

MatCloud, a high-throughput computational materials infrastructure: Present, future visions, and challenges

Xiaoyu Yang(杨小渝), Zongguo Wang(王宗国), Xushan Zhao(赵旭山), Jianlong Song(宋健龙), Chao Yu(虞超), Jiaxing Zhou(周嘉欣), Kai Li(李凯)
Citation: Chin. Phys. B . 2018, 27(11): 110301. **doi:** 10.1088/1674-1056/27/110301

Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Spin manipulation in semiconductor quantum dots qubit

Ke Wang(王柯), Hai-Ou Li(李海欧), Ming Xiao(肖明), Gang Cao(曹刚), Guo-Ping Guo(郭国平)
Chin. Phys. B . 2018, 27(9): 090308. **doi:** 10.1088/1674-1056/27/9/090308

Solid-state quantum computation station

Fanming Qu(屈凡明), Zhongqing Ji(姬忠庆), Ye Tian(田野), Shiping Zhao(赵士平)
Chin. Phys. B . 2018, 27(7): 070301. **doi:** 10.1088/1674-1056/27/7/070301

Quantum information processing with nitrogen-vacancy centers in diamond

Gang-Qin Liu(刘刚钦), Xin-Yu Pan(潘新宇)
Chin. Phys. B . 2018, 27(2): 020304. **doi:** 10.1088/1674-1056/27/2/020304

Qubits based on semiconductor quantum dots

Xin Zhang(张鑫), Hai-Ou Li(李海欧), Ke Wang(王柯), Gang Cao(曹刚), Ming Xiao(肖明), Guo-Ping Guo(郭国平)
Chin. Phys. B . 2018, 27(2): 020305. **doi:** 10.1088/1674-1056/27/2/020305

A high-fidelity memory scheme for quantum data buses

Bo-Yang Liu(刘博阳), Wei Cui(崔巍), Hong-Yi Dai(戴宏毅), Xi Chen(陈希), Ming Zhang(张明)
Chin. Phys. B . 2017, 26(9): 090303. **doi:** 10.1088/1674-1056/26/9/090303

TOPICAL REVIEW — Physics research in the materials genome

MatCloud, a high-throughput computational materials infrastructure: Present, future visions, and challenges*

Xiaoyu Yang(杨小渝)[1,2,†], Zongguo Wang(王宗国)[1], Xushan Zhao(赵旭山)[1], Jianlong Song(宋健龙)[1], Chao Yu(虞超)[1,2], Jiaxin Zhou(周嘉欣)[1,2], and Kai Li(李凯)[1]

[1] *Computer Network Information Centre, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *University of Chinese Academy of Sciences, Beijing* 100049, *China*

(Received 5 June 2018; revised manuscript received 17 August 2018; published online 19 October 2018)

MatCloud provides a high-throughput computational materials infrastructure for the integrated management of materials simulation, data, and computing resources. In comparison to AFLOW, Material Project, and NoMad, MatCloud delivers two-fold functionalities: a computational materials platform where users can do on-line job setup, job submission and monitoring only via Web browser, and a materials properties simulation database. It is developed under Chinese Materials Genome Initiative and is a China own proprietary high-throughput computational materials infrastructure. MatCloud has been on line for about one year, receiving considerable registered users, feedbacks, and encouragements. Many users provided valuable input and requirements to MatCloud. In this paper, we describe the present MatCloud, future visions, and major challenges. Based on what we have achieved, we will endeavour to further develop MatCloud in an open and collaborative manner and make MatCloud a world known China-developed novel software in the pressing area of high-throughput materials calculations and materials properties simulation database within Material Genome Initiative.

Keywords: high-throughput materials simulation, materials informatics

PACS: 03.67.Lx, 93.85.Bc **DOI:** 10.1088/1674-1056/27/11/110301

1. Introduction

MatCloud (http://matcloud.cnic.cn) provides high-throughput computational materials infrastructure for materials simulation, data, and computing resources integrated management.[1] It has been developed by borrowing the idea of cloud computing, hence the user only needs a Web browser to setup jobs, submit jobs, and do data management. MatCloud is directly connected to a computing cluster, a central file system, and a materials properties database. Currently, MatCloud fully supports the Vienna *ab initio* simulation package (VASP) and the user must have a valid VASP license (i.e.,VASP 5.X) before running VASP simulations through MatCloud.

There exist some tools and technologies that support high-throughput materials simulation and data management, such as AFLOW,[2] Material Project,[3] and so on. However, the users usually have to download and install them onto their local computers. The cloud-based mode for licensed users running large amount of first-principles simulations is not well supported and does not provide a graphical environment for users to create customised workflows, submit and monitor workflows and simulations.

While MatCloud is a computational platform where users can setup, submit and monitor a job via a Web browser, and once job finishes all the required data are preserved, MatCloud is also a materials properties simulation database that provides long-term data storage and archival of simulated data. The data curation activity is transparently integrated into the workflow of creating data, managing data, and other digital assets in an end-to-end manner without direct human control, rather than requiring curation activities to happen at the post simulation stage separately. MatCloud also provides a workflow framework to facilitate the automation of multi-scale materials simulations.

MatCloud has been on-line for about one year since April of 2017. In just this one year, MatCloud has received a considerable number of registered users, and it has received their feedbacks and encouragement. The number of MatCloud registered users exceeds 800, and that of organisations exceeds 350. Many users also provided valuable input and requirements. They hope that MatCloud can provide more powerful functionalities. This paper reviews the present state of MatCloud, and it provides a future vision and analyses some major challenges.

2. Present

2.1. MatCloud architecture

We think the integrated materials design e-infrastructure should consist the following key constructs (Fig. 1): (i) work-

*Project supported by the National Key Research and Development Program of China (Grant Nos. 2017YFB0701702 and 2016YFB0700501), the National Natural Science Foundation of China (Grant Nos. 61472394 and 11534012), and Science and Technology Department of Sichuan Province, China (Grant No. 2017JZ0001).

†Corresponding author. E-mail: kxy@cnic.cn

© 2018 Chinese Physical Society and IOP Publishing Ltd

flow system including a front-end workflow designer and a back-end workflow engine, (ii) material property database, (iii) file system that stores the most of the important simulation files, (iv) crystal structure modelling that handles producing large amount of crystal structures files, (v) data extraction engine, and (vi) job scheduler that is responsible for scheduling simulation jobs to computing clusters that may situate at geographically different locations. In addition, MatCloud can also support the setup of exchange–correlation functions, pseudopotential, and k-points. The whole simulation process with post-simulation calculations can be automatically completed without human intervention. The key constructs of workflow designer and material property database are depicted as follows. The user-concerned security management of simulation data is also introduced.

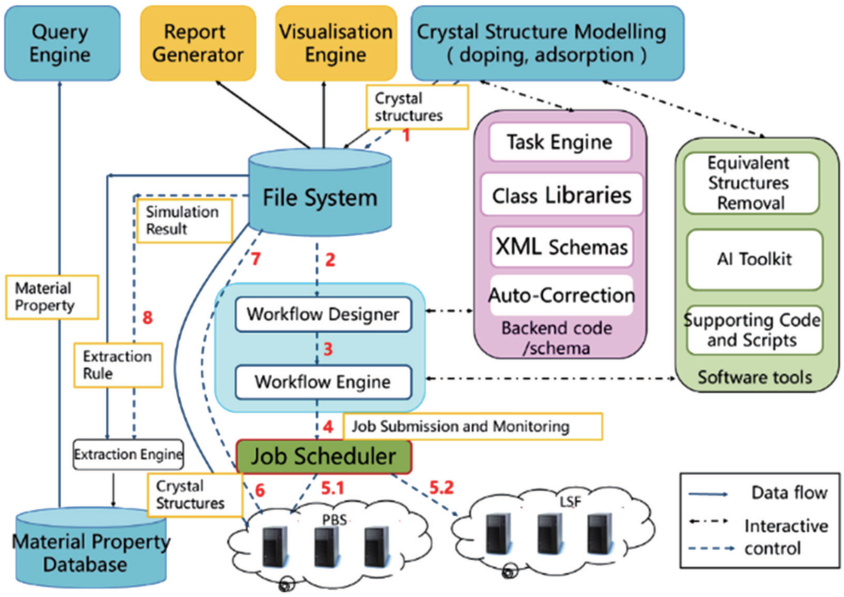


Fig. 1. (color online) Integrated materials design: effective integration of material simulation, materials database, and materials big data.[1]

2.2. MatCloud workflow system

MatCloud provides a graphical user interface (GUI) based environment for users to intuitively create, enact, and monitor a workflow, as shown in Fig. 2. A set of toolkits have been provided, including a dataset toolkit, crystal structure modelling toolkit, simulation toolkit, analysis toolkit, and workflow templates. The dataset toolkit is used as a container to hold crystal structures to be simulated. The crystal structure modelling toolkit is mainly responsible for generating simulation input for various manipulations on initial crystal structures (e.g., create a super cell, transform to a primitive cell, doping, surface, adsorption). The simulation toolkit provides the break-down simulation activities (e.g., geometry optimisation, static calculation).

Creating a workflow is straightforward. Two approaches for creating a workflow have been provided. One approach is to start from scratch and another is to use a pre-defined template. Regarding starting from scratch, users only need to drag & drop the data container component, the required individual simulation component to the canvas, and connect them by lines. Regarding the use of a template, the users can just select the template with some switches on (e.g., geometry optimisation, elastic constant), and then drag & drop it onto the canvas.

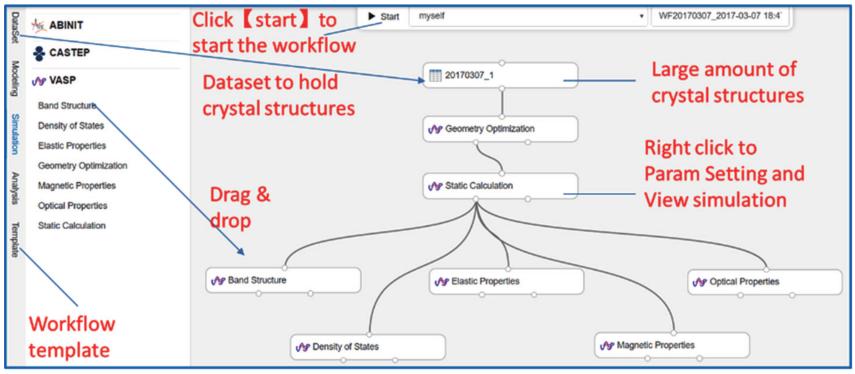


Fig. 2. (color online) MatCloud workflow designer.[1]

Click start button to enact the workflow. Before starting the workflow, the users can set simulation parameters (e.g., precision, exchange–correlation functions, cut-off energy, *k*-points) for each simulation. Once satisfied with the setting, the users just click the start button to enact the workflow. The workflow will then automate the procedures of job submission, monitoring, property extraction & calculation, and store properties data into the database.

2.3. MatCloud workflow classification

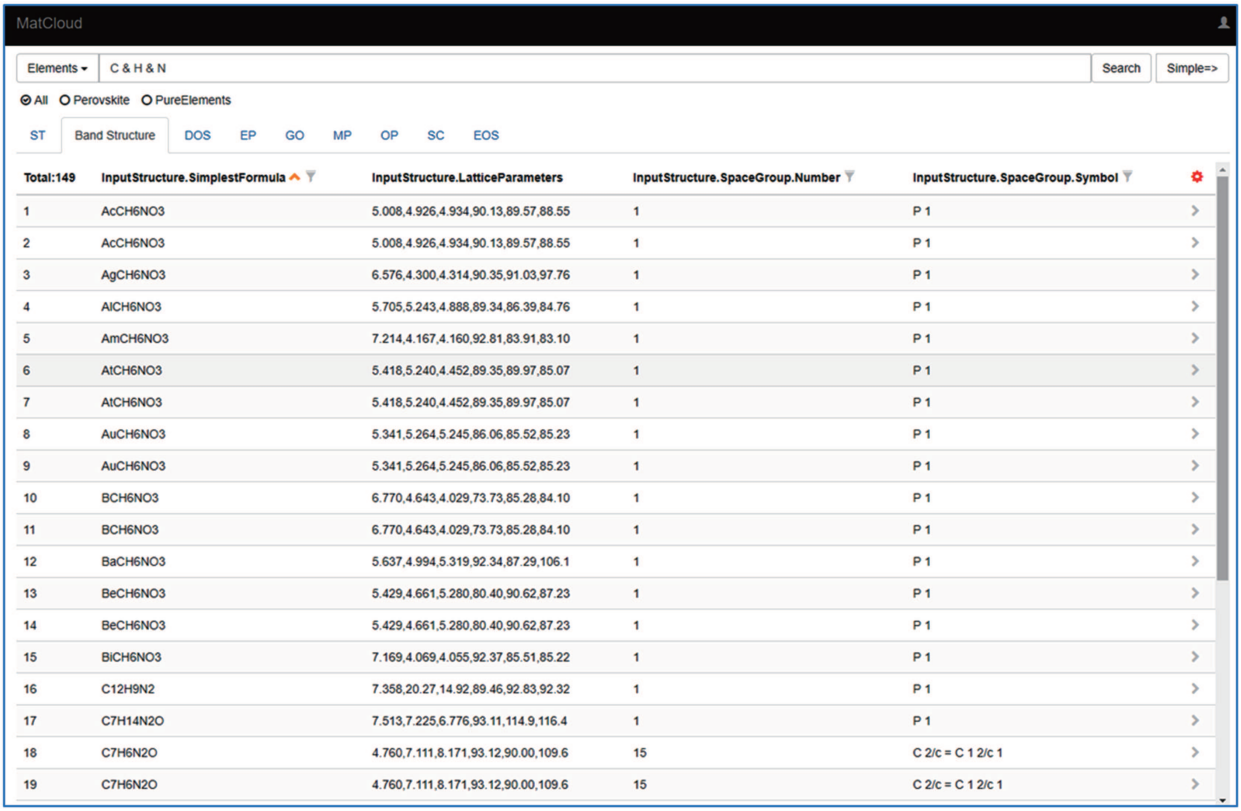
One goal of material simulation is to predict material properties, hence how to effectively obtain material properties from large amount of DFT simulations is vital. The approach of simulation properties acquisition can be complicated because it varies with running DFT simulation one or more times.[1] For example, while some properties can be obtained by only running DFT simulation once (e.g., total energy, force constant) and based on which more properties can be derived (e.g., elastic modulus, band gap*), some material properties can only be acquired by running DFT simulation several times (e.g., equation of state, phonon spectrum) and more properties can be derived (e.g., diffusion coefficient, temperature dependent thermodynamic properties such as Gibbs free energy).[1]

To provide a unified approach, we have classified the workflows into the following four categories: (i) properties extracted by single DFT simulation, (ii) properties derived through theoretical/empirical models over single DFT simulation, (iii) properties acquired by aggregating multiple DFT simulations, and (iv) properties derived through theoretical/empirical models over multiple DFT simulations.[1]

2.4. A case study: building materials simulation database for data mining

Traditionally, material data is acquired by experiments. However, the acquisition of material data from experiment data is sometimes expensive and inefficient and it can sometimes be difficult to get data only by experiment (e.g., doping at a low concentration). Consequently, materials simulation can be of assistance in acquiring materials data.

Materials simulation can also be used to obtain materials data as a counterpart to materials data obtained by experiment. One of major uses of MatCloud is that it can help users to build a material simulation database efficiently. As described previously, acquiring a material property involves several procedures (e.g., job setup, DFT calculations, data extraction). MatCloud can automate these procedures without human interactions. The high-throughput features of MatCloud can also facilitate the calculation of large amount of crystal structures.



Total:149	InputStructure.SimplestFormula	InputStructure.LatticeParameters	InputStructure.SpaceGroup.Number	InputStructure.SpaceGroup.Symbol
1	AcCH6NO3	5.008,4.926,4.934,90.13,89.57,88.55	1	P 1
2	AcCH6NO3	5.008,4.926,4.934,90.13,89.57,88.55	1	P 1
3	AgCH6NO3	6.576,4.300,4.314,90.35,91.03,97.76	1	P 1
4	AlCH6NO3	5.705,5.243,4.888,89.34,86.39,84.76	1	P 1
5	AmCH6NO3	7.214,4.167,4.160,92.81,83.91,83.10	1	P 1
6	AtCH6NO3	5.418,5.240,4.452,89.35,89.97,85.07	1	P 1
7	AtCH6NO3	5.418,5.240,4.452,89.35,89.97,85.07	1	P 1
8	AuCH6NO3	5.341,5.264,5.245,86.06,85.52,85.23	1	P 1
9	AuCH6NO3	5.341,5.264,5.245,86.06,85.52,85.23	1	P 1
10	BCH6NO3	6.770,4.643,4.029,73.73,85.28,84.10	1	P 1
11	BCH6NO3	6.770,4.643,4.029,73.73,85.28,84.10	1	P 1
12	BaCH6NO3	5.637,4.994,5.319,92.34,87.29,106.1	1	P 1
13	BeCH6NO3	5.429,4.661,5.280,80.40,90.62,87.23	1	P 1
14	BeCH6NO3	5.429,4.661,5.280,80.40,90.62,87.23	1	P 1
15	BiCH6NO3	7.169,4.069,4.055,92.37,85.51,85.22	1	P 1
16	C12H9N2	7.358,20.27,14.92,89.46,92.83,92.32	1	P 1
17	C7H14N2O	7.513,7.225,6.776,93.11,114.9,116.4	1	P 1
18	C7H6N2O	4.760,7.111,8.171,93.12,90.00,109.6	15	C 2/c = C 1 2/c 1
19	C7H6N2O	4.760,7.111,8.171,93.12,90.00,109.6	15	C 2/c = C 1 2/c 1

Fig. 3. (color online) Using MatCloud to build perovskite material simulation database.

*The value of band gap cannot be directly extracted from VASP output. This requires further calculation based on eigenvalues and Fermi energy values extracted from the VASP output.

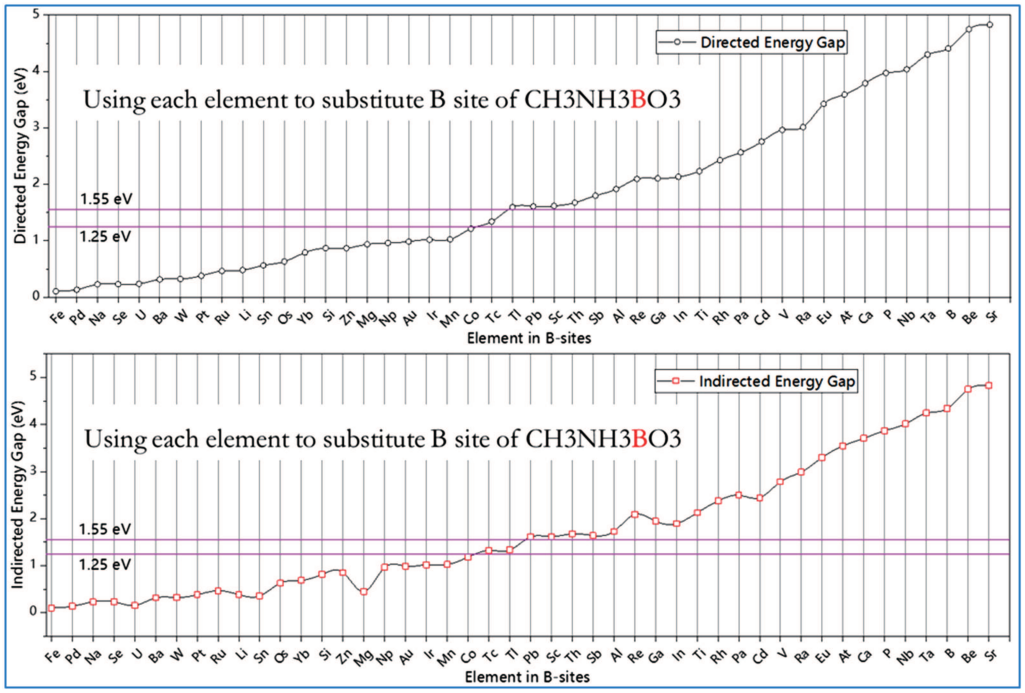


Fig. 4. (color online) Using the developed perovskite materials database, we can quickly find a pattern of element substitution in the B site of $CH_3NH_3BO_3$.

Currently, MatCloud is being used to build a perovskite material simulation database (mainly targeting photovoltaic perovskite materials), photocatalytic materials database, semiconductor storage materials, and liquid metal database. For example, we have used MatCloud to calculate more than 30 materials properties of 243 standard perovskite crystal structures using 840 CPU cores (14 CPU core per simulation) and store them in a database. All these tasks were finished within 2 days without human interactions. We have also heard that MatCloud was used to acquire properties for 153 standard perovskite compounds within 5 h using 840 CPU cores, where most of the time was consumed by VASP calculation.

By using MatCloud, a perovskite material simulation dataset is primarily developed, as shown in Fig. 3. The database can be used for data mining for new discoveries. For example, using the database we developed, a pattern of how direct band gap and indirect gap vary with element substitution in the B site of $CH_3NH_3BO_3$ has been discovered, as shown in Fig. 4.

3. Future visions

From the MatCloud user training and survey, we understand that users hope MatCloud can provide more useful functionalities, as follows: crystal structures modelling, high-throughput screening, more properties calculation, multi-scale simulation, materials simulation database, and simulation ecosystem.

3.1. Crystal structures modelling

Crystal structure modelling requires a crystal unit cell to be built from its constituent parts or modifying existing crystal structures by a number of ways (e.g., doping, surface operation). The development of crystal structures modelling includes the following aspects: (i) create a new structure, or derive a structure based on an initial crystal structure; (ii) source of initial crystal structure; (iii) modelling type such as doping, surface operation, and so on; and (iv) modelling approach: graphically interactive, and so on. All of these aspects require deep thinking. Currently, MatCloud only supports substitutional doping, and some basic operations, including creation of supercell, conversion between unit cell primitive conventional representation, generating large amount of new crystal structures through deformation manipulation over a unit cell, and so on. MatCloud will support more substitutional doping modelling, interstitial doping modelling, and atom adsorptions modelling. Surface modelling, nanostructure modelling, and interface modelling are also important tools in modelling physical phenomena, and MatCloud will develop them in the future.

MatCloud will soon support the following crystal structure modelling.

MatCloud will be able to support three doping-based modelling. As a substitutional doping calculation proposed in Ref. [1], for a given crystal structure, using different dopant elements X $(x_1, x_2, \ldots, x_i)$ to replace different target element species Y $(y_1, y_2, \ldots, y_i)$ that contain certain number of atoms Z $(z_1, z_2, \ldots, z_i)$ that involve certain series of sites U $(u_1, u_2, \ldots, u_i)$ to reach different doping concentrations V

$(v_1, v_2, \ldots, v_i)$. Note that one kind of target element can be substituted by either single dopant atom or multiple different dopant atoms at the same time. The high throughout calculation can happen at loop through of X, Y, Z, U, and V.

Currently, MatCloud only supports the substitutional doping case of using one dopant atom (X is fixed) to replace one target species (Y is fixed) that contains certain number of atoms that involve a series possible sits to reach one concentration (V is fixed). The main restriction of this approach is caused by the doping-filtering working approach used on MatCloud[4] in the process of doped structures generation. The maximum computational time happens at the stage where half of the total number of target atoms are doped.

MatCloud will be able to support interstitial doping modelling (i.e., interstitial doping builder). As an interstitial doping builder, for a given crystal structure, placing different dopant elements (S) in the symmetrically distinct interstitial sites. To reach different doping concentrations (C), certain number of interstitial sites (N) will be considered. Interstitial doping builder can implement that one interstitial site can be filled by either single dopant atom or multiple different dopant atoms at the same time. At the same time, one dopant element can fill in one or more interstitial sites. The high throughout calculation can happen at loop through of S, C, and N. Currently, MatCloud only supports the interstitial doping case of using one dopant atom (S is fixed) to fill certain number of interstitial sites to reach one or more concentration.

MatCloud will be able to support atom adsorption modelling. The aim of adsorption modelling is to get a slab according to the specific miller index for a given crystal structure. Then, we find the surface sites on the slab and put dopant elements (S) onto one or more sites to identify the adsorption sites on the slab by calculation. MatCloud now supports one dopant element onto certain number of surface sites.

MatCloud will be able to support surface modelling, which can include two types: the first is generating a slab according to the miller index for a given crystal structure, and the second is constructing a 2D periodic system. For the former case, user can obtain miller planes using a tool provided by MatCloud named cleave the surface. For the latter case, group and lattice parameters should be supplied. This tool is going to be developed soon.

Nanostructure modelling: MatCloud will develop some nanostructure builders, including single-/multi-walled nanotubes (1D) and nanoclusters (0D), to construct atomistic models of the structures of relevance to nanotechnology.

Interface modeling: MatCloud will also develop a tool to build an interface from two or more source crystals or surfaces.

3.2. Supporting more high-throughput screening functionalities

Cluster expansion based: Currently, MatCloud only supports formation energy based high-throughput screening by using cluster expansion approach. The core idea of cluster expansion approach is that it only calculates the total energies of some particularly selected structures through first-principles simulations, and then use those values to train a model.[5] That model can then be used to quickly predict energy values of new structures (these structures must have one parent structure), hence for energy-based screening. For multi-doped system, MatCloud currently only works for doping atoms one by one. One main restrictions of using high-throughput screening on MatCloud is the limitation of cluster expansion; i.e., it can only be applied to a fixed parent lattice structure.

Next, MatCloud will supply more high-throughput screening functions based on cluster expansion method which has been implemented in the UNCLIE,[6,7] such as vacancy concentrations, adsorption energies, also MatCloud will consider to generalize this method to multiple elements doped systems.

Data model based: Materials informatics provides a novel methodology for high-throughput screening development, and MatCloud is going to develop a new high-throughput screening method based on data mining. For a given structure system, MatCloud can be used train a structure-property model, and this model can be optimized with the increasing number of data. With this model, users can input some structures and MatCloud can quickly predict their properties by using the optimized model, hence screening for required materials.

3.3. Supporting multi-scale simulation

Multi-scale models and simulations play a significant role in the Materials Genome Initiative. For robust, accurate, predictive simulations of materials behavior, bridging materials models and passing materials-related data and information across different scales simulation are critical for the quantitative & predictive modelling to support the development of advanced materials and processes. However, a lack of acceptable linkage software and tools and this lack of versatile, user-friendly linking tools prevent the effective transmission of information between models from various length scales. Establishing an infrastructure for multiscale materials data and developing the associated APIs and standards for connecting different computational tools across length scales is highly recommended in many previous studies.

The MatCloud infrastructure has already laid a foundation to develop an atomistic material database and microstructural material database. The workflow system of MatCloud

now well supports the quantum mechanics simulation, and primarily supports the molecular dynamics simulation. It can now graphically connect different computational tools across length scales through a GUI-based interface to support multi-scale materials simulation. In the future, MatCloud will not only have the electronic precision but will also exceed its range of application from micro-scale to macro-scale. Currently, several multi-scale and multi-dimension simulation methods—including excited states modelling, thermodynamics/kinetics calculation, and morphology simulation—are under development on the basis of MatCloud workflow engine.

Apart from VASP, the support of ABINIT[8] and LAMMPS[9] through MatCloud is now under development. ABINIT is a software suite that is used to calculate the optical, mechanical, vibrational, and other observable properties of materials. ABINIT has some advanced features with perturbation theories based on DFT and many-body Green's functions. ABINIT can easily compute excited state properties via time-dependent density functional theory and many-body perturbation theory using the GW approximation and Bethe–Salpeter equation.

LAMMPS is a software package that performs classical molecular dynamics simulations. It is popular due to its versatility and support for a wide range of potential energy models, long range solvers, and simulation options. It is widely used to study the time evolution of a system of particles, typically atoms or molecules with defined properties. The fundamental steps for LAMMPS simulation include: calculation of the force on each particle as the gradient of the energy, time integration to calculate the new particle positions/velocities with respect to the force, and thermostat/barostat calculations for NPT/NVT simulations. In contrast to VASP (quantum molecular dynamics), LAMMPS (classical molecular dynamics) uses less computationally cost empirical potentials to determine the system energy allowing for better time complexities, more efficient parallel decompositions, and ultimately a capability to simulate much larger systems for longer time scales.

3.4. Supporting more properties calculation

MatCloud now offers efficient and user-friendly quantum mechanical calculations, such as the prediction of formation energy, dielectric constants, optical properties, mechanical properties, and electronic structures (band structure and density of states). MatCloud also supplies some tools to deal with the direct calculated results from VASP; for example, MatCloud can directly obtain refractive index, energy loss, extinction index, adsorption coefficient, reflection coefficient, and optical conductivity after the dielectric function calculation is finished.

In the future, MatCloud will support more materials properties calculation. The new calculations to be developed will include calculations of transition state search, phonon spectra, phonon density of states, electron-phonon coupling, thermal conductivity, electron transport, and diffusion coefficient, etc. Some calculation will also be improved, which include calculations of non-linear magnetic properties, non-linear optical properties, electron localization function, and population analysis, etc.

3.5. Materials simulation data

Provide API to support interrogation with other databases. High-throughput calculations are used to create large databases containing the calculated properties of existing and hypothetical materials. These databases can then be intelligently interrogated, searching for materials with desired properties and so removing the guesswork from materials design. There are various materials simulation databases such as Materials Project, OQMD,[10] NoMad,[11] NIMS CompES-X,[12] and so on. The ultimate goal is to enable the query of several databases simultaneously with common APIs. This would greatly benefit the materials science community (e.g., by enhancing opportunities for data mining) and clearly contribute to fostering innovation more effectively. MatCloud plans to make its data deliverable via APIs, web services, and unified materials data description. Currently, an OPTiMaDe API[13] has been proposed for materials database interrogation and MatCloud has participated the development of specification.

MatCloud will be able to support the use of artificial intelligence. Currently, MatCloud has primarily been formulated as a material property database by storing the material properties extracted from the VASP simulation output. As described previously, MatCloud will support more simulation codes such as LAMMPS, ABINIT, CASTEP, and so on. (For licensed software, users must have valid license.) The ultimate goal of materials simulation database is to assist the materials design and materials informatics can help with this.

Next, MatCloud will develop a general artificial intelligence framework for developing materials property prediction models, mainly targeting materials structure-property model. The framework should support feature variable selection, model training, model validation and use, and so on. It should also connect to and support the use of GPU clusters. The predicted value and associated parameters should be stored in a database.

MatCloud will be able to assign a unique digital identifier to materials simulation data. At the time of writing this manuscript, it is difficult to get a digital identifier to scientific data in China. For example, an organization must register

with ISTIC and CNKI first before their users can obtain digital identifiers of their data. Once registered, the users have to fill certain form(s) to apply for digital IDs, both of which may take some time. To tackle this problem, we developed a technique that can allocate a handle digital identifier to materials simulation data in a fast and efficient manner. A data or dataset labelled with this handle digital identifier can be globally resolved.

3.6. MatCloud materials simulation eco-system

Just as Materials Studio was developed by many contributors, MatCloud in the future will also be extended and developed in an open and collaborative manner by the community people from different universities and research institutes to become a world known China-developed novel software to support high-throughput and multi-scale materials simulations. This also means MatCloud will become a materials simulation eco-system, where different simulation code, user proprietary code, algorithms, etc. can be wrapped or adapted into the MatCloud framework, where his/her intellectual properties will be acknowledged.

4. Challenges

The major challenges of MatCloud future development are described in the following subsections.

4.1. How materials simulation data can be presented in a unified format

Materials data have characteristics of different scale and variety of data types including numerical data, text data, images, and animations. In developing a unified format, the following issues must to be considered: simple to understand and use; easy to manage; compatible with all type of computers, operation systems, and programming languages; fast and easy retrieval, and so on.

Because different simulation code has different simulation output format, data exchange has become a barrier. As shown in Fig. 5, we think that the properties extracted from simulation output should be presented in a unique data format that has a well-defined data schema. From the literature, it is identified that Material Information File (MIF)[14] has a potential to be adopted. Extending MatCloud to a MIF to formulate unified materials simulation data specification to meet the different simulation code presents a considerable challenge.

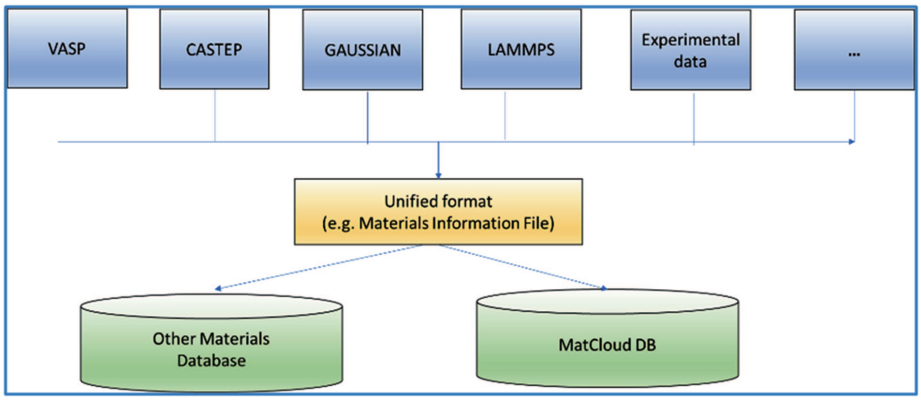


Fig. 5. (color online) A unified data representation layer in MatCloud.

4.2. Data quality management

How to ensure the quality of materials simulation data is also difficult. Experimental data will be the primary criterion for the quality management of corresponding calculation results. The reported measurement data would be collected and collated with simulation property data in the initial database. If no experimental data is collected, then key simulation parameters such as simulation precision, cut-off energy, k-points, validity of formulas/models will be checked, and data meeting certain criteria are put into the validated database. How experimental data and simulation can collate together is also difficult.

4.3. Automation of across scale simulations

Although MatCloud provides a framework for across scale simulations, the challenges still remain in bridging materials models and passing the materials data across different length scales. Currently, MatCloud can support running VASP and LAMMPS simulation code in workflow environment individually, but what would happen when bringing them is still not clear. Answering this question may involve some further investigation and development.

5. Conclusion

This paper describes the present, future visions, and challenges of MatCloud high-throughput computational materi-

als infrastructure. The current MatCloud provides a workflow system, material property database, file system, crystal structure modelling, data extraction engine, and job scheduler. The future perspective of MatCloud has been presented from the crystal structures modelling, high-throughput screening, more properties calculation, multi-scale simulation, materials simulation database&data mining, and simulation eco-system. Three major challenges in MatCloud future development have been discussed.

References

[1] Yang X Y, Wang Z G, Zhao X S, Song J L, Zhang M M and Liu H 2018 *Comput. Material Sci.* **146** 319

[2] Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor R H, Nelson L J, Hart G L W, Sanvito S, Buongiorno-Nardelli M, Mingo N and Levy O 2012 *Comput. Mater. Sci.* **58** 227

[3] Jain A, Hautier G, Moore C J, *et al.* 2011 *Comput. Mater. Sci.* **50** 2295

[4] Zhang M M and Yang X Y 2018 *Comput. Mater. Sci.* **150** 381

[5] Wang Z G, Yang X Y, Wang L G, Wang J, Zhang M M, Zhao X S, Ren J and Zeng Z 2018 *Comput. Mater. Sci.* **143** 55

[6] Lerch D, Wieckhorst O, Hart G L W, Forcade R W and Muller S 2009 *Mater. Sci. Eng.* **17** 055003

[7] Atomic Scale Simulation Final Project, http://publish.illinois.edu/atomicscale/

[8] ABINIT, https://www.abinit.org/

[9] LAMMPS, http://lammps.sandia.gov/

[10] Saal J E, Kirklin S, Aykol M, Meredig B and Wolverton C 2013 *JOM* **65** 1501

[11] NoMad, https://repository.nomad-coe.eu/

[12] NIMS CompES-X, http://compes-x.nims.go.jp/index_en.html

[13] Optimade API, http://www.optimade.org/spec/

[14] Material Information File, http://citrineinformatics.github.io/mif-documentation/