

基于材料基因组计划的计算和数据方法

杨小渝^{1,2}, 任杰^{1,2}, 王娟^{1,2}, 赵旭山¹, 王宗国¹, 宋健龙¹

1. 中国科学院计算机网络信息中心, 北京 100190

2. 中国科学院大学, 北京 100049

摘要 材料基因组计划核心理念, 是通过计算、数据和实验“三位一体”的方式, 变革传统的主要基于经验和实验的“试错法”材料研发模式, 把发现、开发、生产和应用新材料的速度提高到目前的两倍。它旨在建立一个新的以计算模拟和理论预测优先、实验验证在后的新材料研发文化, 从而取代现有的以经验和实验为主的材料研发的模式。本文论述如何通过计算和数据的方法加快新材料研发, 介绍帮助加快新材料发现的高通量集成计算基础平台和软件框架 MatCloud。

关键词 高通量材料计算; 材料基因组计划; 材料信息学; MatCloud

传统以实验为主的“试错法”^[1-3]材料研发模式, 周期长、效率低, 从新材料的最初发现到最终工业化应用一般需要 10~20 年。同时, 由于一些实验很难开展, 也导致了材料的性质至今仍属于未知。

2011 年美国提出了“材料基因组计划”, 其目的就是借助计算手段和数据方法并结合实验, 以三位一体的方式, 帮助企业把发现、开发、生产和应用新材料的速度提高到目前的两倍, 变革传统以实验为主的“试错法”材料研发模式。它强调通过计算与数据的集成, 计算数据与实验数据的集成, 高通量计算与多尺度模拟的集成从而加快材料研发方法和方式的一种材料研发文化的转变。实际上, 材料基因组计划是一个典型的科研信息化理念^[4]在材料科学中的应用与实践。

2014 年 10 月, 中国科学院将清华大学教授、中国科学院院士王崇愚牵头撰写的“实施材料基因组战略研究, 推进我国高端制造业材料发展”咨询报告递交国务院, 提出要建立材料计算(理论)、数据库和实验相融的和材料研发模式。2016 年年初, “材料基因工程关键技术及支撑平台”被列为国家重点研发计划。在随后发布的 2016 年和 2017 年项目指南中, 均强调了对高通量计算和数据的集成化管理关键共性技术的研发。

然而, 材料高通量计算和数据的集成化管理, 需要开展材料计算和数据管理的研究基础设施的支撑^[5], 而开发这些基础设施又涉及材料、物理及计算基多学科交叉, 极具挑战性, 需要对计算和数据方法如何可帮助材料研发有较深入的了解。材料理论设计方式往往包括结构筛选、元素替代、性

能与成分的优化等, 而这些均需要大规模的材料计算, 而大规模计算作业的生成、数据格式的转换、数据的存储、数据处理等又极为容易出错, 从而限制了科研人员通过计算和数据方法帮助开展研究。本文从计算方法和数据方法两个方面, 分析信息化技术如何帮助加快新材料的研发, 介绍高通量材料计算和数据相结合的平台 MatCloud。

1 计算方法如何加快新材料的研发

材料计算模拟被广泛应用在材料性质的预测中, 能够从不同的尺度上对材料进行多尺度模拟计算, 定性定量地描述材料的特征, 从而帮助从多个角度了解材料。1932 年, 物理学家已经得知薛定谔方程能够主导电子系统, 但不能准确地求解该方程。早期的用于第一性原理电子结构计算的方法主要是 Hartree-Fock 理论。但在 Hartree-Fock 理论中, 很重要的一项误差就是电子之间的关联项被全部忽略了。而在量子蒙特卡罗 (quantum Monte Carlo) 方法中, 体系的基态波函数显式地被写成关联的波函数, 也就是说波函数是电子-电子之间距离的显式函数。该方法主要采用量子蒙特卡罗方法对积分进行数值解析, 非常耗时, 但却可能是目前精确度最高的第一性原理方法。1964 年, Kohn 等^[6]提出了密度泛函理论 (density function theory, DFT), 用电子密度描述波函数, 比基于 Thomas-Fermi 模型的密度泛函有了更为坚实的理论基础。1965 年, Kohn 等^[6]发表论文, 描述了求解电子密度和能量的方法, 其核心在于假设了能量和密度之间的泛函关系。他们给这种关系提供了一种简单的近似, 即 Kohn-Sham

收稿日期: 2015-10-22; 修回日期: 2016-11-28

基金项目: 国家自然科学基金面上项目 (61472394); 国家发改委高技术服务业研发与产业化专项 (科发计函字 [2013]8 号)

作者简介: 杨小渝, 研究员, 研究方向为材料信息学, 高通量材料集成计算, 电子信箱: kxy@cnic.cn

引用格式: 杨小渝, 任杰, 王娟, 等. 基于材料基因组计划的计算和数据方法[J]. 科技导报, 2016, 34(24): 62-67; doi 10.3981/j.issn.1000-7857.2016.24.008

方程。1992年Pople^[7]将密度泛函加入了Gaussian程序,提高了精度和计算速度,因此这种方法引起了更多的关注。Kohn和Pople于1998年获得诺贝尔化学奖,自此Kohn-Sham方程在DFT中获得了广泛应用。基于密度函数理论DFT的第一性原理计算在预测材料性质方面具有比较好的精确度,它通过第一性原理计算,求解Kohn-Sham方程,迭代自洽得到体系的电子密度,然后求解体系的基态性质。

分子动力学计算^[8]是从分子尺度上计算物质的热力学性质,有着非常广泛的应用,该方法主要依靠牛顿力学来模拟分子体系的运动,从由分子体系的不同状态构成的系统中抽取样本,计算体系的构型积分,以构型积分的结果为基础进一步计算体系的热力学量和其他宏观性质。2013年,基于量子力学和分子动力学(QM/MM)解决复杂系统的多尺度方法又获诺贝尔奖^[9],其意义在于,计算模拟作为一种工具,将微观尺度上的基本认识转化为宏观尺度上的预测能力。

相图计算方法^[10](calculation of phase diagram, Calphad)利用相关体系的相平衡和热力学性质的综合信息,建立热力学模型,来预测相图和热力学性质。相图以几何图形反映出物质的相平衡规律,表示出物质体系的状态与温度、压力及组成的关系。通过相图可以了解一定体系在不同条件下的相组成情况,被人们称为材料研究的地图。但相图测量的工作量和难度都较大,并且由不同实验室用各种方法所测量的相图还经常存在差异,不易判断其正误。多元相图的测量更加困难,往往通过大量的实验工作,只能得到体系相平衡的一小部分信息。Calphad方法不仅可以计算多元体系的各种相图,还可以进一步得到多种热力学性质,提供相变驱动力等动力学基础数据。目前Calphad方法,已成为辅助材料设计的重要手段。

在以计算模拟为方法加快材料设计方面,取得成功的例子较多,如Ceder等^[11-13]利用高通量第一性原理计算从3万余种化合物中筛选出高效安全的锂电池阴极材料,Curtarolo等^[14]用第一原理计算方法计算了80个二元合金的176个晶体结构的14080个总能,并通过比较计算结果和实验数据,发现176个计算的化合物晶体结构中,除去不稳定结构,有89个化合物的结构是和实验数据一致的,也展示了第一原理计算的方法在预测新材料方面较高的精度。

因此,通过不同尺度的计算可以了解材料的特征,加快材料研发。然而不同尺度计算的集成,以及计算和数据的集成等方面,都面临诸多问题。如开展大规模第一性原理计算,就存在一些制约因素,如计算作业的生成,数据格式的转换,数据的存储,数据处理等极为容易出错,不利于科研人员通过开展大规模计算来帮助材料研究。虽然随着云计算的出现,以及超级计算机的大量使用,计算资源已经不是最主要的问题,但是开展大规模第一性原理计算仍然存在以下4个挑战:1)如何生成大规模的计算作业;2)如何管理这些大规模的计算作业;3)如何管理这些海量的数据;4)如何进行计算和数据的集成管理。

因此,基于“材料基因组计划”的理念,采用材料信息学相关方法和技术,研发高通量材料集成计算关键技术和平台,进行批量的结构建模、大批量计算任务的生成与提交、计算结果的自动处理及采用有效的方法进行数据的自动抽取和规范化加工,存储并管理计算数据,将计算和数据集成为一体,显得愈发重要。

2 数据方法如何加快新材料的研发

通过对已有计算数据和实验数据的挖掘来发现一些模式,基于这些模式进而获得对材料性质的定量或者定性描述,也是一个加快材料研发的重要方法,这种基于数据挖掘的方法可被称为数据驱动的方法(以下又称数据方法)。数据驱动方法的有别于那种有着因果关系和物理内涵的模型方法,它更强调基于大量的数据,寻找数据之间的关联。

2016年5月5日,Raccuglia等^[15]关于机器学习算法改变材料发现方式的文章,被《Nature》以封面论文的形式刊出,其中提出“从失败中学习”:哈佛大学研究者利用机器学习算法,用失败或不成功的实验数据预测了新材料的合成,“我们的机器学习模型获得了比传统人类合成策略更好的效果,并成功预测了有机模板合成的无机物形成条件,成功率达到89%”,这意味着机器学习也能帮助加快材料研发,发明新材料的可能性也大幅提高。这种算法的好处在于,失败的实验数据也能用作下一轮的输入,继而不断完善算法。随着材料科学的发展,数据积累越来越庞大,如何从数以亿万计纷繁复杂的数据中提取有用信息,分析并梳理材料成分-工艺-结构-组织-性能的关系,必然成为材料研究的核心和关键。机器学习方法带领研究人员进入材料空间更深入、更复杂的认知新模式^[16],可以更加科学高效地推动材料设计的发展。

通过数据驱动方法帮助新材料研发已有很多成功案例。1992年,Makishima介绍了利用玻璃材料数据库和知识库开发的一个材料设计专家系统:它能从数据库中选满足一定要求的玻璃组份,再利用知识库中有关玻璃生成规则,预报这些组份生成玻璃的可能性,并利用玻璃组份的原子半径和分解能的数据计算该玻璃材料的物性^[17]。1992年,Yasui和Futagami利用回归处理方法对数据库中的数据进行挖掘,建立起玻璃材料的组份与性能之间的关系,并构造了一个专家系统来预测钙钛矿的生成^[18-19]。1999年,Ashby^[20]提出了“材料选择器的概念”(material selector),即利用已知物质的数据库帮助特定用途材料的选择。

此外,通过数学方程建立结构和材料性质之间的函数关系,即通过定量构效关系(quantitative structure activity relationships, QSAR)也能帮助发现材料的一些性质。Rajan在2011年提到,他的团队没有通过第一性原理计算,而是借用了QSAR的理念,通过数据驱动的方法,识别与已知晶体结构和电子结构相关的离散标量描述符和观察到的材料性能之间的关系,从而找到了决定高温压电钙钛矿关键性能的“结构-性能”描述符^[21]。

高通量材料计算和机器学习的集成,即QM/ML的方法(quantum mechanics/machine learning)更是引起目前业界的普遍关注。QM/ML主要强调通过量子力学计算,产生大量的数据,然后从该数据中学习一些模式,利用该模式来预测材料的性质。2006年,Ceder课题组利用结构之间的关联性发展了数据挖掘结构预测(data mining structure predictor, DMSP)的方法,并成功实现了银镁合金的基态结构预测^[22]。2012年,Saad和Bobbitt等结合计算机与材料化学领域的优势共同研究了数据挖掘在材料结构预测的几种常用算法,利用监督学习方法以平均95%的准确率实现了二元合金的结构预测。此外,该课题组还利用机器学习的统计回归方法基于第一性原理计算结果进行了材料熔点的预测,平均相对误差小于12.8%^[23]。Wolverton研究组于2013年就开始在已经建立的OQMD中进行数据挖掘,发展了巨正则线性规划(grand canonical linear programming GCLP)的机器学习方法,通过组分就可以实现对材料稳定结构的预测^[24],并将该方法成功应用于锂离子电池阳极材料^[25]和镁基三元长周期堆垛有序(long-period stacking ordered LPSO)结构^[26]的预测。陈冠华研究组是开展QM/ML方法较早也是做得比较成功的团队之一,该研究组在2003年便提出了利用第一性原理计算与神经网络(机器学习方法)相耦合的方法提高材料计算的精度,其关键技术在于利用神经网络挖掘实验数据与计算结果之间的定量关系,从而对第一性原理计算结果进行校正^[27]。Bligaard利用含有64000有序金属合金(ordered metallic alloys)的数据库,利用经济学里的帕累托优化方法(Pareto optimal),寻找到了低压缩性,高稳定性,并且低成本的合金优化方法。他们首先利用高通量第一性原理计算,计算了64149种多达4个原子晶胞结构的面心立方和体心立方结构的合金状态方程,建立数据库^[28],然后利用该数据库并结合帕累托优化法进行多目标优化,寻找到满足特定应用需要的优化合金方案。

3 MatCloud:高通量材料计算与数据及资源的一体化管理

3.1 MatCloud 平台简介

通过计算方法和数据方法加快材料的研发,需要良好的信息化技术以及相关信息化平台的支撑,也就是国外常提到的e-Infrastructure或cyber-Infrastructure^[4]。例如,多层平行计算和自动流程的计算方法,能加快计算的效率。有效地统一管理和存储计算材料的结构、物性数据能帮助采用数据挖掘的方法帮助开展相关研究。因此建立计算、储存一体化的集成平台十分必要。

MatCloud是中国科学院计算机网络信息中心材料基因信息技术实验室在国内研发的用于帮助新材料发现的高通量集成计算基础平台和软件框架。它主要基于云计算SaaS和IaaS理念^[29],支持高通量材料集成计算。

高通量材料集成计算(QM/ML或QMDM)目前国内尚无明确的定义。本文认为,高通量材料集成计算,从计算科学的角度,应该具有大规模计算任务的并发式自动流程计算,容错和纠错机制,计算和数据的有效集成。同时具有多用户、多任务、多通道、高并发等计算特点,这些关键字可以用来从计算的角度解读高通量材料计算。而从材料科学的角度,高通量材料集成计算其实质就是探讨如何将组合化学中的“构建单元”和“高通量筛选”理念用于材料计算模拟,通过材料计算寻找、筛选、替代、或优化材料组成的基本构建单元,从而“构建”新的化合物,并计算其相关物性,同时结合材料信息学相关技术将数据、代码、和材料计算软件进行集成,通过数据挖掘尽可能建立起材料组份、结构和性能的定量关系模型用于指导新材料设计。一些用于帮助新材料发现的方法和手段,如结构筛选,元素替代,性能与成分优化等,均涉及到多通道,多任务,高并发的材料计算模拟。

MatCloud其主要功能包括:与高性能计算机实现了有效的对接,能够实现作业的智能调度;能够以课题组为单位进行管理,包括组、组员的用户创建、用户权限以及访问控制的管理等;实现了第一性原理高通量作业的在线生成、提交和作业监控等功能,所有的计算流程都可以自动化实现,并将计算结果按一定的规则存储到数据库中;此外,MatCloud还针对第一性原理计算支持根据结构动态选择参数,包括贻势的推荐功能、根据计算任务和结构产生K点文件及一些主要参数的关联等。

MatCloud更是一个软件框架:提供软件接口和规范,支持插件的研发,针对各课题组高通量计算的实际需求,快速地定制开发“插件”,满足个性化要求。图1^[29]揭示了MatCloud的主要模块和它们之间的有机协同工作关系。主要的核心模块包括:1)高通量材料计算驱动引擎;2)高通量计算操作环境;3)材料计算数据库生态系统;4)材料大数据分析。图2展示了MatCloud的一个用户界面。

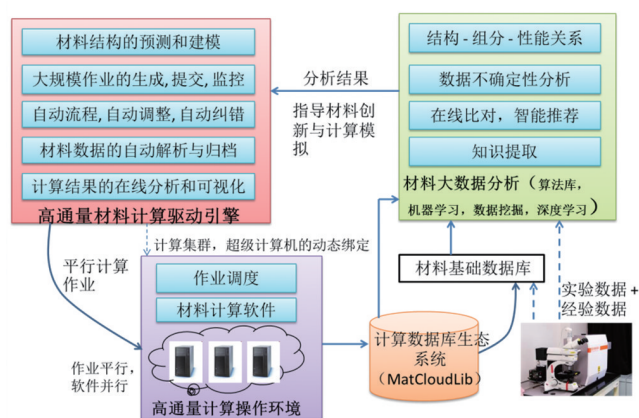


图1 高通量材料集成计算各模块

Fig. 1 Software modules in the integrated high-throughput computational material platform



图2 高通量材料集成计算平台 MatCloud
Fig. 2 MatCloud: an integrated high-throughput computational material platform

3.2 MatCloud 使用案例

热电材料是一种较有前景的能源材料。本文以热电材料 YbZn2Sb2 为例,通过 MatCloud 计算其几何优化、静态计算、能带和态密度的自动流程计算及数据管理等过程,如图3所示。

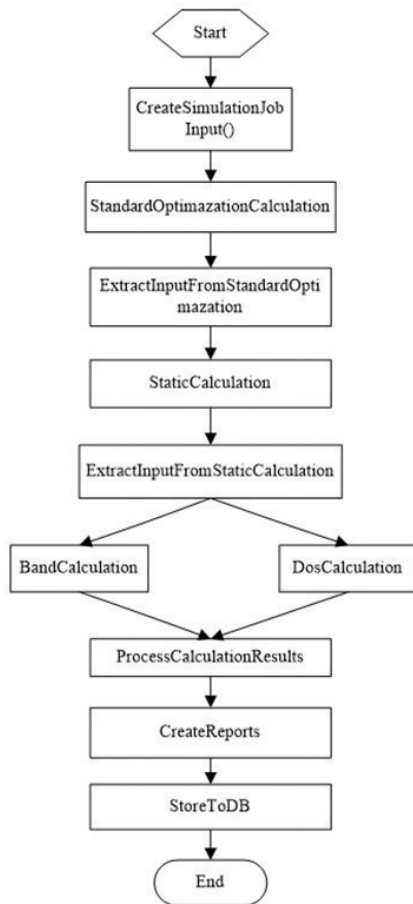


图3 YbZn2Sb2 第一性原理计算流程
Fig. 3 Process of calculating YbZn2Sb2 properties

由图3可知,通过 MatCloud,用户只需一个浏览器登陆指定的 URL(不需安装和下载软件,不需要专门购买计算资源,仅需向 MatCloud 申请一个账号),上传指定 YbZn2Sb2 结构(表1),选择相关计算任务即可。MatCloud 会根据结构和计算任务,自动为每个任务配置计算所需要的最优参数,如截断能, K 点等(表2)。用户也可以通过 Web 页面修改默认参数。一旦确认参数后,点击“提交”,生成的作业通过 MatCloud 被自动提交到与 MatCloud 所连接的计算集群上。一旦计算结束,能带结构(图4)和态密度数据(图5)也将同时生成并被自动保存到数据库中。

表1 晶体结构基本信息

Table 1 Crystal structural information

结构和密度	值
Formula(化学式)	YbZn2Sb2
Space Group(空间群)	p-3m1
Lattice Type(晶系)	Hexagonal
Lattice Parameter(晶胞参数)	$a=4.4451$ $b=4.44510041365$
	$c=7.427$ $\alpha=90.0$ $\beta=90.0$
	$\gamma=119.999996922$
Density(密度)	5.705 g/cm^3

表2 主要参数设置

Table 2 Main calculation parameters

主要参数	参数取值
Energy Cutoff(截断能)	291.78800000 eV
Number of Cores(计算核数)	16
PP-Species(赝势)	PAW_PBE Yb 24Feb2003
	PAW_PBE Zn 06Sep2000
	PAW_PBE Sb 06Sep2000
EDIFF(电子弛豫的收敛精度)	0.0001 eV
K Points(K点)	Structure Optimization(结构优化) 15 15 8
	Properties Calculation(性质计算) 17 17 11

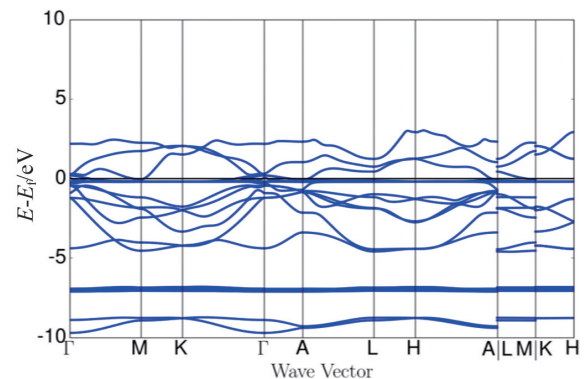


图4 通过 MatCloud 生成的能带图
Fig. 4 Diagram of band structure by MatCloud

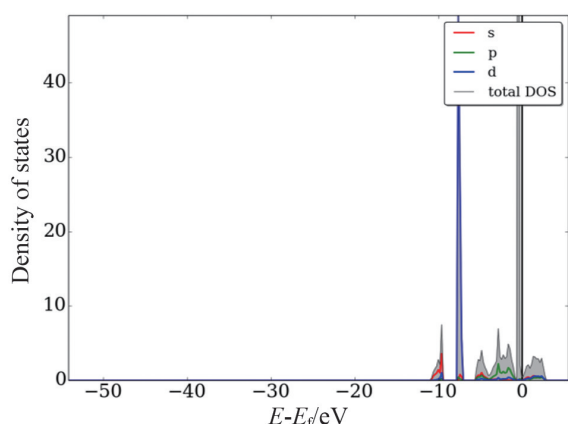


图5 通过MatCloud生成的态密度图

Fig. 5 Diagram of density of states by MatCloud

图4揭示了计算出的YbZn₂Sb₂晶体中电子所具有的能量范围。图4曲线表示亚轨道上电子在空间位置(即倒易空间)的各个能量值。横坐标 Γ 、M、K、A、L、H反映了倒易空间的位置(通过 $\Gamma \rightarrow M \rightarrow K \rightarrow A \rightarrow L \rightarrow H$ 就定义了计算电荷密度的积分路径)。纵坐标表示能量(该图中纵轴能量为特征能量 E 减去费米能 E_f 。费米能表示基态系统电子所占有的能级最高的能量),能量愈大,曲线的位置愈高,一定能量范围内的许多曲线(即能级)彼此相邻很近,这些相邻很近的曲线可以形成价带,导带或禁带,即形成了能带。

图5为计算出的YbZn₂Sb₂晶体的电子态密度图(DOS, Density of States),横坐标为能量(图5中能量为特征能量 E 减去费米能 E_f),纵坐标为单位能量范围内的电子密度。s、p、d为原子核外电子排布的亚轨道。图5中d和p轨道的DOS是很大的尖峰,说明d和p电子相对比较局域,相应的能带也比较窄。

可以看出,MatCloud通过将第一原理计算任务的流程集成,使多个计算任务自动地被提交到与MatCloud集成的计算集群上,辅之以数据的自动提取和管理,极大地降低了第一原理计算的门槛,省去了用户搭建集群或租用集群,以及数据管理等的麻烦,对以做实验为主的材料研发人员尤其有帮助。尤其是MatCloud即将提供的高通量建模和筛选功能,更加方便用户开展大规模的计算和数据管理。将材料高通量集成计算与材料高通量制备和表征相连,更能加快材料研发速度和降低成本,这正是材料基因组计划所追求的材料全链条研发思路。

5 结论

从计算和数据两个方面讨论了材料基因组计划中,材料信息学相关方法和技术将如何加快材料的研发。MatCloud平台旨在解决材料大规模计算面临的计算分散、数据分散等问题,为材料计算模拟提供了一个计算和数据集成的一体化平台。随着大规模科学计算技术的进步、密度泛函理论的发

展、专业软件的完善与开放,材料信息学相关方法和技术将在新材料的研发中起着越来越重要的作用。

致谢 中国科学院固体物理研究所、北京应用物理与计算数学研究所、北京计算科学研究中心、中国科学院上海硅酸盐研究所透明光功能无机材料重点实验室对MatCloud平台研发提供协助。

参考文献(References)

- [1] 田恬. 材料基因组工程——推动材料创新的引擎[J]. 科技导报, 2015, 33(10): 3.
Tian Tian. Material Genome Engineering—An engine to accelerate new material innovation[J]. Science & Technol—材料研发新模式[J]. 科技导报, 2015, 33(10): 13-19.
- [2] Wang Hong, Xiang Yong, Xiagn Xiaodong, et al. Materials genome enables research and development revolution[J]. Science & Technology Review, 2015, 33(10): 13-19.
- [3] Yang X, Wallom D, Waddington S, et al. Cloud computing in e-Science: Research challenges and opportunities[J]. The Journal of Supercomputing, 2014, 70(1): 408-464.
- [4] 杨小渝. 加快材料基因组工程信息化基础设施的建设[J]. 科技导报, 2016, 34(8): 13-14.
Yang Xiaoyu. Accelerating the development of material genome initiative cyberinfrastructure[J]. Science & Technology Review 2016, 34(8): 13-14.
- [5] Hohenberg P, Kohn W. Inhomogeneous electron gas[J]. Physical Review, 1964, 136(3B): B864.
- [6] Kohn W, Sham L J. Self-consistent equations including exchange and correlation effects[J]. Physical Review, 1965, 140(4): 1133-1138.
- [7] Pople J A. Nobel lecture: Quantum chemical models[J]. Reviews of Modern Physics, 1999, 71(5): 1267.
- [8] Allen M P. Introduction to molecular dynamics simulation[J]. Computational soft matter: from synthetic polymers to proteins, 2004, 23: 1-28.
- [9] Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme[J]. Journal of Molecular Biology, 1976, 103(2): 227-249.
- [10] Kaufman L, Bernstein H. Computer calculation of phase diagrams[M]. New York: Academic Press, 1970.
- [11] Kang B, Ceder G. Battery materials for ultrafast charging and discharging[J]. Nature, 2009, 458(7235): 190-193.
- [12] Hautier G, Jain A, Ong S P, et al. Phosphates as lithium-ion battery cathodes: An evaluation based on high-throughput ab initio calculation [J]. Chemistry of Materials. 2011, 23(15): 3495-3508.
- [13] Kang K, Meng Y S, Bréger J, et al. Electrodes with high power and high capacity for rechargeable lithium batteries[J]. Science, 2006, 311(5763): 977-980.
- [14] Curtarolo S, Morgan D, Ceder G. Accuracy of ab initio methods in predicting the crystal structures of metals: A review of 80 binary alloys [J]. Calphad, 2005, 29(3): 163-211.
- [15] Raccuglia C, Elbert K, Adler P, et al. Machine-learning-assisted materials discovery using failed experiments [J]. Nature, 2016, 533: 73-76.
- [16] Kristian S, Karsten W. Making the most of materials computations [J]. Science, 2016, 354(6309): 180-181.
- [17] Makishima A, Uo M, Inoue H. Improved expert system for materials

- design of glasses[C]// Proceedings of the Second International Conference and Exhibition on Computer Applications to Materials and Molecular Science and Engineering, Yokohama, Japan: Computer Aided Innovation of New Materials II, 1992: 22–25.
- [18] Yasui I, Utsuno F. Material design of glasses based on database-INTERGLAD[C]// Proceedings of the Second International Conference and Exhibition on Computer Applications to Materials and Molecular Science and Engineering, Yokohama, Japan: Computer Aided Innovation of New Materials, 1993: 1539–1544.
- [19] Futagami T, Makishima A, Yasui I, et al. Expert system for materials Design of PTC Thermistors [C]//Proceedings of the Second International Conference and Exhibition on Computer Applications to Materials and Molecular Science and Engineering, Yokohama, Japan: Computer Aided Innovation of New Materials II, 1993: 1565–1568.
- [20] Ashby M F. Materials selection in mechanical design[M]. 2nd, Oxford: Butterworth-Heinemann, 1999.
- [21] Balachandran P V, Broderick S R, Rajan K. Identifying the ‘inorganic gene’ for high-temperature piezoelectric perovskites through statistical learning[J]. Proceedings of Royal Society A: Mathematical, Physics, and Engineering Science, 2011, 467(2132): 2271–2290.
- [22] Fischer C C, Tibbetts K J, Morgan D, et al. Predicting crystal structure by merging data mining with quantum mechanics[J]. Nature Materials, 2006, 5(8): 641–646.
- [23] Saad Y, Gao D, Ngo T, et al. Data mining for materials: Computational experiments with AB compounds[J]. Physical Review B, 2012, 85(10): 104104.
- [24] 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)[J]. Journal of the Minerals, Metals & Materials Society, 2013, 65(11): 1501–1509.
- [25] Kirklin S, Meredig B, Wolverton C. High-throughput computational screening of new Li-Ion battery anode materials[J]. Advanced Energy Materials, 2013, 3(2): 252–262.
- [26] Saal J E, Wolverton C. Thermodynamic stability of Mg-based ternary long-period stacking ordered structures [J]. Acta Materialia, 2014, 68: 325–338.
- [27] Hu L H, Wang X J, Wong L H, et al. Combined first-principles calculation and neural-network correction approach for heat of formation[J]. Journal of Chemical Physics, 2003, 119(22): 11501–11507 .
- [28] Bligaard T, Johannesson G H, Ruban A V, et al. Pareto-optimal alloys [J]. Applied Physics Letters, 2003, 83(22): 4527–4529.
- [29] 王卓, 杨小渝, 郑宇飞, 等. 材料基因组框架下的材料集成设计及信息平台初探[J]. 科学通报, 2013, 58(35): 3733–3742.
Wang Zhuo, Yang Xiaoyu, Zheng Yufei, et al. Integrated materials design and informatics platform within the materials genome framework [J]. Chinese Science Bulletin, 2013, 58(35): 3733–3742.

Computational and data management based on Material Genome Initiative

YANG Xiaoyu^{1,2}, REN Jie^{1,2}, WANG Juan^{1,2}, Zhao Xushan¹, WANG Zongguo¹, SONG Jianlong¹

1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The core philosophy of Material Genome Initiative is the transition of the way of new material design from traditional “try-and-error” approach to the in-silico material design approach which employs intensive computing and material informatics. It aims to speed up discovery, development, production and deployment of new material two times faster than it is now. It means a culture shift of the material discovery, development and deployment: simulation and prediction first, followed by the experiment. This paper depicts how computational approach and informatics can discover new materials. A high throughput computational material platform and software framework, namely, MatCloud, is discussed.

Keywords high-throughput computational material; Material Genome Initiative; material informatics; MatCloud

(责任编辑 田恬)