2-21-2022

Large Scale Dataset of Real Space Electronic Charge Density of Cubic Inorganic Materials from Density Functional Theory (DFT) Calculations

Fancy Qian Wang

Kamal Choudhary

Jianjun Hu
University of South Carolina - Columbia, jianjunh@cse.sc.edu

Ming Hu
University of South Carolina - Columbia, hu@sc.edu

scientific **data**

OPEN

DATA DESCRIPTOR

Large scale dataset of real space electronic charge density of cubic inorganic materials from density functional theory (DFT) calculations

Fancy Qian Wang[1] ✉, Kamal Choudhary [2,3], Yu Liu[1], Jianjun Hu[4] & Ming Hu[5] ✉

Driven by the big data science, material informatics has attracted enormous research interests recently along with many recognized achievements. To acquire knowledge of materials by previous experience, both feature descriptors and databases are essential for training machine learning (ML) models with high accuracy. In this regard, the electronic charge density $\rho(r)$, which in principle determines the properties of materials at their ground state, can be considered as one of the most appropriate descriptors. However, the systematic electronic charge density $\rho(r)$ database of inorganic materials is still in its infancy due to the difficulties in collecting raw data in experiment and the expensive first-principles based computational cost in theory. Herein, a real space electronic charge density $\rho(r)$ database of 17,418 cubic inorganic materials is constructed by performing high-throughput density functional theory calculations. The displayed $\rho(r)$ patterns show good agreements with those reported in previous studies, which validates our computations. Further statistical analysis reveals that it possesses abundant and diverse data, which could accelerate $\rho(r)$ related machine learning studies. Moreover, the electronic charge density database will also assists chemical bonding identifications and promotes new crystal discovery in experiments.

Background & Summary

The electronic charge density (ECD) in real space, denoted as $\rho(r)$, is a basic yet informative observable quantity of materials in physics. Dated back to 1964, density functional theory (DFT) was discovered by Hohenberg and Kohn, who proved that the properties of materials at their ground state can be entirely and exclusively determined by $\rho(r)$[1,2]. This theorem has been widely used and applied to multiple physical systems from microscale molecules to macroscale crystals, significantly accelerating our understanding and manipulation of the nature of the world. For example, the bonding characters between neighbouring atoms, a very fundamental yet complex concept (covalent, ionic, metallic bonds), can be fully described by ECD, rather than the simple bonding chemistry. With this property, novel materials with specific structures and target properties[3,4] could be artificially designed, which is one of the most crucial issues in modern crystallography and inverse-materials design[3]. In addition, most electronic, magnetic, optical properties and their couplings, such as electrostatic moments, potentials and interaction energies, spin susceptibility, light absorption, and electromagnetic responses, etc, could be directly obtained starting from the ECD $\rho(r)$ according to the modern band structure theory[2]. Because of these advantages, the basic quantity $\rho(r)$ possesses broad applications, such as identifying the binding sites in host-guest compounds[5], computing infrared intensities[6,7], revealing structural stability, simulating scanning tunnelling microscopy images[8] and so on[9–11]. For instance, Shen et al. have recently reported a charge-density-based general cation insertion algorithm for generating and designing new Li-ion cathode materials[11], and Hu et al. have made efforts to predict mechanical elastic properties of materials from their $\rho(r)$[10].

[1]State Key Laboratory of High-end Server & Storage Technology, Inspur Electronic Information Industry Co., Ltd, Beijing, 100085, China. [2]Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA. [3]Theiss Research, La Jolla, CA, 92037, USA. [4]Department of Computer Science and Engineering, University of South Carolina, Columbia, 29208, South Carolina, United States. [5]Department of Mechanical Engineering, University of South Carolina, Columbia, 29208, South Carolina, United States. ✉e-mail: fancywang@pku.edu.cn; hu@sc.edu
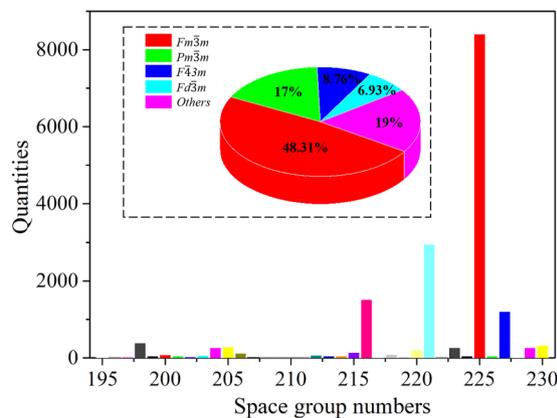
Fig. 1 The number of materials with respect to each space group in ECD-cubic database. The insertion is the corresponding percentage distribution.
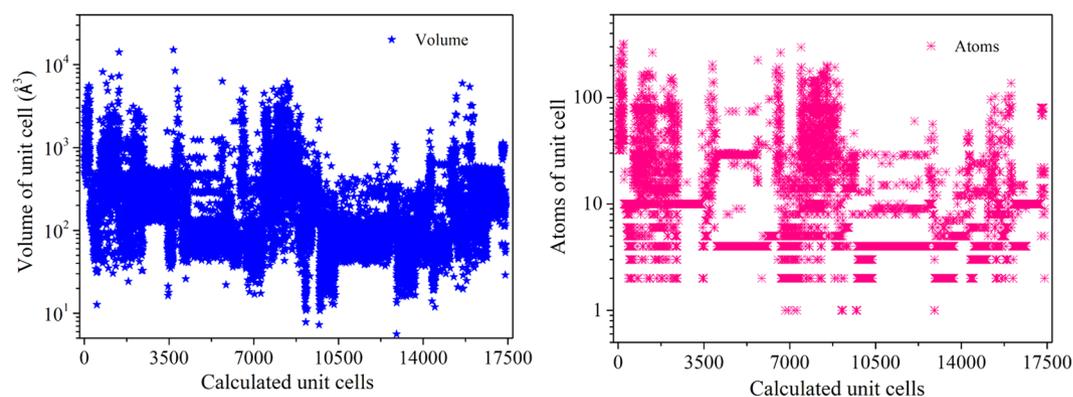
Fig. 2 (**a**) The distributions of volume and (**b**) atom numbers in a unit cell for all the materials in ECD-cubic database.

From the experimental perspective, the ECD $\rho(r)$ can be probed by high-resolution electron diffraction[4,12–14] and transmission electron microscopy, and then subjected to Bader's Quantum Theory of Atoms in Molecules (QTAIM)[15,16] analysis for extracting bonding information based on multipole models (MM) density[14,17]. However, the flaw of such technique resides in collecting raw and accurate metadata of inorganic materials, especially for those high-symmetric dense solids with extended structures and heavy atoms. This is because, in such materials, the scattering from the valence electrons is minute compared to that from core electrons. Consequently, systematic errors such as extinction and absorption are critical[17]. Accordingly, unlike the widely explored molecular systems, very few ECD $\rho(r)$ of inorganic materials were acquired even using the modern techniques. On the contrary, from the theoretical point of view, the ECD $\rho(r)$ of inorganic materials can be accurately simulated by using *ab initio* calculations according to the density functional theory (DFT). Even though quantum chemistry approaches are more immune to artificial parameters, the computational demanding of DFT is more appropriate for realistic materials, rather than small molecules. Hence, the DFT method with sufficient predictive ability is powerful in dealing with ECD $\rho(r)$ of inorganic materials, and thus giving better assistance to experiments.

In the modern science and technology world, big data-driven science has become the fourth science paradigm[18,19], thanks to the significantly increased computing power and the huge data generated every day. Utilizing the machine learning (ML) algorithms to solve problems in material science has significantly boosted the development of material informatics[19–23], among which bypassing the DFT calculations to directly predict the fundamental $\rho(r)$ of materials has attracted immense attentions[24–27]. Nevertheless, most of the recent progresses were confined to the molecular systems[28,29], mainly because the systematic ECD $\rho(r)$ database of inorganic materials is still in its infancy.

Based on the above discussions, in this study, we construct the ECD-cubic, an ECD $\rho(r)$ database of inorganic materials by using the state-of-the-art first-principles calculations. Because of the expensive yet limited computational resources, we focus on materials in cubic symmetry where the $\rho(r)$ are more efficient to calculate (in computation time per atom). The $\rho(r)$ of materials with other symmetries, such as hexagonal, orthorhombic and monoclinic, will be successively included in our future work. So far, the ECD-cubic database contains the $\rho(r)$ of 17,418 materials, which are all selected from the Materials Project database[30,31] and possess cubic
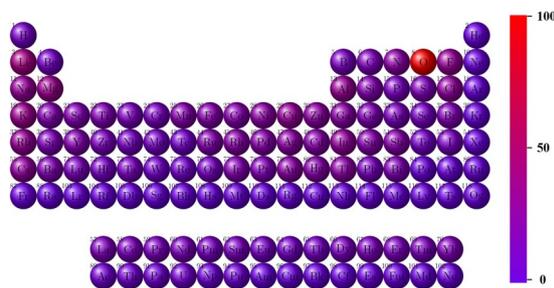
Fig. 3 The distribution of elements for all the materials in ECD-cubic database.

structures. The distributions of space group, volume and number of atoms in the unit cell and elements of the ECD-cubic database are statistically analysed and the results are given in Figs. 1–3. We can see that the space groups of the materials in the ECD-cubic database span from $P23$ (195) to $Ia\overline{3}d$ (230), where the proportions of the top three, namely the $Fm\overline{3}m$ (225), $Pm\overline{3}m$ (221), $F\overline{4}3m$ (216) are 48.31%, 17% and 8.76%, respectively. It is found that the volumes of the unit cells in ECD-cubic database span five orders of magnitude, from the smallest one with 5.6Å^3 (mp-998866) to the largest one with 15,120Å^3 (mp-1172909), and most of them are in the range of 30–1,000Å^3 (~93% in the database). Consistently, the number of atoms in the unit cell also vary in a wide range. For instance, the $Rb_3Sc_2(AsO_4)_3$ (mp-1205185) and $Te(CF_2)_4$ (mp-1204577) have extraordinarily complicated atomic geometry with 320 and 312 atoms in the unit cell, respectively, while a lot of metals with the face-centered cubic (FCC) structure (mp-81, mp-10740, mp-20483, mp-612118, etc) only have one atom in their primitive cells. The colour bar displayed in Fig. 3. distinguishes the percentage of each element involved in materials, where the red and purple indicate higher and lower ratios, respectively. We can see that the ECD-cubic database consists of 89 kinds of elements, where the O element (6.62%) far exceeds others, and the alkali metals, namely, the Li (2.72%), K (2.34%) and Mg (2.33%) elements afterwards.

All the features above demonstrate that the constructed ECD-cubic database is formed by high-quality, abundant, and diverse data, which is an essential prerequisite for ECD related ML studies in material informatics. Simultaneously, it will also provides good opportunities to crystal engineering and gives better guidance for the chemical bonding identifications in experiments.

Methods

In this section, we give a brief introduction to the methods employed in this study, including the general theory of ECD $\rho(r)$ adopted by VASP (Vienna ab-initio simulation package) software, the workflow that we select materials, and all the parameters used in the DFT calculations.

General theory. In DFT, the relationship between ECD $\rho(r)$ and wave function $\Psi(r, r_2, \cdots, r_N)$ is given by:

$$\rho(r) = N \int d^3 r_2 \cdots \int d^3 r_N \Psi^*(r, r_2, \cdots r_N) \Psi(r, r_2, \cdots r_N)$$

The total number of electrons (N) in a unit cell are equal to the integral of ECD $\rho(r)$ over the entire volume (V_{un}). After discretization, the relationship could be expressed as:

$$N = \int_V \rho(r) dV = \sum_{i=1}^{NGXF*NGYF*NGZF} \rho(r_i) \frac{V_{un}}{NGXF * NGYF * NGZF}$$

where the $NG(X, Y, Z)F$ are the fine Fast Fourier Transform (FFT) grids in the reciprocal space along the x, y and z directions, respectively. A series of discrete values of $\rho(r_i) V_{un}$ at each fine FFT-grid are recorded in the CHG file, carrying all the desired information of $\rho(r)$. In non-spin polarized calculations for non-magnetic materials, the CHG only contains total electronic charge density ($\rho(r) = \rho(r)_{spinup} + \rho(r)_{spindown}$), while for those spin-polarized calculations for magnetic materials, a spin electronic charge density ($\rho(r) = \rho(r)_{spinup} - \rho(r)_{spindown}$) will be additionally given.

Workflow. All materials are downloaded from the Material Project database[30,31], which is one of the widely used databases in material informatics since it contains more than 144,595 inorganic compounds with three-dimensional structural information. On account of the extremely expensive computational cost of the DFT calculations, the cubic symmetry is treated as the criterion when we select materials, leaving 18,494 candidates. Out of these candidates, we have calculated the electronic charge density $\rho(r)$ of 17,418 materials after filtering out the structures with incomplete information or the calculations cannot achieve good self-consistent field convergence after several tries. The ECD $\rho(r)$ of the remaining materials with non-cubic symmetries would be available in the future.

Density functional theory calculations. All the first-principles calculations are carried out using the projector augmented wave (PAW) method[32,33] as implemented in the Vienna Ab initio Simulation Package (VASP) based on the density functional theory (DFT). We start by optimizing each crystal structure, where both the atomic positions and lattice constants are fully allowed to relax in spin-unrestricted mode and without any

Keys	Description
system	The name of calculated material, the same as the content in the 1st line of CHG file.
vector	Vector, usually 1.0, the same as the content in the 2nd line of CHG file.
lattice	Lattice constants along the x, y and z directions, respectively, the same as the content from 3rd - 4th line of CHG file.
elements	The elements involved in materials, the same as the content in the 5th line of CHG file.
elements_number	The quantities of each element listed above, the same as the content in the 6th line of CHG file.
coor_type	The type, usually direct, the same as the content in the 7th line of CHG file.
coordinates	The atomic coordinates along x, y and z directions in materials, the same as the content in the 8th-$(8+3N)$th line of CHG file, where N is the number of atoms in the cell.
FFT	The FFT grids ($NGXF$, $NGYF$ and $NGZF$) used in calculations along the x, y and z directions, the same as the content after coordinates in CHG file.
charge	The calculated electronic charge density components based on the FFT grids. Note: if the materials without magnetism, such entry only contain the total charge density, the number of components equal to $NGXF*NGYF*NGZF$; while for materials with magnetism, such entry additionally contain the spin charge density behind the total charge density, the number of components equal to $2*NGXF*NGYF*NGZF+1$

Table 1. The keys and their corresponding descriptions of the JSON file for each material.

symmetry constraints. These calculations are performed until the maximal Hellmann-Feynman force component smaller than 10^{-3} eVÅ^{-1}, and the total energy convergence tolerance is set to be 10^{-6} eV. To obtain accurate lattice parameters[34–36], the Opt-B88vdW functional[37] is taken into account to deal with the long term interactions in the exchange-correlation interaction. The k-point grids (KPOINTS) for each material used to calculate their CHGs are the same KPOINTS files as used in static calculations in Materials Project database (downloaded in around June 2020), which have been proved to achieve good convergence in previous works. All the required KPOINTS files are obtained by using the Pymatgen (Python Materials Genomics), which is a robust, open-source python library for materials analysis. The choice of kinetic energy cut-off of planewave functions for each material is also according to the Materials Project database. After fully converged, the ECD $\rho(r)$ of all the materials are calculated separately with energy convergence threshold is set to be 10^{-6} eV.

Data Records

As mentioned above, the ECD-cubic database is formed by the $\rho(r)$ of 17,418 inorganic materials along with their atomic structures. To be consistent with the Material Project database, we continue to use the same material ID for identifying them. The metadata of each material is stored in the Javascript Object Notation Files (JSON) format, denoted as mp-id.json, which can be easily integrated with other databases such as MongoDB. All the entries including the keys and their corresponding descriptions of the JSON file are listed in Table 1. Moreover, we also provide a python script to parse the JSON file to the standard CHG format for visualizing or restarting the VASP calculations. According to the previous study, the quality of the data is determined by its coverage of the chemical-property space of interest as well as the uncertainty associated with the data[38]. In the ECD-database, all the $\rho(r)$ are created by consistent DFT calculations, thus largely removing its uncertainty. While for separating the chemical-property space of interest, herein, we provide a structural list recording the IDs with the corresponding energy above hull of each material. Such a list and the ECD-database are made available through Figshare repository[39]. The same copy is also uploaded to our Carolina Materials Database (http://www.carolinamatdb.org/). Besides, to enhance the reproducibility of this work, all the raw input files, namely the POSCAR, KPOINTS and INCAR for calculating the electronic charge density of each material can be acquired via Carolina Materials Database.

We would like to give some comparison between our calculations and other publicly available datasets. The Material Project database includes several datasets such as band structures, piezoelectric tensors, and elastic properties, yet totally excludes the electronic charge densities. Although a few CHGCARs of materials could be found in the NOMAD repository[40], there are some uncertainties in terms of the quality of those CHGCARs. First, the CHGCARs stored in NOMAD are generated in the calculations of structure optimization, not the self-consistent process. Such CHGCARs are usually used to restart VASP calculations, not for the electronic charge densities analysis. Second, the parameters used for calculating the CHGCARs are missing, thus users may not perform reliable subsequent processing from these data. Besides, we extract initial structures of each material from the Material Project database and re-optimize them using the opt-B88vdW functional instead of the GGA/PBE functional used in the Material Project database. This would be more accurate because the opt-B88vdW functional has been proved to give much improved lattice parameters for both van der Waals (vdW) and non-vdW solids[34], which are essential to the calculations of accurate electronic charge density. In addition, such functional is usually adopted in the construction of many other properties related database[8,34,36,41], hence our datasets would be useful for further analysis and comparisons since consistent computational procedures are adopted.

Technical Validation

The calculated electronic charge density $\rho(r)$ is a widely accepted quantity for predicting physical properties[34] benefitting from the significant predictive power of the state-of-the-art DFT calculations. Here we elaborate several patterns of the calculated $\rho(r)$ and compare them with those reported in previous theoretical or experimental studies, to further verify the correctness of our simulations. Since there are over 17,418 electron density files calculated in our work, and it is unlikely to visualize all of them in the manuscript. Instead, we only show a
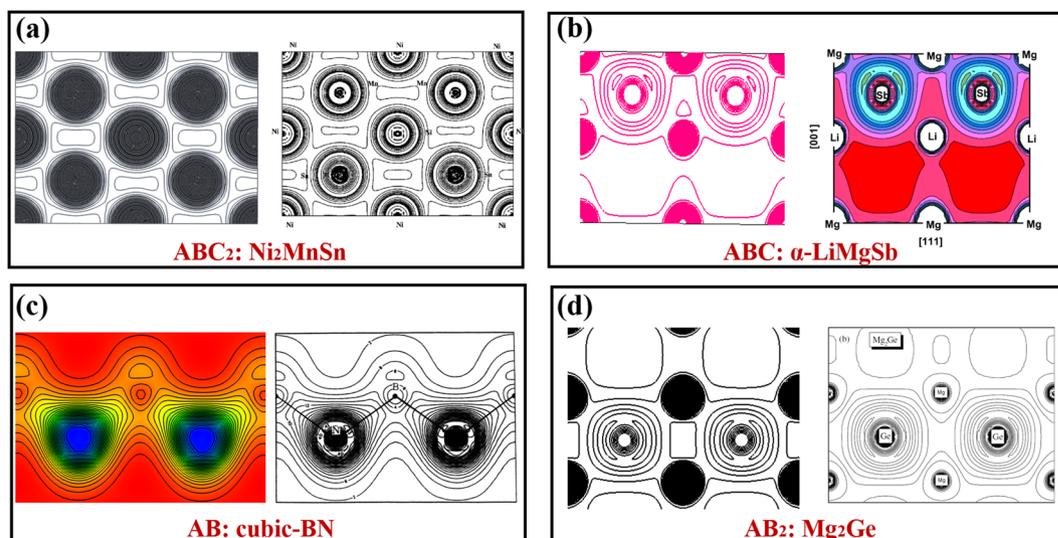
Fig. 4 The calculated patterns of electronic charge density $\rho(r)$ of (**a**) ABC_2 type, Ni_2MnSn [The right panel reprint with permission from ref. [51]. Copyright 2001 American Physical Society], (**b**) ABC type, α-LiMnSb [The right panel reprint with permission from ref. [47]. Copyright 2010 Elsevier], (**c**) AB type, cubic-BN [The right panel reprint with permission from ref. [52]. Copyright 1986 American Physical Society] and (**d**) AB_2 type, Mg_2Ge [The right panel reprint with permission from ref. [48]. Copyright 2005 Wiley], respectively.
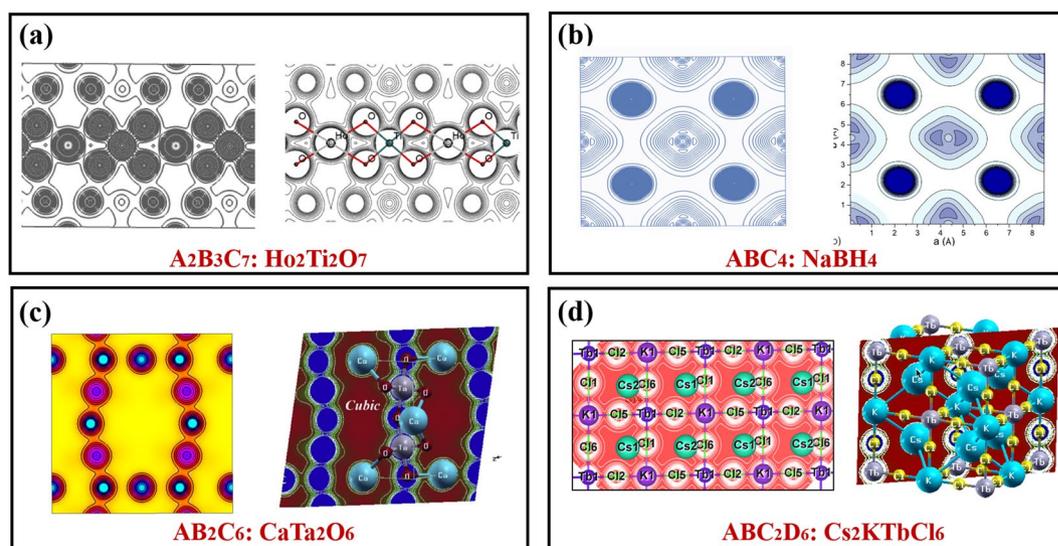
Fig. 5 The calculated patterns of electronic charge density $\rho(r)$ of (**a**) $A_2B_3C_7$ type, $Ho_2Ti_2O_7$ [The right panel reprint with permission from ref. [53]. Copyright 2017 Elsevier], (**b**) ABC_4 type, $NaBH_4$ [The right panel reprint with permission from ref. [50]. Copyright 2010 American Physical Society], (**c**) AB_2C_6 type, $CaTa_2O_6$ [The right panel reprint with permission from ref. [42]. Copyright 2019 Elsevier] and (**d**) ABC_2D_6 type, Cs_2KTbCl_6 [The right panel reprint with permission from ref. [47]. Copyright 2010 Elsevier], respectively.

few visualizations from scratch, without any biased selections. Hence, a screening process is essential for choosing the representative materials for visualizing. We start with figuring out the simplest chemistry formulae of all materials in the ECD-cubic database, to specify the differences among materials. The results show that there are 507 types of formulae with the simplest chemistry in the ECD-cubic database, where the proportions of the formulae ABC_2, ABC_2D_6 and AB_3 ranking top three with the value of 24.92%, 12.11% and 9.86%, respectively. Besides, the total proportions of the top 13 types account for 77% and each of them contributes more than 1%, thus screened as our target materials. Ultimately, fourteen materials with well-characterized experimental or computational images in the literature, are chosen and their corresponding electronic charge density $\rho(r)$[42–54] are shown in Figs. 4–7. To further clarify, the patterns of each material are framed in a black rectangle, where the left panel shows our simulated pattern while the right panel is the reference reprinted from previous
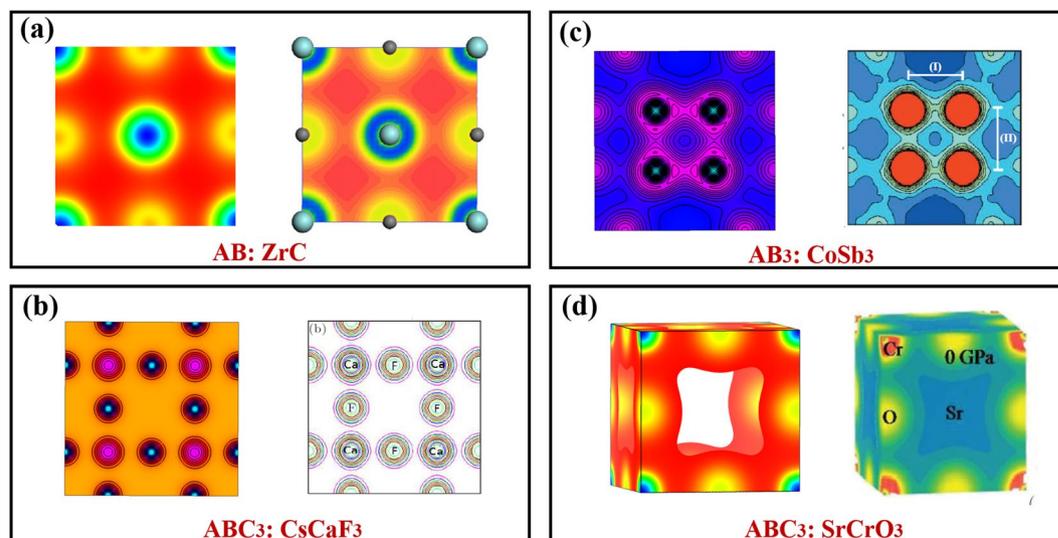
Fig. 6 The calculated patterns of electronic charge density $\rho(r)$ of (**a**) AB type, ZrC, [The right panel reprint with permission from ref. [46]. Copyright 2011 Elsevier], (**b**) AB$_3$ type, CoSb$_3$ [The right panel reprint with permission from ref. [55]. Copyright 2007 American Physical Society] and (**c,d**) ABC$_3$ type, CsCaF$_3$ [The right panel reprint with permission from ref. [43]. Copyright 2012 IOP Science], and SrCrO$_3$[The right panel reprint with permission from ref. [54]. Copyright 2020 IOP Science], respectively.

studies. Notably, the space group, elements, the size of unit cells or structures of all the selected materials are not restricted, thus making the technical validation reliable.

The selected materials with diverse crystal structures and elements, cover extensive research areas, such as the full Heusler alloys Ni$_2$MnSn (mp-20440)[51], the host-guest thermoelectric materials CoSb$_3$[54,55] (mp-1317), the alkaline tetrahydroborides NaBH$_4$ (mp-976181)[50], the spinel oxides MgAl$_2$O$_4$ (mp-3536)[49], even the rare earth pyrochlores Ho$_2$Ti$_2$O$_7$ (mp-33948)[53], etc. Each material stands for one type, which has similar structures yet different chemical constituents. For example, the ABC type[47] mainly consists of Nowotny-Juza A^{I}B^{II}C^V, where A^I = Li, Na, Cu, Ag; B^{II} = Be, Mg, Zn, Cd; C^V = N, P, As, Sb, Bi. Meanwhile, the typical materials of ABC$_3$ and AB$_2$ type are cubic perovskites as well as IIA–IV antifluorite compounds[48], respectively. This clearly proves that the materials we screened out could give comprehensive descriptions about the whole database, even if the limited number of $\rho(r)$ of materials are visualized. We observe that all the electronic charge density $\rho(r)$ patterns based on our calculations are in good agreement with those in the previous studies, confirming the technical validation of our results.

Next, several sources of inaccuracy during the dataset construction need to be discussed. The first limitation is that the magnetic orders of each case are not included in our massive calculations, mainly due to the absence of experimental data. In fact, the electron spin whether in ferromagnetic (FM) or antiferromagnetic (AFM) configurations, almost has no influence on the total electronic charge density ($\rho = \rho_{spinup} + \rho_{spindown}$) concerned there, yet slightly affects the spin electronic charge density ($\rho_{spin} = \rho_{spinup} - \rho_{spindown}$) in some cases. The reason is that the electron bonding (or antibonding) energy scale is on the order of a few electron volt (eV), while the spin polarization energy scale is usually a few tens of millielectron volt (meV). For example, except for considering the fine structure near the nuclei, the spin electronic charge density of CaMnO$_3$[56] shows much similarities between its AFM and FM phases, which is akin to that of other cubic perovskites, namely the KMnF$_3$[57,58], KNiF$_3$[59], KCuF$_3$[60], etc.

The second drawback is that the DFT calculation is conducted at absolute zero temperature (0 K), entirely ignoring the temperature effects on both electron and ion subsystems. This may induces some inconsistency between the theoretical and experimental results, especially when the latter is performed under high temperature. For instance, the covalent bond strength of NiO is increased with the sintering temperature, which can be clearly seen through its electronic charge density mapping[61]. The lattice constants of many materials would increase a bit from 0 K to finite temperature, which usually results in slight change in the spatial distribution of electronic charge density.

Third, some materials would undergo phase transition from 0 K to finite temperature. A typical example is that some cubic halide perovskites in ABC$_3$ formula, where the B-sites ions possess s^2 long pair electrons, such as CsPbX$_3$ and CsSnX$_3$, exhibit dynamic off-centering effect[62,63]. The ion position fluctuates between eight energetically favourable asymmetric configurations. Hence, the B site ions in these cubic systems is not located at the center of an octahedral coordination environment[64]. Such phenomenon is enhanced with the reduced temperature, which is beyond the discussion in this study.

Finally, the U parameters may induce some discrepancy of $\rho(r)$ for strong correlation materials, such as transition metal oxides. However, we did not inherit the U parameters in this study due to the following reasons. First, only a few U parameters are provided in the Materials Project for transition metal oxides. This will make the calculations inconsistent if we only add the U parameters for some materials. Second, the U value is
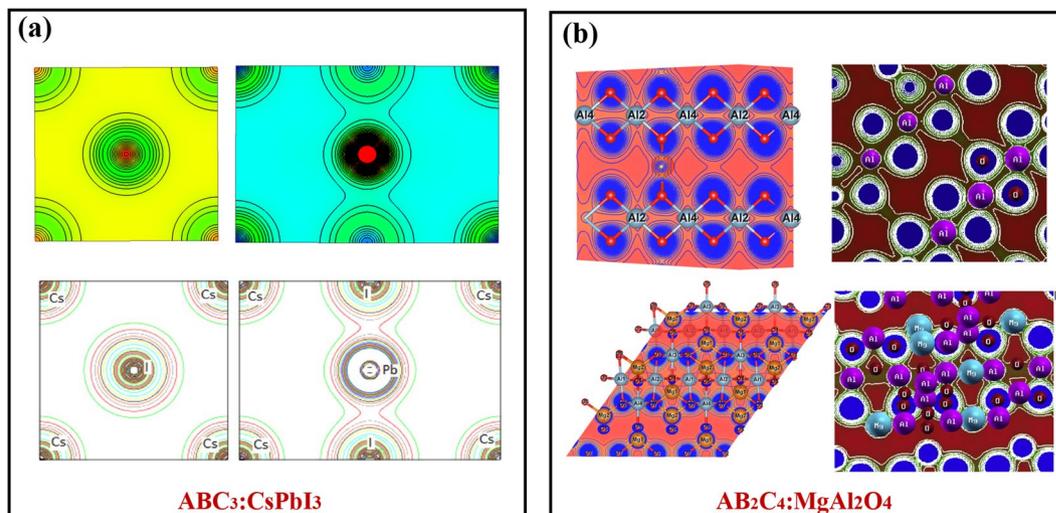
Fig. 7 The calculated patterns of electronic charge density $\rho(r)$ of (**a**) ABC$_3$ type, CsPbI$_3$ [The right panel reprint with permission from ref. [44]. Copyright 2011 Elsevier] and (**b**) AB$_2$C$_4$ type, MgAl$_2$O$_4$ [The right panel reprint with permission from ref. [49]. Copyright 2014 Elsevier].

an empirical parameter in practice, so its value can be quite different across different studies or databases[65]. We would like to emphasize that there is no universal U values for any material that can well reproduce all physical properties, including lattice constants, elastic coefficients, band dispersion, phonon spectrum, etc. Hence, the choice of the U values may be determined case-by-case, which is beyond the scope discussed here.

Code availability

Python-language-based codes for converting the JSON file to the original CHG file generated by VASP of each case are provided, which can be downloaded through Figshare[39] and our Carolina Materials Database (http://www.carolinamatdb.org/).

References

1. Grisafi, A. *et al*. Transferable Machine-Learning Model of the Electron Density. *ASC Central Sci* **5**, 57–64 (2019).
2. Fabrizio, A., Grisafi, A., Meyer, B., Ceriotti, M. & Corminboeuf, C. Electron density learning of non-covalent systems. *Chem. Sci.* **10**, 9424–9432 (2019).
3. Krawczuk, A. & Macchi, P. Charge density analysis for crystal engineering. *Chem. Cent. J.* **8**, 68 (2014).
4. Tolborg, K. & Iversen, B. B. Electron Density Studies in Materials Research. *Chem.-Eur. J.* **25**, 15010–15029 (2019).
5. Gilson, M. K. & Honig, B. H. Calculation of electrostatic potentials in an enzyme active-site. *Nature* **330**, 84–86 (1987).
6. Porezag, D. & Pederson, M. R. Infrared intensities and Raman-scattering activities within density-functional theory. *Phys. Rev. B* **54**, 7830–7836 (1996).
7. Choudhary, K. *et al*. High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. *npj Comput. Mater.* **6**, 64 (2020).
8. Choudhary, K. *et al*. Computational scanning tunneling microscope image database. *Sci. Data* **8**, 57 (2021).
9. Overgaard, J. *et al*. Relationships between Electron Density and Magnetic Properties in Water-Bridged Dimetal Complexes. *Inorg. Chem.* **53**, 11531–11539 (2014).
10. Zhao, Y. *et al*. Predicting Elastic Properties of Materials from Electronic Charge Density Using 3D Deep Convolutional Neural Networks. *J. Phys. Chem. C* **124**, 17262–17273 (2020).
11. Shen, J.-X., Horton, M. & Persson, K. A. A charge-density-based general cation insertion algorithm for generating new Li-ion cathode materials. *npj Comput. Mater.* **6**, 161 (2020).
12. Schmokel, M. S. *et al*. Atomic properties and chemical bonding in the pyrite and marcasite polymorphs of FeS2: a combined experimental and theoretical electron density study. *Chem. Sci.* **5**, 1408–1421 (2014).
13. Coppens, P., Iversen, B. & Larsen, F. K. The use of synchrotron radiation in X-ray charge density analysis of coordination complexes. *Coord. Chem. Rev.* **249**, 179–195 (2005).
14. Chopra, D. Advances in Understanding of Chemical Bonding: Inputs from Experimental and Theoretical Charge Density Analysis. *J. Phys. Chem. A* **116**, 9791–9801 (2012).
15. Bader, R. F. W. Atoms in Molecules: A Quantum Theory. *J. Mol. Struc.-theochem* **360**, 1–3 (1994).
16. Bader, R. F. W. A quantum-theory of molecular-structure and its applications. *Chem. Rev.* **91**, 893–928 (1991).
17. Schmokel, M. S., Overgaard, J. & Iversen, B. B. Experimental Electron Density Studies of Inorganic Materials. *Z. Anorg. Allg. Chem.* **639**, 1922–1932 (2013).
18. Agrawal, A. & Choudhary, A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Mater.* **4**, 053208 (2016).
19. Agrawal, A. & Choudhary, A. Deep materials informatics: Applications of deep learning in materials science. *Mrs Commun.* **9**, 779–792 (2019).
20. Cai, J., Chu, X., Xu, K., Li, H. & Wei, J. Machine learning-driven new material discovery. *Nanoscale Adv.* **2**, 3115–3130 (2020).
21. Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M. & Fazzio, A. From DFT to machine learning: recent approaches to materials science-a review. *J. Phys. Mater.* **2**, 3 (2019).

22. Morgan, D. & Jacobs, R. in Annual Review of Materials Research, Vol 50, 2020 Vol. 50 Annual Review of Materials Research (ed D. R. Clarke) 71–103 (2020).

23. Schleder, G. R., Padilha, A. C. M., Rocha, A. R., Dalpian, G. M. & Fazzio, A. Ab lnitio Simulations and Materials Chemistry in the Age of Big Data. *J. Chem. Inf. Model.* **60**, 452–459 (2020).

24. Brockherde, F. *et al.* Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **8**, 872 (2017).

25. Dick, S. & Fernandez-Serra, M. Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Commun.* **11**, 3509 (2020).

26. Chandrasekaran, A. *et al.* Solving the electronic structure problem with machine learning. *npj Comput. Mater.* **5**, 22 (2019).

27. Kamal, D., Chandrasekaran, A., Batra, R. & Ramprasad, R. A charge density prediction model for hydrocarbons using deep neural networks. Mach. *Learn.: Sci. Technol.* **1**, 025003 (2020).

28. Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **11**, 2336–2347 (2020).

29. Tsubaki, M. & Mizoguchi, T. Quantum Deep Field: Data-Driven Wave Function, Electron Density Generation, and Atomization Energy Prediction and Extrapolation with Machine Learning. *Phys. Rev. Lett.* **125**, 206401 (2020).

30. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

31. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci* **68**, 314–319 (2013).

32. Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

33. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

34. Choudhary, K., Ansari, J. N., Mazin, I. I. & Sauer, K. L. Density functional theory-based electric field gradient database. *Sci. Data* **7**, 362 (2020).

35. Choudhary, K., Cheon, G., Reed, E. & Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Phys. Rev. B* **98**, 014107 (2018).

36. Choudhary, K. *et al.* The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Comput. Mater.* **6**, 173 (2020).

37. Klimes, J., Bowler, D. R. & Michaelides, A. Chemical accuracy for the van der Waals density functional. *J. Phys.-Condens. Mat.* **22**, 022201 (2010).

38. Chen, C. *et al.* A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **10**, 1903242 (2020).

39. Fancy, Q. Wea Large scale dataset of real space electronic charge density of cubic inorganic materials from density functional theory (DFT) calculations. *figshare* https://doi.org/10.6084/m9.figshare.c.5368343 (2021).

40. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phy.: Mater.* **2**, 036001 (2019).

41. Choudhary, K. *et al.* Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms. *Sci. Data* **5**, 180082 (2018).

42. Subhan, F. *et al.* Elastic and optoelectronic properties of CaTa2O6 compounds: Cubic and orthorhombic phases. *J. Alloys Compd.* **785**, 232–239 (2019).

43. Babu, K. E., Veeraiah, A., Swamy, D. T. & Veeraiah, V. First-Principles Study of Electronic Structure and Optical Properties of Cubic Perovskite CsCaF3. *Chin. Phys. Lett.* **29**, 117102 (2012).

44. Murtaza, G. & Ahmad, I. First principle study of the structural and optoelectronic properties of cubic perovskites CsPbM3 (M=Cl, Br, I). *Physica B* **406**, 3222–3229 (2011).

45. Azam, S. & Reshak, A. H. Study of electronic structure, charge density, Fermi energy and optical properties of Cs2KTbCl6 and Cs2KEuCl6. *Physica B* **431**, 102–108 (2013).

46. Li, H. *et al.* Structural, elastic and electronic properties of transition metal carbides TMC (TM = Ti, Zr, Hf and Ta) from first-principles calculations. *Solid State Commun.* **151**, 602–606 (2011).

47. Bouhemadou, A., Khenata, R., Rached, D. & Amrani, B. Theoretical prediction of the elastic, electronic and optical properties of the filled tetrahedral semiconductor alpha-LiMgSb. *Comput. Mater. Sci* **49**, 64–69 (2010).

48. Benhelal, O. *et al.* First-principles calculations of the structural, electronic and optical properties of IIA-IV antifluorite compounds. *Phys. Status Solidi B* **242**, 2022–2032 (2005).

49. Reshak, A. H., Khan, S. A. & Alahmed, Z. A. Investigation of electronic structure and optical properties of MgAl2O4: DFT approach. *Opt. Mater.* **37**, 322–326 (2014).

50. Remhof, A. *et al.* Rotational motion of BH4 units in MBH4 (M=Li,Na,K) from quasielastic neutron scattering and density functional calculations. *Phys. Rev. B* **81**, 214304 (2010).

51. Deb, A. *et al.* Spin-dependent electron momentum density in the Ni2MnSn Heusler alloy. *Phys. Rev. B* **63**, 205115 (2001).

52. Wentzcovitch, R. M., Chang, K. J. & Cohen, M. L. Electronic and structural-properties of BN and BP. *Phys. Rev. B* **34**, 1071–1079 (1986).

53. Chernyshev, V. A., Petrov, V. P., Nikiforov, A. E., Agzamova, P. A. & Avram, N. M. Elastic properties of rare earth pyrochlores R2Ti2O7 (R = Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu): Ab initio calculations. *Opt. Mater.* **72**, 565–570 (2017).

54. Hasan, M. Z. & Rasheduzzaman, M. & Monower Hossain, K. Pressure-dependent physical properties of cubic SrBO3 (B = Cr, Fe) perovskites investigated by density functional theory*. *Chin. Phys. B* **29**, 123101 (2020).

55. Ohno, A. *et al.* X-ray charge density study of chemical bonding in skutterudite CoSb3. *Phys. Rev. B* **76**, 064119 (2007).

56. Fava, F. F., D'Arco, P., Orlando, R. & Dovesi, R. A quantum mechanical investigation of the electronic and magnetic properties of perovskite. *J. Phys.: Condens. Matter* **9**, 489 (1997).

57. Dejongh, L. J. & Block, R. Exchange interactions in some 3D-metal ionic compounds: I. The 180 superexchange in the 3d-metal fluorides XMF3 and X2MF4(X = K, Rb, Tl; M = Mn, Co,Ni). *Physica B & C* **79**, 568–593 (1975).

58. Pisani, C., Dovesi, R., Erba, A. & Giannozzi, P. in Modern Charge-Density Analysis (eds C. Gatti & P. Macchi) 79–132 (Springer Netherlands, 2012).

59. Ricart, J., Dovesi, R., Roetti, C. & Saunders, V. Electronic and magnetic structure of KNiF3 perovskite. *Phys. Rev. B* **52**, 2381 (1995).

60. Towler, M., Dovesi, R. & Saunders, V. R. Magnetic interactions and the cooperative Jahn-Teller effect in KCuF 3. *Phys. Rev. B* **52**, 10150 (1995).

61. Saravanakumar, S., Saravanan, R. & Sasikumar, S. Effect of sintering temperature on the magnetic properties and charge density distribution of nano-NiO. *Chem. Pap.* **68**, 788–797 (2014).

62. Remsing, R. C. & Klein, M. L. A new perspective on lone pair dynamics in halide perovskites. *APL Mater.* **8**, 050902 (2020).

63. Fabini, D. H. *et al.* Dynamic Stereochemical Activity of the Sn2+ Lone Pair in Perovskite CsSnBr3. *J. Am. Chem. Soc.* **138**, 11820–11832 (2016).

64. Radha, S. K., Bhandari, C. & Lambrecht, W. R. L. Distortion modes in halide perovskites: To twist or to stretch, a matter of tolerance and lone pairs. *Phys. Rev. Mater.* **2**, 063605 (2018).

65. Jie, J. *et al.* A new MaterialGo database and its comparison with other high-throughput electronic structure databases for their predicted energy band gaps. *Sci. China Technol. Sc.* **62**, 1423–1430 (2019).

Acknowledgements

Author contributions

Conceptualization, F.Q.W., M.H.; methodology, F.Q.W., M.H.; DFT calculation, F.Q.W., K.C., M.H.; investigation, F.Q.W., Y.L., M.H.; original draft preparation, F.Q.W.; manuscript editing and revision, F.Q.W., J.H., M.H.; supervision, F.Q.W., M.H.; funding acquisition, Y.L., J.H., M.H.

Competing interests

The authors declare no competing interest.

Additional information

Correspondence and requests for materials should be addressed to F.Q.W. or M.H.

Reprints and permissions information is available at www.nature.com/reprints.

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