

# A new MaterialGo database and its comparison with other high-throughput electronic structure databases for their predicted energy band gaps

JIE JianShu<sup>1</sup>, WENG MouYi<sup>1</sup>, LI ShunNing<sup>1</sup>, CHEN Dong<sup>1</sup>, LI ShuCheng<sup>1</sup>, XIAO WeiJi<sup>1</sup>, ZHENG JiaXin<sup>1</sup>, PAN Feng<sup>1\*</sup> & Wang LinWang<sup>2\*</sup>

<sup>1</sup> School of Advanced Materials, Peking University Shenzhen Graduate School, Shenzhen 518055, China;

<sup>2</sup> Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley 94720, USA

Received February 3, 2019; accepted April 19, 2019; published online June 5, 2019

Recently, many high-throughput calculation materials databases have been constructed and found wide applications. However, a database is only useful if its content is reliable and sufficiently accurate. It is thus of paramount importance to gauge the reliabilities and accuracies of these databases. Although many properties have been predicted accurately in these databases, electronic band gap is well known to be underestimated by traditional density functional theory (DFT) calculations under local density approximation (LDA), which becomes a challenging problem for materials database building. Here, we introduce MaterialGo (<http://www.pkusam.com/data-base.html>), a new database calculating the band structures of crystals using both Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional and Heyd-Scuseria-Ernzerhof (HSE) hybrid functional. Comparing different PBE databases, it is found that their band gaps are consistent when no  $U$  parameter is used for transition metal d-state or heavy element f-state to correct their self-interaction error, but rather different when PBE+ $U$  are used, mostly because of the different values of  $U$  used in different database. HSE calculations under standard parameters will give larger band gaps that are closer to experiment. Based on the high-throughput HSE calculations over 10000 crystal structures, we might have a better understanding of the relationship between crystal structures and electronic structures, which will help us to further explore material genome science and engineering.

**high-throughput, hybrid functional, calculation database**

**Citation:** Jie J S, Weng M Y, Li S N, et al. A new MaterialGo database and its comparison with other high-throughput electronic structure databases for their predicted energy band gaps. *Sci China Tech Sci*, 2019, 62: 1423–1430. <https://doi.org/10.1007/s11431-019-9514-5>

## 1 Introduction

With the recent progress of machine learning techniques, big data are becoming a major tool for material research [1–3]. High-throughput *ab initio* calculations have fueled the recent frenzy for material science database building. While previous databases like Inorganic Crystal Structure Database (ICSD) [4,5], The National Institute of Standards and Technology (NIST) database (<https://www.nist.gov/data>),

American Mineralogist Crystal Structure Database [6], just to name a few, are mostly built on top of experimental results over decades, materials databases built in the last few years are based on *ab initio* calculations. The maturity of *ab initio* methods, especially the ones for plane wave density functional theory (DFT), the availability of high throughput computation hardware and automatic workflow framework made the rapid developments of such databases possible. The framework development for automatic calculation and data storage is important. For example, AiiDA (Automated Interactive Infrastructure and Database for Computational

\*Corresponding authors (email: [panfeng@pkusz.edu.cn](mailto:panfeng@pkusz.edu.cn); [lwwang@lbl.gov](mailto:lwwang@lbl.gov))

Science) [7] offers a framework with which one can easily manage the calculation and perform the analysis and storage of the results. The materials database building effort was originally initialized by the Materials Genome Initiative (MGI) by U.S. government started in 2011 (<https://www.mgi.gov/>). The MGI resulted in several widely used high throughput databases like Materials Project (MP) [8,9], Automatic-FLOW for Materials Discovery (AFLOW) [10–12] and Open Quantum Materials Database (OQMD) [13]. While OQMD focuses on structural information and phase diagram of compounds, Materials Project and AFLOW provide a wider array of properties, like binding energy, phase diagram and energy band gap. Each of these databases has tens of thousands of materials data entries. In Europe, the newly developed Novel Materials Discovery (NOMAD) project (<https://www.nomad-coe.eu/>) maintains a repository to store detailed input and output file during calculation and call for inputs from the whole electronic structure community. In China, the Chinese Material Genome Engineering (MGE) has been started two years ago, resulting many works on high-throughput computation, machine learning and big data [14–17].

Given the wide use of all these databases, it is thus imperative to know the reliability of these databases. Although theoretically speaking, data in all these databases should be the same since they are all generated using DFT and plane wave DFT codes, in reality different calculation details and procedures have made them different. One major factor is atomic positions. For a given compound, although most databases take the initial atomic structures from experimental results, atomic relaxations can lead to different structures. Another major factor is the use of PBE+ $U$  method [18]. The  $U$  parameter is often used for transition metal  $d$ -state or heavy element  $f$ -state to correct their self-interaction error. In practice, the  $U$  is used as an empirical parameter. However, for a given system, some databases use  $U$  while some others do not use  $U$ . For example, MP only uses  $U$  parameter for a small set of transition metals, while AFLOW uses  $U$  also for some p-block metals and Lanthanides. Even when both databases use PBE+ $U$ , the  $U$  values can be different. For this purpose, a cross-database analysis will be very useful. It will provide guidance for the users in regards to the reliability and accuracy of these databases. It will also be useful to compare these databases with results from more accurate methods like HSE [19]. For this purpose, we will include our own new database which uses HSE to calculate the electronic structure. Lastly, it is important to compare the calculated database results with experiments when the experimental values are available.

In this work, we introduce MaterialGo (MG), a new database calculating the band structures of crystals not only by PBE but also by HSE method. Then we compare the band gap predictions of MP, AFLOW, and MG databases. We

focus on band gap since this is a very difficult aspect of the current electronic structure calculations and it is useful for a wide range of applications, from solar cells to light emission diodes [20,21]. For the HSE functional, there are many previous works [22,23] compared its band gaps to experiments, but they are done in relatively small sets of materials. Here, we present HSE band gaps for tens of thousands of systems by high-throughput calculation of electronic structure. Through the comparison, we would like to address the following questions: (1) how large are the band gap differences between different databases? (2) what is the cause of these band gap differences: atomic structures, PBE+ $U$  parameters, or some other factors? (3) how much does the HSE improve band gap calculations? (4) what is the general error of the band gaps compared with experimental values?

## 2 Calculation workflow and method

### 2.1 The high-throughput workflow

As a database focusing on electronic structures, during our calculations for MaterialGo, crystal atomic structures are taken from MP database without further relaxations. These are the DFT optimized structures in MP starting from the initial experimental structures. Input files are generated according to structure files by scripts, and then batches of jobs are submitted to a PBS queueing system. Other scripts are used to automatically check whether a calculation is successful, re-adjust input parameters for unsuccessful jobs, extract useful information from finished jobs, and show the results on web.

Our database calculations are carried out using the PWmat [24,25] code. For the treatment of the exchange-correlation effect, both PBE and HSE06 are used. NCPP-SG15-PBE pseudopotential [26,27] is used for both PBE and HSE06 calculation. A plane wave basis set with a cutoff of 50 Ryd is used. The Brillouin Zone  $k$ -meshes are generated by Monkhorst-Pack method with the criterion that the number of  $k$ -points multiplied by the number of atoms is roughly around 1000.

The most time-consuming step is the calculation of the hybrid functional under plane wave basis, as it needs many Fourier transformations to carry out the Fock exchange integral. The recently proposed ACE [28] has been used to accelerate the hybrid functional self-consistent calculation. This method has been implemented in the PWmat code, and the GPU usage of the PWmat code also speeds up its calculation significantly. Due to such advances, it becomes practical to use hybrid functional in high throughput calculations. In the part of hybrid functionals, there are many versions of the hybrid method with different mixing parameters between the Fock exchange term and the semilocal exchange functional. We have decided to use the original

HSE06 parameter of Krukau, Vydrov, Izmaylov and Scuseria [29],  $\alpha=0.25$  and  $\omega=0.1058$ , which is mostly designed for relatively low band gap semiconductors.

To test the accuracy of HSE in PWmat, calculated band gaps are validated with the SC/40 test set as proposed by Heyd et al. [30] (Figure S1 (Supporting Information)). As one can see, the PWmat calculated HSE06 band gaps agree well with those reported in original literature, where localized orbitals are used. Such agreement verifies the correctness of PWmat HSE calculations. The speed of our workflow can be shown in the following data: for structures that have MP ID less than 21000, exactly 2000 have been calculated successfully. These structures have in total 15553 atoms, and calculations take 18279125 s (approximately 212 days) in total. So that an average of about 2.5 h is needed for a structure with an average of about 8 atoms, covering both PBE and HSE calculations.

## 2.2 The presentation of our data

All the data which are calculated by our workflow are open in public in our website (<http://www.pkusam.com/data-base.html>). One can click the name of elements from the periodic table to search materials (Figure 1). The webpage has been linked to our database. After selecting a name of one material, 3D atomic structures, cell parameters, electronic band structures are listed in the webpage. For each crystal, we have also provided a link to the original database (the Materials Project) from which we downloaded the atomic structures from. All input files in PWmat format can also be downloaded from our website. Graphics which can be zoomed in and out of band structures and density of states calculated by HSE and PBE are also shown in our web site separately.

## 3 Results and discussion

### 3.1 MG database

More than 10000 structures have been calculated so far in MG. The distribution of calculated structures is illustrated in Figure 2. In Figure 2, a structure is classified into alloy if it contains only metals in its chemical formula. Similarly, oxides contain O and any number of metals, halides contain one of F, Cl, Br, I and any number of metals, chalcogenides contain one of S, Se, Te and any number of metals, oxo and chalcogenides contain one of O, S, Se, Te, one of F, Cl, Br, I and any number of metals, oxysalts contain O, one of B, C, N, Si, P, S, As, Se, Te and any number of metals.

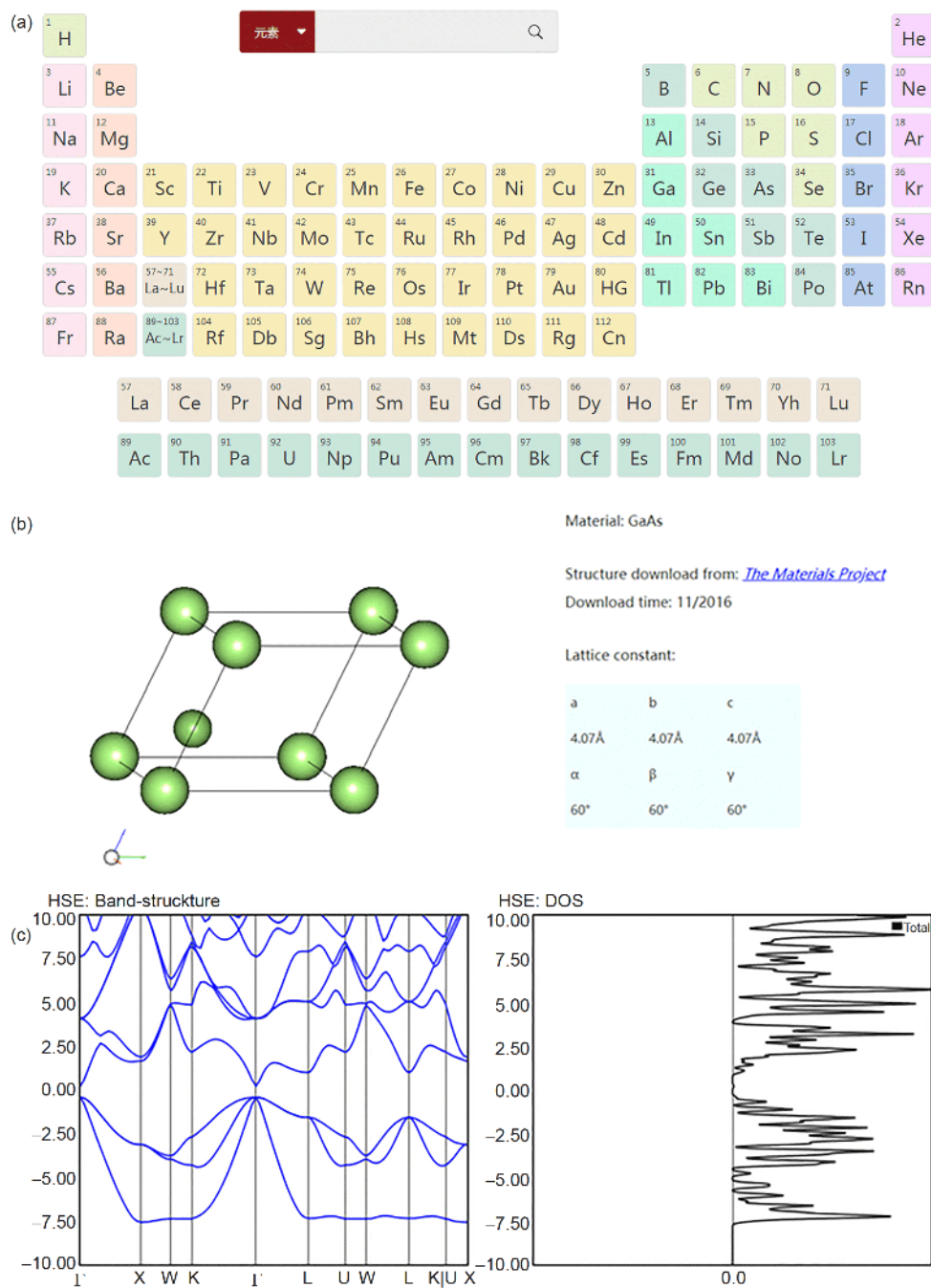
### 3.2 Comparison between MP, AFLOW and MG all with PBE calculations

The first comparison we made is between different PBE-

related database results, namely MP, AFLOW and PBE part of MG. MG uses no  $U$  parameters for PBE calculation, while the MP and AFLOW use quite different  $U$  values, as described in Table S1 (Supporting Information). These three databases also vary in the choice of other calculation parameters (Table S2 and Sect. 2.1). While the MG is calculated with the PWmat code with the optimized Vanderbilt norm conserving (OVNC) pseudopotential NCPP-SG15-PBE, both AFLOW and MP are calculated with the VASP code using augmented plane wave (PAW) pseudopotential. The PWmat yields the result almost the same as the PWscf calculation, and NCPP-SG15-PBE generally can give a band gap very close to that of PAW result. Thus, a major source of difference might come from the difference of using  $U$ . In Figure 3, we have shown the comparison with different categories of the  $U$  usages.

In Figure 3(a)–(c), one small point corresponds to one crystal structure. In the comparison, there are about 10000 crystal structures used. We first notice that there are structures for which one database predicts their band gaps as zero (metals), while the other database predicts their band gaps as nonzero (insulators). It is worth knowing what causes such dramatic differences. For this, we have divided the comparison into three different cases, and listed the results in Table 1. One case is both databases use  $U$ ; another case is only one database uses  $U$ ; and the third case is none of the databases use  $U$ . We see that, if none of them use  $U$ , the metal/insulator assignment between two data bases are much closer, only about 3% to 7% different assignment. On the other hand, if only one data base uses  $U$ , there could be 27% different assignment. For the cases, which both use  $U$  (MP vs AFLOW), the difference is 15%. This shows the significant uncertainty caused by the use of  $U$ . We also see that, in terms the metal/insulator assignment, the AFLOW/MG-PBE has the smallest difference, while MP/MG-PBE and MP/AFLOW has similar differences.

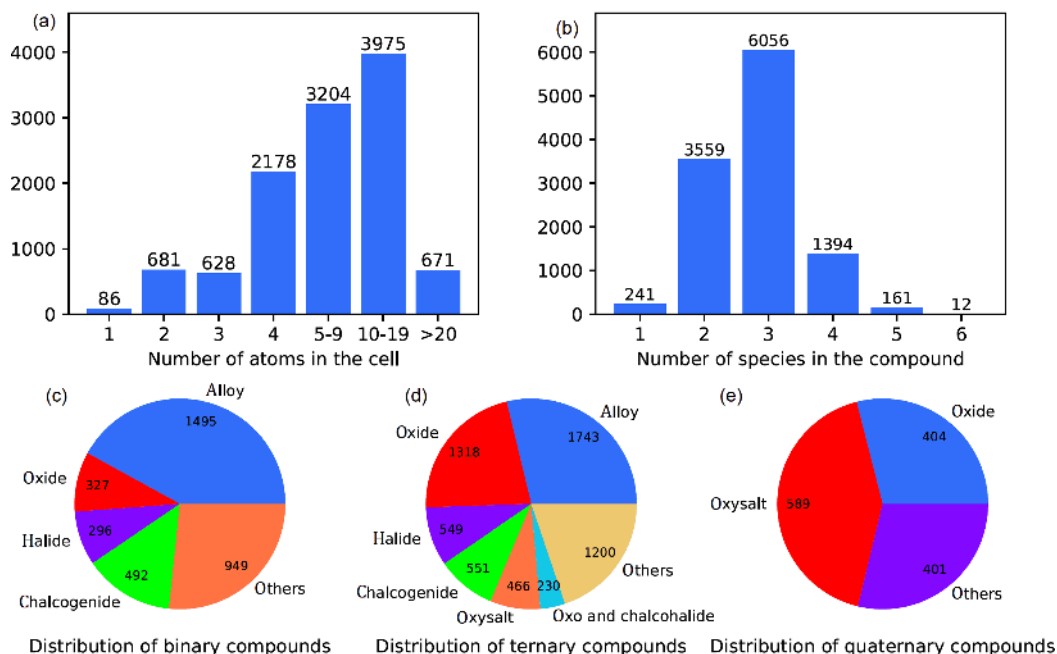
We then look at the overall agreements between databases shown in Figure 3(a)–(c). Once again, we divide the cases into three categories: both- $U$ ; one- $U$ , none- $U$ . The closest agreement comes from the none- $U$  category represented by the red dots in the figures. In Figure 3(a)–(c), the red dots are quite close to the diagonal, and the Pearson correlation coefficients for them are 0.918, 0.941 and 0.931, respectively. If we only consider the structures that are assigned as insulators by both databases and only for none- $U$ , the correlation coefficients further rise to 0.984, 0.978 and 0.983, respectively. This shows that in the absence of the influence of  $U$  parameter, these databases give highly reliable results, and their mutual correlations are high. The correlation coefficient represents whether two data sets are linearly correlated, but even if the correlation coefficient equals one, it does not mean they are the same, since they can be differed by a scaling factor and a shift. We thus have also calculated



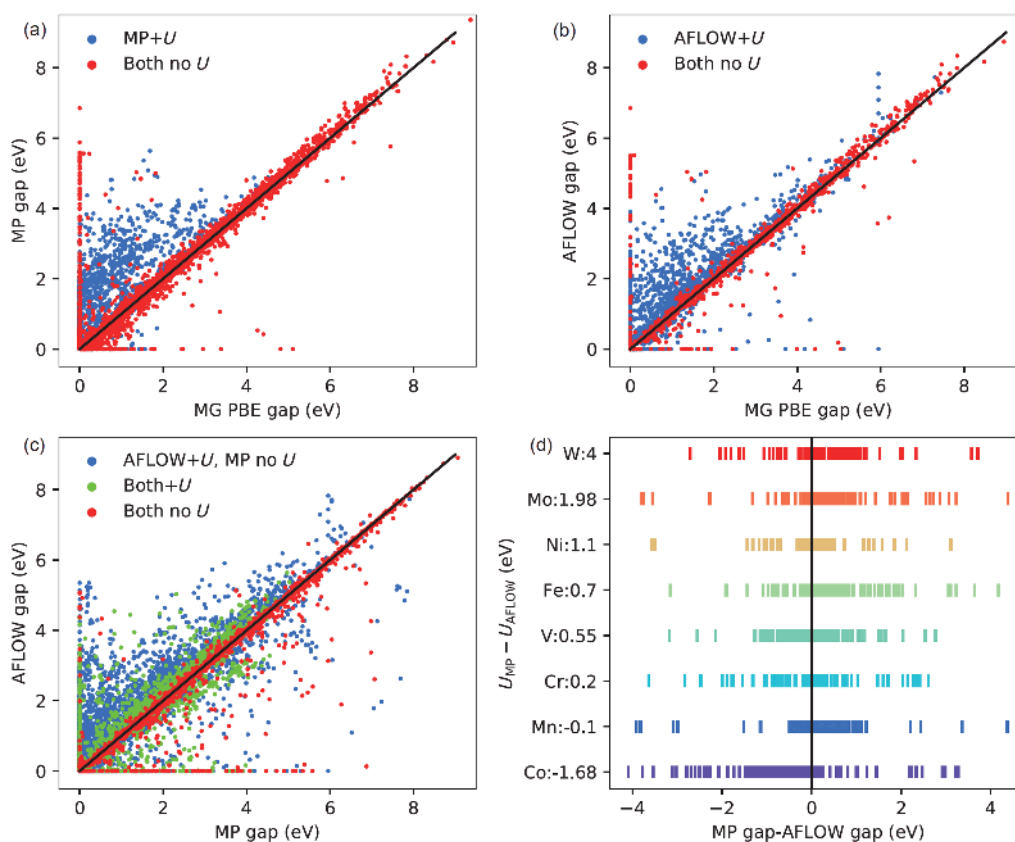
**Figure 1** (Color online) (a) The home webpage of periodic table to search materials; (b) 3D materials structures and cell parameters shown in our webpage; (c) structures can be rotated in webpage HSE band structures and electronic density of states. Band structures can be zoomed in or out for details and whole pictures.

their root mean square (RMS) band gap differences, only taken into account the insulator ones in both databases and none- $U$  cases. The RMS band gap differences are 0.309, 0.388, 0.328 eV for the three comparisons in Figure 3(a)–(c) respectively. The relatively small RMS differences mean these databases are not only linearly correlated, they are indeed close to each other. We do see slightly more outliers (away from the  $y=x$  black line) for the MP-AFLOW comparison for the red dot in the higher band gap region, where

AFLOW tends to predict smaller band gaps. Given that they both use VASP code, the difference might come from the different DFT relaxed structure. On the other hand, the MG and MP have the same crystal structures, they have less such outliers and also have the smallest RMS of the band gap differences, despite of the fact that one is calculated with VASP while the other is calculated with PWmat and different pseudopotentials are used. This means the different codes and pseudopotential are less of a problem, and the crystal



**Figure 2** (Color online) (a) The distribution of number of species of calculated structures in our database; (b) the distribution of number of atoms in the cell of calculated structures in our database; (c)–(e) the distribution of different kinds of compounds of binary, ternary and quaternary compounds in our database. Structures with other numbers of species are fewer and not further analyzed.



**Figure 3** (Color online) Comparison of the three PBE-related databases. (a), (b), (c) are comparisons between MP and MG PBE, AFLOW and MG PBE, AFLOW and MP, respectively. In each subplot, red points correspond to structures that have no +U species in either database, blue points are those that have +U species in only one database, and green points are structures that have +U species in both databases. The black line is the  $y=x$  identity line. (d) shows structures that have exactly one +U specie in both MP and AFLOW. It shows the relationship between the  $U$  difference of a cation atom (vertical axis) and the predicted band gap difference (horizontal axis) between MP and AFLOW. Each short vertical bar represents one material containing the transition metal elements listed on the vertical axis.

**Table 1** The influence of Hubbard Parameter on the consistency of metal-insulator classification between different PBE-related databases

	MP vs MG PBE		AFLOW vs MG PBE		MP vs AFLOW	
	Same <sup>d)</sup>	Different <sup>e)</sup>	Same	Different	Same	Different
Both+ $U$ <sup>a)</sup>	–	–	–	–	84.5%	15.5%
One + $U$ <sup>b)</sup>	73.3%	26.7%	92.0%	8.0%	89.5%	10.5%
Non + $U$ <sup>c)</sup>	93.2%	6.8%	95.2%	4.8%	96.6%	3.4%
All	90.1%	9.9%	92.9%	7.1%	90.8%	9.2%

a) refers to structures that have elements applied with  $U$  parameters in both databases; b) refers to structures with only one database using  $U$  in some elements in the structure; c) refers to structures without any element that is applied with  $U$  in either database; d) means the structure is considered as either conductor or insulator in both databases; e) means it is considered as conductor in one database and insulator in another.

structure difference might play a bigger role when getting different results. But overall, when no  $U$  parameter is used, and both databases predict a crystal as insulator, the results are consistent.

We next look at the comparison where one database uses  $U$  and the other does not use  $U$ , as represented by the blue dots in Figure 3(a)–(c). The trend is clear, the database with  $U$  predicts much bigger band gaps than the one not using  $U$ . The correlation coefficients for the blue dots in Figure 3(a)–(c) are 0.638, 0.870 and 0.856, while the RMS band gap differences are 1.379, 0.770 and 0.796 eV, respectively. Thus, there is no simple correlation between the + $U$  gaps and non- $U$  gaps, it is thus impossible to use a simple scaling to yield the + $U$  result from the non- $U$  result.

We have also shown the results where both AFLOW and MP use  $U$ , as represented by the green dots in Figure 3(c). We see that the spread is relatively large, although smaller than the one- $U$  cases. One of the reasons is that the  $U$ 's used in AFLOW and MP can be rather different for the same element containing d-state. To illustrate that, we have shown the difference of  $U$  between AFLOW and MP for different d-containing elements as the vertical axis in Figure 3(d), and the predicted band gap differences as the horizontal axis. The difference in  $U$  values varies from  $-1.68$  eV for Co (3.32 eV in MP and 5.0 eV in AFLOW) to 4 eV for W (6.2 eV in MP and 2.2 eV in AFLOW). Surprisingly, we do not see any obvious trend for the corresponding band gap difference. The small  $U$  difference case (Cr) has similar band gap spread as the large  $U$  difference cases (Co and W). Nevertheless, there is a weak trend that larger  $U$  tends to lead to bigger band gap, as represented by a slight center of mass shift towards right from Co to W.

As a summary, different PBE-related databases sometimes disagree on whether a structure is a metal or an insulator, and significant band gap difference can exist for the insulator cases. However, the main source of the difference comes from the different usage of Hubbard parameter  $U$  when PBE + $U$  is used. There are cases which one database uses  $U$ , but the other database does not use, or when both use  $U$ , but their values are very different. When neither of the data base uses  $U$ , the relative agreement is good: the correlation coefficients

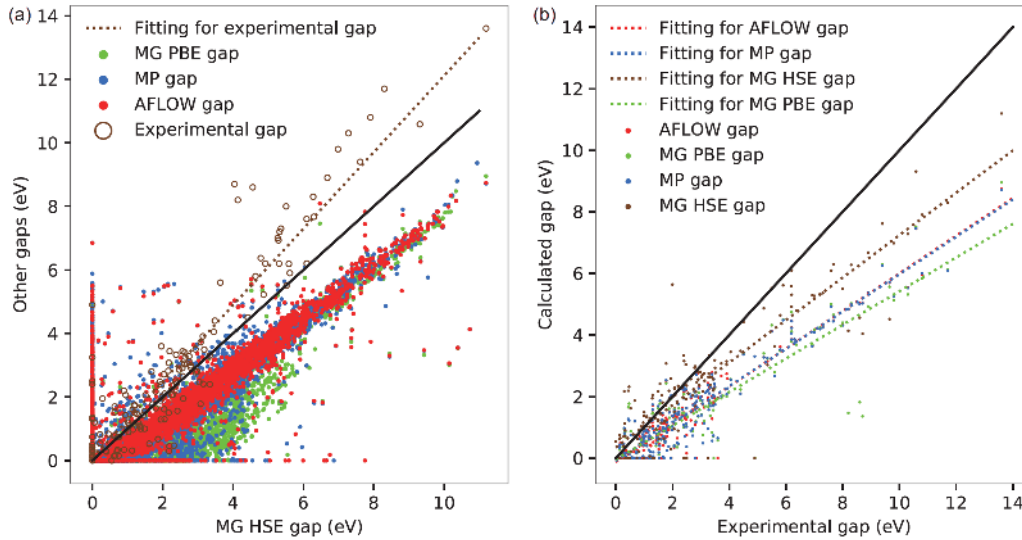
are in the range of 0.98; the RMS band gap difference is 0.3–0.4 eV; and the mis-identification rate between metal/insulator is about a few percent. We thus identify the uncertainty of  $U$  as the main source of scatters in the predicted band gaps. On the other hand, the discrepancies caused by different atomic structure (due to different relaxation scheme), or different pseudopotentials and codes, are relatively small.

### 3.3 The effect of HSE calculation and comparison with experiment

One of the major advances for MG is the use of hybrid exchange functional in addition to the PBE calculations. The hybrid functional is designed to provide a better prediction for the semiconductor band gap, although its computational cost is much higher than that of PBE especially for plane wave-based codes.

We have plotted the comparison between MG+HSE, MG+PBE, AFLOW+PBE (w/o  $U$ ) (with or without  $U$ ), MP+PBE (w/o  $U$ ) in Figure 4(a). We have used the MG+HSE as the  $x$ -axis since this is the one clearly different from the other three PBE results. First, comparing MG+HSE with MG+PBE, we see that the different identification rate between metal/insulator is about 8% (Table 2). Thus, the difference in calculation methods does not dramatically change the metal/insulator identification rate. As for all the insulator cases, there is a clear trend that HSE band gaps are systematically larger than PBE gaps. If only the no- $U$  data are used for the three PBE results, then the correlation coefficients of them to the MG+HSE are 0.944, 0.938, 0.919. These are rather high, indicating a simple scaling correction can be used to convert the PBE (without  $U$ ) results to HSE results. The multiplication factor for the three cases are 0.782, 0.765, 0.793 for MP/MG+HSE, AFLOW/MG+HSE, and MG+PBE/MG+HSE, respectively.

It will be more interesting to compare the calculated results with experiments. Unfortunately, the experimental band gaps for the 10000 structures are not all available. Nevertheless, we have collected about 100 structures with their band gaps listed in Table S3. They are also represented in Figure 4(a).



**Figure 4** (Color online) MG HSE gaps vs MG PBE, MP, AFLOW and experimental gaps. Both no- $U$  and with- $U$  data are included for MP and AFLOW. The black lines in (a) and (b) are  $y=x$  line. In (b), the fitted AFLOW and MP gaps (including the data with  $U$ ) are almost on top of each other.

**Table 2** HSE vs PBE for the metal-insulator classification

	MG HSE vs MG PBE		MG HSE vs MP		MG HSE vs AFLOW	
	Same	Different	Same	Different	Same	Different
PBE+ $U$	–	–	84.0%	16.0%	91.3%	8.7%
No $U$	89.6%	10.4%	89.5%	10.5%	93.2%	6.8%
ALL	89.6%	10.4%	88.6%	11.4%	91.9%	8.1%

As one can see, there are some scattering, but the correlation coefficient between the MG+HSE result and the experimental result is still 0.927, indicating a linear transformation can still be used to convert the HSE result to the experimental result. It is similar for the PBE results. The correlation coefficients between MG+PBE, AFLOW+PBE, MP+PBE and experimental values are 0.894, 0.965, 0.941, respectively, not much different from the correlation coefficient for the MG+HSE. In the above comparison with experiment, both with  $U$  and no- $U$  cases are used for AFLOW+PBE and MP+PBE, the comparison will be worse if only the no- $U$  cases are used (as for the case of MG+PBE). This means, although the use of  $U$  parameters increases the discrepancies between different databases, it does reduce the error relative to experimental values. If one draws an average linear fit to the experimental data, it can be represented by the brown dotted line in Figure 3. One can then see, the MG+HSE (presented by the black line in Figure 4) are in the middle between the PBE results and the experimental result. To understand the situation from another perspective, we have plotted the comparison for the 100 structure with experimental band gaps in Figure 4(b), using experimental gap as the  $x$ -axis and the calculated gap as the  $y$ -axis. As we can see, although the use of hybrid functional significantly improves calculated band gaps, in an average sense, the HSE still

significantly underestimates the band gaps especially for the large band gap cases. This is because the original HSE is fitted to yield the correct band gap for small band gap (1 to 2 eV) semiconductors. The fixed Fock exchange mixing parameter can be understood as a fixed screening dielectric effect (in screened exchange formalism) [31]. For large band gap materials, the dielectric constant is usually much smaller, hence one should use a larger mixing parameter. The use of a fixed mixing parameter means an overestimation of the screening for large band gap materials, which leads to underestimation of the band gaps.

## 4 Conclusions

In conclusion, we introduce MG, a new database containing both PBE and HSE bandgaps and compare it with existing databases like MP and AFLOW and experimental values. We find that, (1) among all calculation databases there are some significant disagreements when classifying a structure into metals and insulators; (2) the prime reason for deviations among different database calculated with PBE is the use of  $U$ ; (3) when no  $U$  is used, PBE results between different databases have correlation coefficients about 0.98, and RMS errors of about 0.35 eV for the cases which are identified by

both databases as insulators; (4) while HSE significantly enlarges calculated band gaps, it still significantly underestimates band gaps of large band gap materials due to the fixed value of mixing parameter. Thus, to build a calculation database which gives decent band gaps, one should either find a way to automatically regulate mixing parameters, or use even more accurate methods, like GW method [32] or the Wannier Koopman's method [33–36].

*This work was supported by the National Key R&D Program of China (Grant No. 2016YFB0700600), the Shenzhen Science and Technology Research Grant (Grant No. ZDSYS201707281026184), and the Guangdong Key-lab Project (Grant No. 2017B0303010130). Wang is supported by the Director Office of Science (SC), Basic Energy Science (BES), Materials Science and Engineering Division (MSED) of the US Department of Energy (DOE) under Contract No. DE-AC02-05CH11231 through the Materials Theory program (KC2301).*

### Supporting Information

The supporting information is available online at [tech.scichina.com](http://tech.scichina.com) and [link.springer.com](http://link.springer.com). The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

- Pilania G, Gubernatis J E, Lookman T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. *Comput Mater Sci*, 2017, 129: 156–163
- de Jong M, Chen W, Notestine R, et al. A statistical learning framework for materials science: Application to elastic moduli of k-nary inorganic Polycrystalline Compounds. *Sci Rep*, 2016, 6: 34256
- Morales-García Á, Valero R, Illas F. An empirical, yet practical way to predict the band gap in solids by using density functional band structure calculations. *J Phys Chem C*, 2017, 121: 18862–18866
- Belsky A, Hellenbrandt M, Karen V L, et al. New developments in the inorganic crystal structure database (ICSD): Accessibility in support of materials research and design. *Acta Cryst Sect A Found Cryst*, 2002, 58: 364–369
- Bergerhoff G, Hundt R, Sievers R, et al. The inorganic crystal structure data base. *J Chem Inf Model*, 1983, 23: 66–69
- Downs R T, Hall-Wallace M. The american mineralogist crystal structure database. *Am Mineral*, 2003, 88: 247–250
- Pizzi G, Cepellotti A, Sabatini R, et al. AiiDA: Automated interactive infrastructure and database for computational science. *Comput Mater Sci*, 2016, 111: 218–230
- Jain A, Hautier G, Moore C J, et al. A high-throughput infrastructure for density functional theory calculations. *Comput Mater Sci*, 2011, 50: 2295–2310
- Jain A, Ong S P, Hautier G, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater*, 2013, 1: 011002
- Curtarolo S, Setyawan W, Hart G L W, et al. AFLOW: An automatic framework for high-throughput materials discovery. *Comput Mater Sci*, 2012, 58: 218–226
- Calderon C E, Plata J J, Toher C, et al. The AFLOW standard for high-throughput materials science calculations. *Comput Mater Sci*, 2015, 108: 233–238
- Curtarolo S, Setyawan W, Shi Dongwang, et al. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput *ab initio* calculations. *Comput Mater Sci*, 2012, 58: 227–235
- Saal J E, Kirklin S, Aykol M, et al. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *JOM*, 2013, 65: 1501–1509
- Liu Y, Zhao T, Ju W, et al. Materials discovery and design using machine learning. *J Materiomics*, 2017, 3: 159–177
- Liu Y, Zhao T, Yang G, et al. The onset temperature ( $T_g$ ) of As Se1 glasses transition prediction: A comparison of topological and regression analysis methods. *Comput Mater Sci*, 2017, 140: 315–321
- Shi S Q, Gao J, Liu Y, et al. Multi-scale computation methods: Their applications in lithium-ion battery research and development. *Chin Phys B*, 2016, 25: 018212
- Wang Y, Zhang W, Chen L, et al. Quantitative description on structure-property relationships of Li-ion battery materials for high-throughput computations. *Sci Tech Adv Mater*, 2017, 18: 134–146
- Perdew J P, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Phys Rev Lett*, 1996, 77: 3865–3868
- Heyd J, Scuseria G E, Ernzerhof M. Hybrid functionals based on a screened coulomb potential. *J Chem Phys*, 2003, 118: 8207–8215
- Shockley W, Queisser H J. Detailed balance limit of efficiency of *p-n* junction solar cells. *J Appl Phys*, 1961, 32: 510–519
- Adjokatse S, Fang H H, Loi M A. Broadly tunable metal halide perovskites for solid-state light-emission applications. *Mater Today*, 2017, 20: 413–424
- Lucero M J, Henderson T M, Scuseria G E. Improved semiconductor lattice parameters and band gaps from a middle-range screened hybrid exchange functional. *J Phys-Condens Matter*, 2012, 24: 145504
- Moussa J E, Schultz P A, Chelikowsky J R. Analysis of the heyd-scuseria-ernzerhof density functional parameter space. *J Chem Phys*, 2012, 136: 204117
- Jia W, Fu J, Cao Z, et al. Fast plane wave density functional theory molecular dynamics calculations on multi-GPU machines. *J Comput Phys*, 2013, 251: 102–115
- Jia W, Cao Z, Wang L, et al. The analysis of a plane wave pseudo-potential density functional theory code on a GPU machine. *Comput Phys Commun*, 2013, 184: 9–18
- Hamann D R. Optimized norm-conserving vanderbilt pseudopotentials. *Phys Rev B*, 2013, 88: 085117
- Schliffl M, Gygi F. Optimization algorithm for the generation of ONCV pseudopotentials. *Comput Phys Commun*, 2015, 196: 36–44
- Lin L. Adaptively compressed exchange operator. *J Chem Theor Comput*, 2016, 12: 2242–2249
- Krukau A V, Vydrov O A, Izmaylov A F, et al. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J Chem Phys*, 2006, 125: 224106
- Heyd J, Peralta J E, Scuseria G E, et al. Energy band gaps and lattice parameters evaluated with the heyd-scuseria-ernzerhof screened hybrid functional. *J Chem Phys*, 2005, 123: 174101
- Lee B, Wang L W, Spataru C D, et al. Nonlocal exchange correlation in screened-exchange density functional methods. *Phys Rev B*, 2007, 76: 245114
- Hybertsen M S, Louie S G. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys Rev B*, 1986, 34: 5390–5413
- Ma J, Wang L W. Using Wannier functions to improve solid band gap predictions in density functional theory. *Sci Rep*, 2016, 6: 24924
- Weng M, Li S, Ma J, et al. Wannier koopman method calculations of the band gaps of alkali halides. *Appl Phys Lett*, 2017, 111: 054101
- Weng M, Li S, Zheng J, et al. Wannier koopmans method calculations of 2D material band gaps. *J Phys Chem Lett*, 2018, 9: 281–285
- Li S, Weng M, Jie J, et al. Wannier-koopmans method calculations of organic molecule crystal band gaps. *Europhys Lett*, 2018, 123: 37002