$  is sensitive to the lattice parameters that are slightly overestimated by PBE. Employing experimental lattice parameters or those relaxed within HSE significantly improves the results<sup>15</sup>. For Cu-bearing materials, it is known that HSE often exhibits substantial errors in  $E_g$  due to nonlocal screening effects in Cu, which requires GW calculations<sup>33,34</sup>. We also note that van der Waals interactions are not described by semilocal functionals, and lattice parameters can be overestimated in layered structures such transition-metal dichalcogenides<sup>35</sup>. This can significantly affect  $E_g$ , and so care is needed in referring to  $E_g$  in layered materials. The present results do not consider finite-temperature effects on  $E_g$ , which can be significant in some materials, for example, hybrid perovskites<sup>36</sup>. More generally,  $E_g$  dataset with the ultimate theoretical accuracy would be obtained by the quasiparticle approaches such as GW or Bethe-Salpeter equations<sup>37,38</sup>.

## Code availability

The AMP<sup>2</sup> package used for constructing the present database is available at <https://github.com/MDIL-SNU/AMP2> and was released under a GPLv3 (GNU General Public License). The package requires pre-installation of numpy, scipy, spglib, and PyYAML modules. Detailed guidelines and examples can be found in the manual (<https://amp2.readthedocs.io/en/latest/>).

Received: 21 August 2020; Accepted: 6 October 2020;

Published online: 11 November 2020

## References

- Shockley, W. & Queisser, H. J. Detailed Balance Limit of Efficiency of p-n Junction Solar Cells. *J. Appl. Phys.* **32**, 11, <https://doi.org/10.1063/1.1736034> (1961).
- Kang, Y., Youn, Y., Han, S., Park, J. & Oh, C.-S. Computational Screening of Indirect-Gap Semiconductors for Potential Photovoltaic Absorbers. *Chem. Mater.* **31**, 4072–4080, <https://doi.org/10.1021/acs.chemmater.9b00708> (2019).
- Gorai, P., McKinney, R. W., Haegel, N. M., Zakutayev, A. & Stevanovic, V. A computational survey of semiconductors for power electronics. *Energy & Environ. Sci.* **12**, 3338–3347, <https://doi.org/10.1039/C9EE01529A> (2019).
- Sofo, J. O. & Mahan, G. D. Optimum band gap of a thermoelectric material. *Phys. Rev. B* **49**, 4565–4570, <https://doi.org/10.1103/PhysRevB.49.4565> (1994).
- Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002, <https://doi.org/10.1063/1.4812323> (2013).
- Curtarolo, S. *et al.* AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235, <https://doi.org/10.1016/j.commatsci.2012.02.002> (2012).
- Saal, J. E., Kirklín, S., Aykol, M., Meredig, B. & Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **65**, 1501–1509, <https://doi.org/10.1007/s11837-013-0755-4> (2013).
- Choudhary, K. *et al.* Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms. *Sci. Data* **5**, 180082, <https://doi.org/10.1038/sdata.2018.82> (2018).
- Park, S., Lee, B., Jeon, S. H. & Han, S. Hybrid functional study on structural and electronic properties of oxides. *Curr. Appl. Phys.* **11**, S337–S340, <https://doi.org/10.1016/j.cap.2010.09.008> (2011).
- Lany, S. Band-structure calculations for the 3d transition metal oxides in GW. *Phys. Rev. B* **87**, 085112, <https://doi.org/10.1103/PhysRevB.87.085112> (2013).
- 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, <https://doi.org/10.1021/co200012w> (2011).
- Tran, F. & Blaha, P. Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange–Correlation Potential. *Phys. Rev. Lett.* **102**, 226401, <https://doi.org/10.1103/PhysRevLett.102.226401> (2009).
- Seiler, D. G. C. L. L. International conference on narrow-gap semiconductors and related materials. *J. Res. Natl. Inst. Standards Technol.* **95**, 13, <https://doi.org/10.6028/jres.095.037> (1990).
- Sawatzky, G. A. & Allen, J. W. Magnitude and Origin of the Band Gap in NiO. *Phys. Rev. Lett.* **53**, 2339–2342, <https://doi.org/10.1103/PhysRevLett.53.2339> (1984).
- Youn, Y. *et al.* AMP<sup>2</sup>: A fully automated program for ab initio calculations of crystalline materials. *Comput. Phys. Commun.* **256**, 107450, <https://doi.org/10.1016/j.cpc.2020.107450> (2020).
- FIZ Karlsruhe Inorganic Crystal Structure Database. <https://icsd.products.fiz-karlsruhe.de/>. 16.02 version.
- Kim, S. *et al.* A band-gap database for semiconducting inorganic materials calculated with hybrid functional. *figshare* <https://doi.org/10.6084/m9.figshare.12839240.v5> (2020).
- SNUMAT: SNU material data center. <https://www.snumat.com/>. Accessed: 2020-08-17.
- Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50, [https://doi.org/10.1016/0927-0256\(96\)00008-0](https://doi.org/10.1016/0927-0256(96)00008-0) (1996).
- Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186, <https://doi.org/10.1103/PhysRevB.54.11169> (1996).
- Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979, <https://doi.org/10.1103/PhysRevB.50.17953> (1994).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868, <https://doi.org/10.1103/PhysRevLett.77.3865> (1996).
- Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. The. *J. Chem. Phys.* **118**, 8207–8215, <https://doi.org/10.1063/1.1564060> (2003).
- Yim, K. *et al.* Novel high-k dielectrics for next-generation electronic devices screened by automated ab initio calculations. *NPG Asia Mater.* **7**, e190, <https://doi.org/10.1038/am.2015.57> (2015).
- Heyd, J., Peralta, J. E., Scuseria, G. E. & Martin, R. L. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. The. *J. Chem. Phys.* **123**, 174101, <https://doi.org/10.1063/1.2085170> (2005).
- Wang, L., Maxisch, T. & Ceder, G. Oxidation energies of transition metal oxides within the GGA+U framework. *Phys. Rev. B* **73**, 195107, <https://doi.org/10.1103/PhysRevB.73.195107> (2006).
- Da Silva, J. L. F., Ganduglia-Pirovano, M. V., Sauer, J., Bayer, V. & Kresse, G. Hybrid functionals applied to rare-earth oxides: The example of ceria. *Phys. Rev. B* **75**, 045121, <https://doi.org/10.1103/PhysRevB.75.045121> (2007).



28. Lee, K., Youn, Y. & Han, S. Identification of ground-state spin ordering in antiferromagnetic transition metal oxides using the Ising model and a genetic algorithm. *Sci. Technol. Adv. Mater.* **18**, 246–252, <https://doi.org/10.1080/14686996.2017.1300046> (2017).
29. Yim, K. *et al.* Computational discovery of p-type transparent oxide semiconductors using hydrogen descriptor. *npj Comput. Mater.* **4**, 17, <https://doi.org/10.1038/s41524-018-0073-z> (2018).
30. Lee, M., Youn, Y., Yim, K. & Han, S. High-throughput ab initio calculations on dielectric constant and band gap of non-oxide dielectrics. *Sci. Reports* **8**, 14794, <https://doi.org/10.1038/s41598-018-33095-6> (2018).
31. Youn, Y. *et al.* Large-Scale Computational Identification of p-Type Oxide Semiconductors by Hierarchical Screening. *Chem. Mater.* **31**, 5475–5483, <https://doi.org/10.1021/acs.chemmater.9b00816> (2019).
32. Fielding, R. T. Principled Design of the Modern Web Architecture. *ACM Transactions on Internet Technol.* **2**, 36, <https://doi.org/10.1145/514183.514185> (2002).
33. Wang, Y. *et al.* Electronic structures of Cu<sub>2</sub>O, Cu<sub>4</sub>O<sub>3</sub>, and CuO: A joint experimental and theoretical study. *Phys. Rev. B* **94**, 245418, <https://doi.org/10.1103/PhysRevB.94.245418> (2016).
34. van Veenendaal, M. A., Eskes, H. & Sawatzky, G. A. Strong nonlocal contributions to Cu 2p photoelectron spectroscopy. *Phys. Rev. B* **47**, 11462–11469, <https://doi.org/10.1103/PhysRevB.47.11462> (1993).
35. Ramasubramaniam, A., Naveh, D. & Towe, E. Tunable band gaps in bilayer transition-metal dichalcogenides. *Phys. Rev. B* **84**, 205325, <https://doi.org/10.1103/PhysRevB.84.205325> (2011).
36. Wiktor, J., Rothlisberger, U. & Pasquarello, A. Predictive determination of band gaps of inorganic halide perovskites. *Phys. Chem. Lett.* **8**, 5507, <https://doi.org/10.1021/acs.jpcclett.7b02648> (2017).
37. Schilfgaarde, M. V., Kotani, T. & Faleev, S. Quasiparticle self-consistent GW theory. *Phys. Rev. Lett.* **96**, 226402, <https://doi.org/10.1103/PhysRevLett.96.226402> (2006).
38. Rohlfing, M. & Louie, S. G. Electron-hole excitations and optical spectra from first principles. *Phys. Rev. B* **62**, 4927, <https://doi.org/10.1103/PhysRevB.62.4927> (2000).
39. *CRC Handbook of Chemistry and Physics, 86th Edition Edited by David R. Lide.* (CRC Press, 2005).
40. Tell, B., Shay, J. & Kasper, H. Some properties of AgAlTe<sub>2</sub>, AgGaTe<sub>2</sub>, and AgInTe<sub>2</sub>. *Phys. Rev. B* **9**, 5203–5208, <https://doi.org/10.1103/PhysRevB.9.5203> (1974).
41. Ashrafi, A. B. M. A., Kumano, H., Suemune, I., Ok, Y.-W. & Seong, T.-Y. Single-crystalline rocksalt CdO layers grown on GaAs (001) substrates by metalorganic molecular-beam epitaxy. *Appl. Phys. Lett.* **79**, 470–472, <https://doi.org/10.1063/1.1387258> (2001).
42. Shay, J. L., Buehler, E. & Wernick, J. H. Electroreflectance Study of the Energy-Band Structure of CdSnP<sub>2</sub>. *Phys. Rev. B* **2**, 4104–4109, <https://doi.org/10.1103/PhysRevB.2.4104> (1970).
43. Robbins, M., Phillips, J. C., Lambrecht, V. G., Laboratories, B. & Hill, M. Solid solution formation in the systems CuM<sup>III</sup>X<sub>2</sub>-AgM<sup>III</sup>X<sub>2</sub> where M<sup>III</sup>=Al, Ga, In and X<sub>2</sub>=S, Se. *J. Phys. Chem. Solids* **34**, 5, [https://doi.org/10.1016/S0022-3697\(73\)80210-0](https://doi.org/10.1016/S0022-3697(73)80210-0) (1973).
44. Benchouk, K. *et al.* New buffer layers, large band gap ternary compounds: CuAlTe<sub>2</sub>. *The Eur. Phys. J. Appl. Phys.* **10**, 9–14, <https://doi.org/10.1051/epjap:2000114> (2000).
45. Radautsan, S. I. & Tiginyanu, I. M. Defect Engineering in II–III<sub>2</sub>–VI<sub>4</sub> and Related Compounds. *Jpn. J. Appl. Phys.* **32**, 5, <https://doi.org/10.7567/JJAPS.32S3.5> (1993).
46. Davydov, V. Y., Ivanov, S. V. & Mudryi, A. V. Absorption and Emission of Hexagonal InN. Evidence of Narrow Fundamental Band Gap. *Phys. Status Solidi (b)* **229**, 3, [https://doi.org/10.1002/1521-3951\(200202\)229:3<R1::AID-PSSB99991>3.0.CO;2-O](https://doi.org/10.1002/1521-3951(200202)229:3<R1::AID-PSSB99991>3.0.CO;2-O) (2002).
47. Goubin, F. *et al.* Experimental and Theoretical Characterization of the Optical Properties of CeO<sub>2</sub>, SrCeO<sub>3</sub>, and Sr<sub>2</sub>CeO<sub>4</sub> Containing Ce<sup>4+</sup> (F<sup>0</sup>) Ions. *Chem. Mater.* **16**, 662–669, <https://doi.org/10.1021/cm034618u> (2004).
48. Forster, C. M. & White, W. B. Optical absorption edge in rare earth sesquisulfides. *Mater. Res. Bull.* **41**, 448–454, <https://doi.org/10.1016/j.materresbull.2005.07.035> (2006).

## Acknowledgements

This work was supported by Nano-Material Technology Development Program through the National Research Foundation of Korea (NRF-2020M3A7C2095492) and Creative Materials Discovery Program through the NRF funded by the Ministry of Science and ICT (2017M3D1A1040689). The computations were carried out at Korea Institute of Science and Technology Information (KISTI) National Supercomputing Center (KSC-2020-CRE-0064).

## Author contributions

S.K., M.L., C.H., H.A., D.L., W.J., D.Y. and Y. Youn performed the high-throughput calculations and data collections. Y. Yoon constructed the SNUMAT site. S.K., M.L., Y.K., Y. Youn, and S.H. analyzed the data and wrote the manuscript. S.H. coordinated the whole work. All authors reviewed and commented on the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Yong Youn or S.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](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 Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020