



开源密度泛函理论软件ABACUS: 助力大原子模型发展

彭星亮¹, 周巍青¹, 郑大也¹, 金祖鑫¹, 黄一珂¹, 路登辉², 李卓远^{3,4}, 文通其^{3,4}, 陈默涵^{1,2*}

1. 北京科学智能研究院, 北京 100080

2. 北京大学应用物理与技术研究中心, 北京 100871

3. 香港大学机械工程系结构材料中心, 香港 999077

4. 香港大学生命科学与能源材料创新研究所, 深圳 518000

* 联系人, E-mail: mohanchen@pku.edu.cn

2024-11-20收稿, 2025-03-04修回, 2025-03-04接受, 2025-03-05网络版发表

国家自然科学基金(12122401, 12074007, 12135002)资助

摘要 基于机器学习的原子间势函数模型可以兼具量子力学的精度和传统经验力场的效率, 近年来已在分子动力学等领域得到了广泛的应用. 随着机器学习势函数方法在应用领域的拓展和相关数据的持续积累, 正如语言大模型的出现一样, 构建可覆盖元素周期表上大部分元素的通用大原子模型(large atomic model, LAM)也成为可能, 并且正逐渐发展成为理解和预测材料性质的全新原子尺度模拟工具. 开放大原子模型项目(The OpenLAM project)旨在构建适用于复杂材料体系的高精度和高效率预训练大模型, 并且通过进一步的下游微调和知识蒸馏来获得高效准确的原子间势函数模型, 而其中的关键步骤是需要第一性原理软件计算出海量的训练数据. 本文介绍开源密度泛函理论软件原子算筹ABACUS对OpenLAM项目的支撑. 该软件是一款国内自主研发的第一性原理材料计算软件, 支持平面波和数值原子轨道双基组, 构建了方便用户使用的赝势和数值原子轨道库, 已在不同硬件平台上成功移植, 为OpenLAM项目提供了大量密度泛函理论计算数据, 应用领域覆盖了合金、半导体、钙钛矿等多种材料体系. 最后, 本文展望了ABACUS在AI for Science时代下的发展.

关键词 密度泛函理论, 原子算筹, 机器学习势, 大原子模型

基于量子力学的第一性原理计算模拟已在材料科学、凝聚态物理、量子化学等领域获得了大量的应用. 在众多第一性原理计算方法中, 密度泛函理论(density functional theory, DFT)^[1,2]因其计算效率和精度的较好平衡获得了广泛使用. DFT理论基于Hohenberg-Kohn定理^[1], 即系统的基态性质可以通过电子密度来唯一确定, 而求解Kohn-Sham方程可以获得电子密度^[2]. 借助高性能计算机和高效求解Kohn-Sham方程的DFT软件, 研究人员能够在合理的计算资源下模拟复杂的物理和化学过程, 已经成为研究和发现新材料、理解复杂化

学过程的重要工具^[3~13]. 因为DFT软件种类较多、特点也不同, 研究人员可以根据不同需求选择不同的软件. 一方面, 使用赝势是常用的DFT材料计算方法, 对于周期性边界条件下的计算, 基于平面波基组的软件表现较好^[5,11,13], 而基于局域轨道基组的软件则可有效减少大体系模拟的计算量, 更适合处理非周期性体系^[7,12]. 另一方面, 全电子计算软件则可以给出较为精确的结果, 可以作为赝势结果的标定^[3]. 目前, 用户数量较多且软件生态较完整的主流DFT软件多起源于国外. 然而, 随着科学研究的不断深入以及AI for Science领域

引用格式: 彭星亮, 周巍青, 郑大也, 等. 开源密度泛函理论软件ABACUS: 助力大原子模型发展. 科学通报, 2025, 70: 4027–4043
Peng X, Zhou W, Zheng D, et al. Open-source density functional theory software ABACUS: promoting the development of large atomic-scale machine learning pre-trained models (in Chinese). Chin Sci Bull, 2025, 70: 4027–4043, doi: [10.1360/TB-2024-1243](https://doi.org/10.1360/TB-2024-1243)

的兴起,对DFT软件的易用性、稳定性、高性能优化水平、代码架构可扩展性、软件更新和维护的能力等方面都提出了更大的需求和挑战。

传统的Kohn-Sham密度泛函理论方法很难处理数百原子或更大尺度的复杂材料体系。虽然经典或半经典方法可有效降低计算量从而处理大体系计算,但其计算精度在一些复杂体系中仍有欠缺。这样一个精度和效率的两难抉择长期阻碍着第一性原理计算在大体系和复杂体系中的实际应用。近年来,深度学习与物理的深度交叉融合为有效解决该难题提供了新的方法,其中具有代表性的工作之一是基于深度学习的原子间势函数模型构建方法DeePMD(deep potential molecular dynamics)^[14]。DeePMD模型基于第一性原理数据训练而成,能以第一性原理的精度预测原子的能量、受力和晶胞所受应力等性质,从而很好地描述原子体系的势能面(potential energy surface, PES)。特别是这种基于深度神经网络的势函数在推理成本上低于传统第一性原理方法数个量级,使其成为大体系长时间模拟的理想工具。然而,基于机器学习的势函数模型仍需第一性原理软件产生大量数据,并且往往是针对单一应用场景生成特定的机器学习势函数,势函数的生产周期较长且成本较高。

随着机器学习势函数在特定应用场景的不断积累,产生的大量第一性原理数据包含有材料领域的公共知识,可以被统一的深度学习框架所学习,从而生成可面向元素周期表的原子间势函数大模型。近两年来,国内外已经出现相关模型。例如,2022年,来自英国剑桥大学的Batatia等人^[15]基于消息传递图神经网络(message passing neural network, MPNN)框架提出了MACE模型,通过使用更高的四体相互作用来减少消息传递的迭代次数,从而获得更快和更准的模型推理能力;随后,他们又提出了MACE-MP-0^[16]模型,该模型对于多种场景展现出了一定的泛化能力。2022年,来自美国加州大学圣地亚哥分校的Chen等人^[17]提出了基于三体作用图神经网络的M3GNet模型,采用了187687个构型训练并可适用于89种元素组成的材料体系。2024年, Qi等人^[18]改进了材料空间的采样,开发出了M3GNet-DIRECT模型,该模型在可移植性方面具有更好的表现。2023年,来自美国国家标准与技术研究院的Choudhary等人^[19]基于ALIGNN^[20]框架,使

用包含约75000种材料和400万帧数据进行训练,提出了ALIGNN-FF模型,可以实现对任意89种元素组合成的物质进行推理。2023年,来自美国加州大学伯克利分校的Deng等人^[21]开发的CHGNet模型也是基于图神经网络的普适机器学习势函数模型,该模型在训练时加入了原子磁矩信息,使得模型可同时描述原子和电子信息,适用于具有可变价态的过渡金属元素体系。2023年, DeepMind公司开发的GNoMe模型^[22]发现了220万种理论上稳定的新晶体材料,其中38.1万种为预测出来的新材料。值得一提的是,尽管上述模型覆盖的材料范围都很广,但实际使用时的精度还远未达到第一性原理的精度。因此,在未来的一段时间内,该领域在数据、模型、场景应用上都还将迎来持续的发展。

尽管上述工作已经展开了对原子间势函数大模型的探索,然而这一领域的工作还存在着一些本质的困难和挑战。首先,原子间势函数大模型的训练需要大量材料第一性原理数据,目前在该领域有许多不同的常用软件,即便是同一种材料,基于不同参数得到的数据也难以对齐,数据质量的不齐可能成为大模型发展的障碍。其次,除了商业大公司,目前的工作多采用公开DFT数据集,当模型精度和适用性在未来进一步推向落地实用时,需要规模更加庞大的数据集,甚至是源源不断产生的新数据,这就需要对底层产生大量材料第一性原理数据的基础设施,即密度泛函理论或多体量子化学软件有精度和效率上的革新。最后,原子间势函数大模型的功能还可进一步完善,目前研究人员主要聚焦的对象还是材料体系的能量、受力、应力等基本性质,将来大模型还可朝着预测更多材料物理性质不断发展,甚至可以纳入实验数据进行训练,因此底层基础设施的建设也亟待不断完善。

国内科学计算软件开源社区DeepModeling联合多家合作单位,针对以上问题发布了开放大原子模型项目(OpenLAM: <https://www.aissquare.com/openlam>)^[23]。OpenLAM项目建立在DPA等系列模型的基础上,希望通过建立围绕DPA模型的开源生态系统,为微观尺度的科学研究提供新的基础设施,并推动材料、能源、生物制药等领域微尺度工业设计的变革。例如, Zhang等人联合多家协作单位,从2022年起相继提出了DPA-1^[24]和DPA-2^[25]的原子间势函数大模型。DPA-1模型是一种基于注意力机制的深度学习势能模型,它的

训练数据除了包含公开的OC20数据集,还包含如铝镁铜合金、高熵合金和固态电解质等,可以适用于56种元素体系,通过预训练结合微调的方式来适用实际的应用场景. DPA-2^[25]在DPA-1的基础上持续演进,采用了多任务的训练策略,能够同时优化不同DFT参数来源的数据集,有效降低了模型对数据质量的要求,使得不同软件得到的数据可以同时训练. 同时, DPA-2模型的设计使得在预训练阶段模型就能学习到化学组分和构型等知识,从而使得在特定的下游任务中需要的微调数据量大大减少. 尽管DPA-2模型较为庞大,但通过模型蒸馏的功能,可以得到效率大幅提升的推理模型,从而可应用于实际研究工作中. 此外,原子尺度大模型发展离不开大量第一性原理计算数据作为支撑,原子算筹(atomic-orbital based *ab-initio* computation at USTC, ABACUS)是一款开源密度泛函理论软件,同时支持周期性边界条件下的平面波和数值原子轨道两种基组,并实现了包含实时演化含时密度泛函理论方法(rt-TDDFT)^[26]、杂化泛函计算(hybrid functional)^[27]、随机波函数密度泛函理论(stochastic-DFT)^[28]、无轨道密度泛函理论方法(OFDFT)^[29]等多种算法,适应不同类型材料体系的计算需求. 特别是,借助DeePKS^[30,31]等AI辅助算法,电子结构软件有望在数据生产的效率上进行革新. 此外, ABACUS已支持异构计算并且适配了多款国产硬件,目前正在国产超算平台上持续为DPA模型产生第一性原理数据.

1 原子算筹ABACUS软件

1.1 ABACUS软件介绍

ABACUS, 中文名原子算筹, 最早由中国科学技术大学的何力新教授团队开发, 目前该软件与DeepModeling社区合作, 采用开源模式发展, 鼓励国内外科研人员参与软件的开发, 在过去几年内极大促进了软件的快速迭代和功能丰富. 此外, ABACUS的用户社区在逐渐扩大, 社区也提供了相关的教程和文档支持.

为方便读者直观地了解不同密度泛函软件的异同, 表1列出了可处理周期性边界条件的多款密度泛函理论软件, 并进行了一些简单的比较. 大多数密度泛函软件仅支持一种基组, 而ABACUS同时支持平面

波和数值原子轨道两种基组, 且两种基组的计算输入文件统一, 无需太多的额外学习成本. 用户可根据研究体系的尺寸和所需精度灵活选择不同基组进行计算. 此外, 同类软件大多只支持一种密度泛函框架, 最为常见的是Kohn-Sham密度泛函理论. ABACUS除了支持Kohn-Sham密度泛函理论之外, 还支持多种密度泛函理论, 如适用于研究温稠密物质的随机波函数密度泛函理论, 适用于大尺寸材料体系的无轨道密度泛函理论, 以及适用于激发态模拟的含时密度泛函理论. 以上方法极大扩展了ABACUS可适用的场景.

在功能层面, ABACUS作为一款国产的材料第一性原理软件, 近年来实现了与多种原子尺度深度学习算法的结合, 支持包括深度势能分子动力学方法(DeePMD)^[39,40]、基于主动学习的构型采样软件DP-GEN^[41]、机器学习哈密顿量预测方法(DeepH)^[42]、机器学习泛函修正算法(DeePKS)^[30,31]、机器学习紧束缚近似模型方法(DeePTB)^[43]等前沿算法. 这些新算法显著改善了传统密度泛函理论在计算精度、计算体系尺寸和复杂程度以及计算效率方面的瓶颈问题, 推动电子和原子尺度的模拟算法发挥了更大的作用. 此外, ABACUS还提供了多个常用材料模拟软件的接口, 方便用户对材料性质进行更广泛的模拟. 例如, 非绝热分子动力学软件Hefei-NAMD^[44]、支持声子计算的Phonopy^[45]和ShengBTE^[46]、最大局域化Wannier函数计算的Wannier90^[47]、用于原子尺度模拟的ASE^[48]、基于紧束缚方法计算电子结构性质的PYATB^[49]、计算交换耦合和磁各向异性的TB2J^[50]等软件.

在高性能计算发展方面, ABACUS软件针对多核处理器和异构计算机(如曙光DCU等国产超级计算机硬件平台等)进行了代码框架的重新设计、算法的移植以及深度的效率优化. 例如, 该软件成功地移植到曙光DCU硬件平台之后, 已为OpenLAM项目提供了大量第一性原理数据. 接下来, 我们先简要介绍ABACUS的基础算法, 之后介绍该软件在OpenLAM项目中的应用, 包括半导体和合金等材料体系.

1.2 平面波基组

ABACUS支持使用平面波(plane wave, PW)或数值原子轨道(numerical atomic orbital, NAO)基组对电子波

函数进行展开. 当使用PW基组时, 自洽场(self consistent field, SCF)方法求解Kohn-Sham方程的流程参见图1(a). 在密度泛函理论面向周期性体系的具体实现中, 使用平面波基组展开是一种常见方式, 记为

$$\psi_{nk}(\mathbf{r}) = \sum_{\mathbf{G}} c_{nk}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \quad (1)$$

其中, $\mathbf{G}$ 和 $\mathbf{k}$ 分别表示平面波的波矢和布里渊区里的采样点, n 表示能带序号, $c_{nk}(\mathbf{G})$ 为平面波的展开系数. 基于上式, 密度泛函理论中的核心Kohn-Sham方程可

写为

$$\sum_{\mathbf{G}'} \left[\frac{1}{2} (\mathbf{k}+\mathbf{G})^2 \delta_{\mathbf{G}\mathbf{G}'} + V_{\text{KS}}(\mathbf{G}-\mathbf{G}') \right] c_{nk}(\mathbf{G}') = \varepsilon_{nk} c_{nk}(\mathbf{G}), \quad (2)$$

这里, ε_{nk} 为本征能量, $V_{\text{KS}}(\mathbf{G}-\mathbf{G}')$ 是平面波基组下的Kohn-Sham有效势, 包含描述电子和电子相互作用的Hartree势、交换关联项势, 以及描述电子和离子相互作用的外势项(例如赝势)等.

平面波基组在密度泛函理论计算中具有较多优势. 首先, 平面波之间相互正交, 用户可以简单通过能量截

表 1 可处理周期性边界条件的密度泛函理论软件对比

Table 1 Comparison of several density functional theory packages that can handle periodic boundary conditions

密度泛函理论软件	支持的基组	离子势的处理	是否开源
ABACUS	平面波、数值原子轨道	赝势	开源
VASP ^[32]	平面波	PAW/赝势	商业
Quantum ESPRESSO ^[33]	平面波	PAW/赝势	开源
ABINIT ^[5]	平面波	PAW/赝势	开源
CASTEP ^[34]	平面波	赝势	开源
CP2K ^[35]	平面波、高斯基组	赝势	开源
OpenMX ^[36]	数值原子轨道	赝势	开源
SIESTA ^[37]	数值原子轨道	赝势	开源
FHI-aims ^[38]	数值原子轨道	全电子	商业
WIEN2k ^[3]	线性缀加平面波+局域轨道	全电子	商业

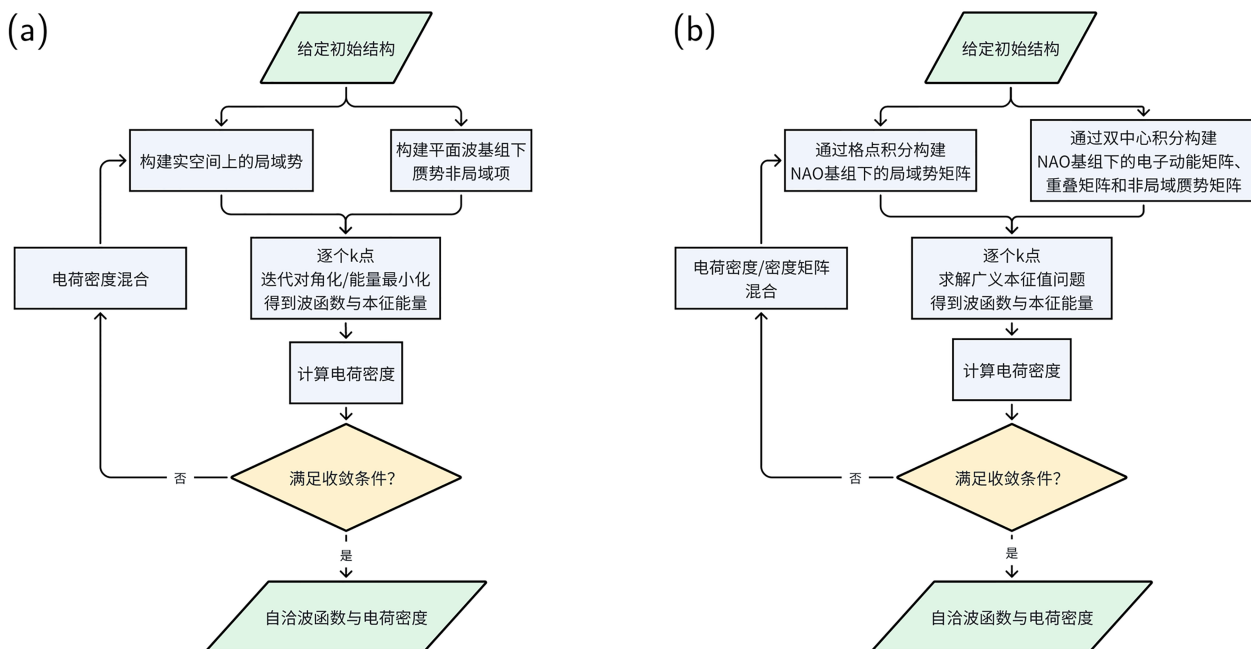


图 1 ABACUS在使用平面波基组(a)和数值原子轨道基组(b)时自洽场(SCF)迭代方法的流程图

Figure 1 Self consistent field method implemented in ABACUS with the usage of plane wave basis set (a) and numerical atomic orbital basis set (b)

断值参数来系统增加基矢量的个数,从而增加计算精度.其次,平面波有良好的解析性质,例如基于动能的表达式较为简单且实现容易,而有效势的计算可以借助快速傅里叶变换(fast-Fourier-transform, FFT)来实现相关物理量在实空间和倒空间的变换.最后,平面波由于数量较大,往往采用迭代法来求解Kohn-Sham方程的特征值,其中平面波的大量使用适合采用异构设备进行高性能加速.

ABACUS在平面波基组下支持使用共轭梯度法(conjugate gradient, CG)或Davidson等迭代对角化方法来求解Kohn-Sham方程的特征值和特征向量,这也是软件在大多数材料体系中计算耗时最多的部分.传统对角化方法的耗时随着体系的增长呈 $O(N^3)$ (N 是系统哈密顿矩阵的维度)快速增加,制约了实际可计算体系的大小,因此可以采用异构加速的方法来加快计算.一般而言,平面波基组适合100个原子以下的密度泛函理论计算.从大原子模型的角度出发,100个原子已经可以具有非常丰富的元素种类选择和空间排布方式,满足这个尺寸大小的周期性材料体系是当下大模型训练数据的主要来源之一.

在求解Kohn-Sham方程的特征值和特征向量的过程中,由于平面波基组的数量过大,通常不显式地构建哈密顿量矩阵而是计算哈密顿算符作用于电子波函数来求得迭代法求解特征值方法中的梯度向量.通常来讲,哈密顿量矩阵乘上波函数这个操作占据Kohn-Sham方程求解的主要时间,其性能热点主要包括:第一,在计算赝势中的局域势对波函数作用时,需执行傅里叶变换(FFT)操作,其时间复杂度为 $O(N\log N)$, N 为平面波数量;第二,在计算赝势中Kleinman-Bylander (KB)非局域项时,需计算大量非局域势投影和平面波的内积.在对平面波基组下求解Kohn-Sham方程的任务进行高性能优化方面,ABACUS支持MPI和OpenMP两种并行方案,物理算法上采取对布里渊区 $\mathbf{k}$ 点、平面波基组和实空间格点进行任务的并行分配和运算.

1.3 数值原子轨道基组

第一性原理计算中基组的选择对计算的精度与效率有直接影响.在使用传统的平面波或实空间格点基组时,基组对精度的影响由单一参数(平面波动能截断

或实空间格点密度)系统地控制,计算结果随该参数增大而呈现的收敛体现了基组趋于完备.尽管这类基组能够系统性地降低由基组的不完备性所带来的精度误差,它们对实际体系的表示效率并不高.以晶体硅为例,一个能量收敛到毫电子伏量级的计算一般需要数万个平面波;在计算几十个原子或更大的体系时,所需的平面波数量往往以百万计,这对计算资源提出了较高的需求.

ABACUS软件还支持基于数值原子轨道的Kohn-Sham方程求解.数值原子轨道能在一定精度下更高效地表示电子波函数,从而大幅提升计算效率.原子轨道基函数通常具有如下形式

$$\phi_{lm\zeta}(\mathbf{r}) = f_{l\zeta}(r)Y_{lm}(\hat{r}), \quad (3)$$

其中, l, m, ζ 分别表示角量子数、磁量子数和径向轨道标号, $f_{l\zeta}$ 表示径向轨道,而角向部分 Y_{lm} 在实践中通常取实球谐函数.实际计算时,对一个角动量 l 可引入多个径向函数以扩大基函数所张成的空间,这些径向函数由 ζ 区分.在周期性体系的原子轨道线性组合(linear combination of atomic orbitals, LCAO)方法中,电子波函数由各原子轨道所对应的布洛赫函数线性叠加而成,具有如下形式

$$\phi_{\mu\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} \mathbf{e}^{i\mathbf{k}\cdot\mathbf{R}} \phi_{\mu}(\mathbf{r} - \mathbf{R}), \quad (4)$$

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mu} c_{\mu n, \mathbf{k}} \phi_{\mu\mathbf{k}}(\mathbf{r}), \quad (5)$$

这里, μ 为元胞内轨道的复合指标(包含元素、原子以及 $lm\zeta$), $\mathbf{R}$ 为布拉维晶格矢量, $\phi_{\mu\mathbf{k}}$ 为原子轨道 ϕ_{μ} 对应的布洛赫函数, c 为展开系数.由于原子轨道基组不具正交性,某个 $\mathbf{k}$ 点下的Kohn-Sham方程对应到广义本征值问题时具有如下公式

$$\sum_{\nu} H_{\mu\nu, \mathbf{k}} \mathbf{c}_{\nu n, \mathbf{k}} = \varepsilon_{n\mathbf{k}} \sum_{\nu} S_{\mu\nu, \mathbf{k}} \mathbf{c}_{\nu n, \mathbf{k}}, \quad (6)$$

其中, $H_{\mu\nu, \mathbf{k}} = \langle \phi_{\mu\mathbf{k}} | \hat{H} | \phi_{\nu\mathbf{k}} \rangle$ 为Kohn-Sham哈密顿量的矩阵元, $S_{\mu\nu, \mathbf{k}} = \langle \phi_{\mu\mathbf{k}} | \phi_{\nu\mathbf{k}} \rangle$ 为交叠矩阵元.上式的求解同样以SCF的方式进行,具体流程参见图1(b).当体系较大时,使用原子轨道基组可以带来基组交叠矩阵与哈密顿量矩阵的稀疏性,使用针对稀疏矩阵求解的线性标度算法^[37,51]可以进一步提升计算效率.因此,在模拟几百乃至更多原子的大体系中,原子轨道线性组合方法相比于平面波基组更具效率优势.

在数值原子轨道基组的计算中,其性能热点主要包含:第一,通过格点积分技术构建局域势的哈密顿量矩阵和电荷密度,以及原子求力和应力的相关操作;第二,进行哈密顿量矩阵的对角化计算,对于基组数为 N 的时间复杂度为 $O(N^3)$. ABACUS实现了在数值原子轨道基组计算中采用MPI和OpenMP并行. 其中,MPI的并行主要包括:使用2D块循环算法对矩阵进行切分的并行,以及进行格点积分运算时对实空间格点的并行划分.

高质量的原子轨道基组是该领域长期追求的目标,对基组的研究与开发贯穿了计算化学与计算材料科学的发展,从早期受类氢原子波函数启发的斯莱特型轨道到数值计算上更具优势的高斯型轨道,原子轨道基组已在众多领域得到广泛应用. 近20多年来,数值原子轨道基组逐渐发展为一个具有竞争力的选项. 与传统基函数不同的是,数值原子轨道的径向函数 $f(r)$ 完全由 $(r_i, f(r_i))$,即一系列格点上的值所表示;矩阵元计算亦需通过数值算法实现. 相较于传统基组,数值原子轨道拥有更大的自由度,但矩阵元计算更为复杂. 随着相应数值算法的成熟^[37,51~53],这类基组相较于传统基组在矩阵元计算上的复杂度得以克服,其高自由度带来了良好的可迁移性,精度与效率的兼顾,可定制性等一系列优点.

数值原子轨道的生成策略不一. 一类常见的方法是利用禁势来产生单个原子的波函数^[37,54~57]. 基于禁势的方法,Blum等人^[51]以能量为判据,从一批候选中逐个选出最优目标能量的函数,以此来生成数值原子轨道. Ozaki等人^[58,59]则提出禁势解出的波函数可作为数值原子轨道的“基函数”,数值原子轨道在这些基函数下的展开系数可一并在自洽场迭代中优化.

除了上述与能量相关的策略,另一类流行的方法是利用(平面波或实空间格点等)高精度基组生成的参考态来生成轨道. 例如, Sanchez-Portal等人^[60,61]提出最小化“溢出函数(Spillage Function)”,如下式所定义:

$$S = \frac{1}{N} \sum_{n=1}^N \sum_{\mathbf{k}} \langle \psi_{n\mathbf{k}} | (1 - P_{\mathbf{k}}) | \psi_{n\mathbf{k}} \rangle$$

$$= \frac{1}{N} \sum_{n=1}^N \sum_{\mathbf{k}} \| (1 - P_{\mathbf{k}}) | \psi_{n\mathbf{k}} \rangle \|^2, \quad (7)$$

其中, N 为参考的能带数, $\{\psi_{n\mathbf{k}}\}$ 为高精度平面波计算得

到的参考电子波函数, $P_{\mathbf{k}}$ 是到晶格动量为 $\mathbf{k}$ 的原子轨道布洛赫函数所张成空间的投影算符:

$$P_{\mathbf{k}} = \sum_{\mu\nu} |\phi_{\mu\mathbf{k}}\rangle S_{\mu\nu,\mathbf{k}}^{-1} \langle \phi_{\nu\mathbf{k}}|. \quad (8)$$

在此方案中,参考系统为给定晶体,而原子轨道则由一系列赝波函数(即禁势内孤立原子的本征态)或者斯莱特型轨道叠加而成.

基于上述溢出函数的框架, ABACUS开发团队的Chen等人^[62]提出原子轨道可取为球面波叠加而成^[63], 参考系统则可取一系列键长不等的二聚体(dimer)与三聚体(trimer)以得到更好的可迁移性. 不久前, Lin等人^[64]提出,对上述目标函数添加梯度项

$$S' = S + \frac{1}{N} \sum_{n=1}^N \sum_{\mathbf{k}} \|\nabla(1 - P_{\mathbf{k}}) | \psi_{n\mathbf{k}} \rangle\|^2, \quad (9)$$

可进一步提升轨道质量. 这两种轨道均经过充分测试,在一系列体系中表现良好. 图2展示了轨道的径向函数以及基组对能量的影响. 上述方法亦可用来生成价层分裂(split valence/multiple-zeta)与具有极化(polarization)轨道的基组;同SG15赝势^[65]配套的double-zeta with polarization (DZP)与triple-zeta with double-polarization (TZDP)轨道可在ABACUS官网(<https://abacus.ustc.edu.cn/pseudo/list.htm>)下载获得.

1.4 赝势-数值原子轨道库

对于第一性原理计算,在尽可能保持精度的条件下降低计算成本是该领域长期关注的方法发展方向. 对于大多数常见元素,人们通常所关心的材料性质,例如物态相图等热力学性质、能隙等光电性质、弹性模量等力学性质基本由原子的外层电子或价电子决定. 因此,对于体系的哈密顿量中描述电子-离子相互作用的外势项,以赝势^[13,65~72]或类似的近似方法,如投影缀加平面波方法^[13](projector augmented-wave, PAW)取代,去除掉对一些内壳层电子的描述,只保留关键的价电子部分,从而显著降低计算成本,是常见的一种电子结构计算方法. 目前被广泛使用的赝势方法主要有模守恒赝势(norm-conserving pseudopotentials, NCPP)^[67]方法,例如SG15^[65]、PseudoDOJO^[68]、GTH^[73]等系列赝势. 此外,还有所需能量截断值较低的超软赝势^[70](ultra-soft pseudopotential, USPP)方

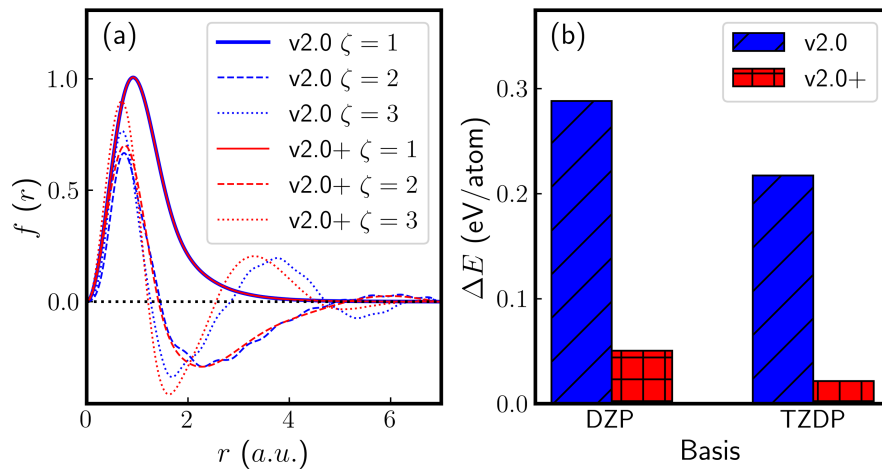


图 2 轨道的径向函数和基组对能量的影响. (a) 铜(In)元素的TZDP数值原子轨道基组中的d轨道径向函数. 蓝线来自基于Lin等人^[64]的工作、可由官网获取的轨道(v2.0), 红线来自在其基础上进一步优化的轨道(此处记为v2.0+). (b) 采用数值原子轨道基组与平面波基组计算晶体铜(ICSD: 53777)所得的能量差. 其中DZP包含2s2p2d1f轨道而TZDP包含3s3p3d2f轨道, 更大的基组可使数值原子轨道基组的计算趋于平面波结果

Figure 2 The radial functions of orbitals and the effect of basis on energy. (a) The radial functions of the d-orbitals in the TZDP numerical atomic orbital basis set for indium (In). The blue line is derived from the orbitals (v2.0) based on the work of Lin et al.^[64], which can be obtained from the official website, and the red line is derived from the orbitals further optimized on this basis (hereafter referred to as v2.0+). (b) The energy difference obtained by calculating crystalline indium (ICSD: 53777) using numerical atomic orbital basis sets and plane wave basis sets. The DZP includes 2s2p2d1f orbitals, while the TZDP includes 3s3p3d2f orbitals; a larger basis set can make the calculations using numerical atomic orbital basis sets approach the results of plane wave calculations

法, 例如PSLibrary^[69]、GBRV^[74]等系列赝势. PAW方法^[13](如VASP^[13]、JTH^[75]等)和超软赝势方法具有一定相似性, 但实现较为复杂. 通常PAW方法所具有的优点是需要较低的能量截断值且具备还原出全电子波函数的方法. 此外, 赝势还和对应的交换关联泛函赝势有关. 因此, 从大原子模型的角度出发, 训练数据如果包含了来自不同赝势的第一性原理数据, 会使得模型训练容易遇到数据来源和质量不统一的问题, 而DPA框架则考虑了不同训练数据来源的兼容性问题.

赝势的生成参数直接决定其精度上限和可迁移性, 影响实际计算的精度和效率, 因此有必要对赝势进行充分测试, 并且发布测试报告作为用户选择赝势的重要依据. Quantum ESPRESSO开发团队^[33]、AiiDA自动化工作流与计算材料数据库团队^[76]和Materials Cloud线上开放材料科学平台团队^[77]在2018年联合发布了基于Quantum ESPRESSO的大规模赝势效率和精度测试工作^[78], 搭建了标准固体赝势数据库(Standard Solid State Pseudopotential library, SSSP), 并以其作为平台开源了全部测试数据. 该工作中提到了一些用于验证赝势精度的指标. 例如, 目前被较为广泛使用的 Δ 指

标^[79]定义如下:

$$\Delta(a, b) = \frac{1}{V_M - V_m} \sqrt{\int_{V_m}^{V_M} [E_a(V) - E_b(V)]^2 dV}, \quad (10)$$

其中, a 和 b 代表采用赝势计算相对于采用全电子计算的差别, V_M 和 V_m 分别对应测试中使用的最大和最小晶胞体积. 此处, E_a 和 E_b 代表赝势和全电子方法计算得到的晶体状态方程中能量随体积的变化. 具体来说, 能量公式为Birch - Murnaghan固体状态方程^[80], 可写为

$$E(V) = E_0 + \frac{9V_0B_0}{16} \left\{ \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^3 B'_0 + \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^2 \left[6 - 4 \left(\frac{V_0}{V} \right)^{2/3} \right] \right\}, \quad (11)$$

其中, V_0 为参考体积, B_0 为体积模量, B'_0 为其关于压力的一阶导数.

除了 Δ 指标之外, 还可以通过比较两种不同方法 a 和 b 计算得到的晶体能带来定义其中价带(valence bands)能级的差别 $\eta_v(a, b)$, 其表达式为

$$\eta_v(a, b) = \min_{\omega} \sqrt{\frac{\sum_{\mathbf{nk}} \tilde{f}_{\mathbf{nk}} (\epsilon_{\mathbf{nk}}^a - \epsilon_{\mathbf{nk}}^b + \omega)^2}{\sum_{\mathbf{nk}} \tilde{f}_{\mathbf{nk}}}}, \quad (12)$$

其中, ε_{nk} 代表在布里渊区 k 点、能带指标为 n 的 Kohn-Sham 方程的本征值. 能级平移量为 ω . 此外, 还有函数定义为

$$\tilde{f}_{nk} = \sqrt{f_{nk}^a(\varepsilon_F^a, \nu, \sigma) f_{nk}^b(\varepsilon_F^b, \nu, \sigma)}, \quad (13)$$

其中, $f_{nk}(\varepsilon, \sigma)$ 是 Fermi-Dirac 分布函数, 而 σ 为展宽系数. ε_F^a 和 ε_F^b 代表两种计算里得到的费米能级, ν 为能级平移量.

进一步, 更多的第一性原理软件开发团队包括 CASTEP^[34]、SIESTA^[37]、VASP^[81]、CP2K^[35]、BigDFT^[82]、Wien2k^[3]、FLEUR(<https://www.flapw.de/>) 等, 于 2024 年发布了针对赝势的软件之间精度对比测试, 覆盖了较大范围的体系^[83]. 以上两个工作切实推动了第一性原理计算软件之间的有效对比, 为第一性原理软件用户在计算效率和精度方面的比较提供了良好的借鉴, 对于将来发展以数据为核心的各种机器学习方法具有重要意义.

为了支持 OpenLAM 项目的稳定和快速发展, 确保 ABACUS 的用户能方便选择和使用赝势及数值原子轨道基组进行第一性原理计算, 并且能在一些关键性质上(如状态方程、能带结构)和全电子方法进行对比, ABACUS 开发团队提出了赝势-数值原子轨道库 APNS(ABACUS-pseudopotential-numerical atomic orbital-square)项目. 该项目旨在提供赝势及赝势-轨道联合测试 workflow、测试数据开源网站和自动化高通量轨道生成 workflow. 目前, 这三方面工作已取得阶段性成果, 可以在 ABACUS 的官方文档找到相关信息(<https://abacus.deepmodeling.com/en/latest/advanced/pp-orb.html>). APNS 网站上展示了使用 APNS workflow 进行赝势计算的测试结果, 例如对于模守恒赝势, 可以通过选取一系列的能量截断值来判断总能量、晶体所受压力和能带上能级的收敛情况, 网站上给出了当这三个指标都收敛时的能量截断值, 方便用户决定所使用的能量截断值. 图 3 展示了硅的模守恒赝势的测试例子. 目前, APNS 网站上的测试已涵盖了元素周期表上超过 80 种元素. 对于每种元素, 我们均测试了其稳定的单质晶体结构, 这些结构是从 Materials Project^[84]、Crystallography Open Database(COD)^[85] 等在线开放材料数据库获取的. 以 ABACUS 为密度泛函理论计算工具, APNS 网站提供

了详尽的赝势测试数据, 目前已为 OpenLAM 项目在赝势选取上提供了大范围支持. 接下来, APNS 网站还将持续建设, 覆盖更大范围的赝势和数值原子轨道测试.

2 ABACUS 的异构并行

2.1 异构并行策略

随着深度学习在过去十几年的快速发展, 异构计算——即利用不同类型的处理器或计算资源协同工作以加速计算任务的计算范式——已成为新的发展趋势. 例如, 可以采用图形处理单元(GPU)和深度计算单元(deep computing unit, DCU)来加速科学计算. 这里提到的 DCU 是由海光(HYGON)公司推出的一款专门用于人工智能(AI)和深度学习的加速卡, 适用于密集型数值计算. 密度泛函理论中最耗时的部分也往往是基于线性代数的矩阵和向量运算, 因此有望通过异构并行的方式来大幅提升软件的计算效率.

近年来, ABACUS 的发展也考虑了面向不同的异构计算设备, 为此我们建立了统一的异构编程计算框架. 对于平面波基组下的 Kohn-Sham 方程求解, 采用共轭梯度法或 Davidson 迭代法求解矩阵的特征值和特征向量往往是最耗时的操作, 因此也是异构加速的重点. 针对这类问题, ABACUS 的异构框架主要包含两大组件: 数学库和算子. 一方面, 数学库提供了统一的数学函数接口, 包括线性代数运算和快速傅里叶变换等. 这些数学库包括 CUDA 加速实现的 cuBLAS、cuSolver 和 cuFFT, ROCm 加速实现的 hipBLAS、hipSolver 和 hipFFT, 以及 CPU 上的 Intel 数学核心库(MKL). 另一方面, 算子(operator)是指在程序中定义的自定义计算操作, 用于封装特定的计算逻辑和内存管理, 例如和向量或矩阵有关的特定操作. 算子简化了在不同异构框架上执行这些自定义计算的过程, 算子的支持覆盖了包括英伟达 CUDA、AMD ROCm 和 Intel oneAPI 等在内的平台. 通过以上最大程度的抽象, ABACUS 里实现迭代法求解矩阵的算法就可以尽可能独立于特定硬件加速器去编写, 使得相同的科学计算代码能够在不同硬件平台上高效运行, 而不需要为每种硬件单独编写代码, 极大提升了核心代码面向不同硬件时的可移植性. 例如, 以共轭梯度法

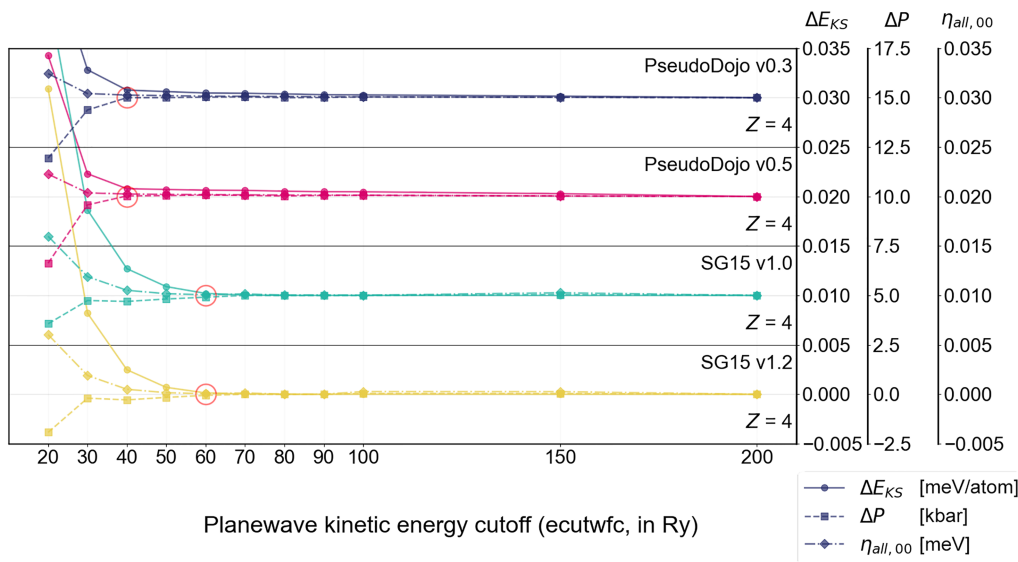


图 3 对硅元素, 采用ABACUS 的平面波基组结合四种不同模守恒赝势的收敛性测试. 横坐标为平面波的能量截断值ecutwfc, 纵坐标代表三个收敛指标. 图中四种模守恒赝势分别为PseudoDojo v0.3、PseudoDojo v0.5、SG15 v1.0、SG15 v1.2. 赝势中价电子数目为Z. 如图右下角所示, 以不同线型区分三种收敛指标. 实线代表 ΔE_{KS} (单位: meV/atom), 表示Kohn-Sham 总能量与最大ecutwfc得到的总能量差别, 点划线 ΔP (单位: kbar)代表相应的两个计算得到的压强差别, 虚线 $\eta_{all,00}$ (单位: meV)则代表能带结构的差别. 将硅元素测试收敛的标准定为 $|\Delta E_{KS}| \leq 1$ meV/atom, $|\Delta P| \leq 0.1$ kbar, $|\eta_{all,00}| \leq 10$ meV. 当标准达到时, 以红色圆圈标记对应的能量截断值ecutwfc, 可以看到前两个赝势达到收敛标准的ecutwfc为40 Ry, 而后两个赝势达到收敛标准为ecutwfc取60 Ry

Figure 3 For silicon, a convergence test was performed using the plane-wave basis set of ABACUS in combination with four different norm-conserving pseudopotentials. The horizontal axis represents the energy cutoff value of the plane wave (ecutwfc), and the vertical axis represents three convergence criteria. The four norm-conserving pseudopotentials in the figure are PseudoDojo v0.3, PseudoDojo v0.5, SG15 v1.0, and SG15 v1.2. The number of valence electrons in the pseudopotential is denoted by Z. As shown in the lower right corner of the figure, we distinguish the three convergence criteria with different line types. The solid line represents ΔE_{KS} (unit: meV/atom), which indicates the difference between the Kohn-Sham total energy and the total energy obtained with the maximum ecutwfc. The dash-dot line ΔP (unit: kbar) represents the difference in pressure between the two corresponding calculations, and the dashed line $\eta_{all,00}$ (unit: meV) represents the difference in the band structure. We set the convergence criteria for silicon as $|\Delta E_{KS}| \leq 1$ meV/atom, $|\Delta P| \leq 0.1$ kbar, $|\eta_{all,00}| \leq 10$ meV. When the criteria are met, we mark the corresponding energy cutoff value ecutwfc with a red circle. It can be seen that for the first two pseudopotentials, the ecutwfc value that meets the convergence criteria is 40 Ry, while for the last two pseudopotentials, the ecutwfc value that meets the convergence criteria is 60 Ry

求解特征值的具体实现为例, 计算操作如gemv(数学库: 矩阵向量乘法)和vecdiv(算子: 向量间的除法操作)在各种异构计算平台(包括CPU、GPU 和DCU)下被封装为统一的调用接口按需调用, 使得算法能够在不关注底层实现细节的情况下, 享受到高效的跨平台高性能计算支持.

除了迭代法求特征值之外, 我们还针对ABACUS 平面波基组下的其他热点函数进行了异构并行, 例如求原子受力和原胞应力. 此外, 2024年ABACUS开发团队也实现了基于数值原子轨道的异构并行. 基于以上工作, ABACUS目前已经基本实现了面向GPU和DCU的异构加速, 大幅提升了计算效率, 使得大原子模型的第一性原理训练数据可以通过在不同硬件上(包括国产硬件)进行高效计算获得.

2.2 异构并行下的精度和效率测试

为了评估由于不同硬件架构带来的精度偏差, 我们采用ABACUS软件的平面波基组求解了包含有95个构型数据的测试集(下载地址: <https://github.com/deepmodeling-activity/abacus-test.github.io/tree/master/datasets/alloy>). 该测试集由66种常见元素组成, 包括二元或三元合金, 以及金属非金属化合物等, 原子个数的范围覆盖从2到53, 电子数的范围覆盖从13到503, 这66种元素具体是H, B, N, O, F, Na, Mg, Al, Si, P, Cl, K, Ca, Ti, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Y, Zr, Nb, Mo, Ru, Rh, Pd, Cd, In, Sn, Sb, I, Cs, Ba, La, Ce, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, W, Re, Os, Ir, Pt, Au, Bi, Th, U, Np.

测试在CPU和曙光DCU硬件平台上进行,其中CPU包含32个核心,型号为Intel(R) Xeon(R) Platinum 9242 CPU @ 2.30 GHz, DCU是曙光超算平台编号为7185-32C-128G-4卡,包含4个DCU加速卡和32个CPU核心.我们测试了自洽场迭代(SCF)计算,其中平面波动能截断设置为100 Ry,倒空间k点采样间隔设置为0.08 bohr⁻¹,自洽迭代的收敛标准为密度差小于10⁻⁸,电荷混合方法使用Broyden方法,并使用0.015 Ry的高斯展宽,和使用CG方法进行对角化求解.图4展示了测试获得的体系能量、原子受力和晶胞应力的偏差.从图中可以看出,无论是能量、力还是应力的计算结果,在CPU和DCU平台上都展现出了高度的一致性.例如,能量偏差保持在10⁻⁴ ~ 10⁻⁶ meV/atom的量级,力的偏差在10⁻² ~ 10⁻⁵ eV/Å的范围内,而应力的偏差则在1 ~ 0.01 kbar的量级.这些偏差主要是由于不同计算平台的编译环境差异,导致波函数的初始化不一致所带来的,但它们均在可接受的误差范围内.这表明ABACUS软件能够在多种计算平台上提供稳定且可靠的结果.

我们进一步采用该测试集测试了ABACUS平面波基组方法的异构并行加速效果.图5展示了ABACUS对不同大小的体系,在CPU(32个核心,型号为Intel(R) Xeon(R) Platinum 9242 CPU @ 2.30GHz)和DCU(包含4个DCU加速卡以及32个CPU核心)平台进行自洽

场迭代(SCF)计算的效率.其中CPU上的计算使用的是Davidson(DAV)和CG方法,DCU上的计算使用的是CG方法.从图中可以看出,随着体系的增大(系统的总价层电子数增加),CPU上的耗时与DCU上的耗时比例呈现增大的趋势.对于小体系(如电子数小于50),32核CPU与4卡DCU的计算效率相当(耗时比值约为1~2),而对于大体系(如电子数超过50),DCU的加速效果更好.对于CG方法,4卡DCU在计算超过200电子的体系时,相比32核CPU,其效率可提升约6~10倍.在OpenLAM这种需要大量DFT计算的项目中,这个效率加速可降低成本.

3 ABACUS与预训练模型

3.1 合金模型数据

在OpenLAM项目中,ABACUS的平面波赝势方法结合PBE交换关联泛函生产了大量合金体系的第一性原理数据.这个数据集涵盖了从轻金属如锂(Li)和铝(Al)到重金属如金(Au)和铅(Pb),还包括了过渡金属和稀土金属,共包含有53种常见金属如下: Ag, Al, Au, Be, Ca, Cd, Ce, Co, Cr, Cu, Dy, Er, Fe, Ga, Gd, Ge, Hf, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, Na, Nb, Nd, Ni, Os, Pb, Pd, Pr, Pt, Re, Rh, Ru, Sc, Si, Sm, Sn, Sr, Ta, Tb, Ti, Tm, V, W, Y, Zn, Zr.

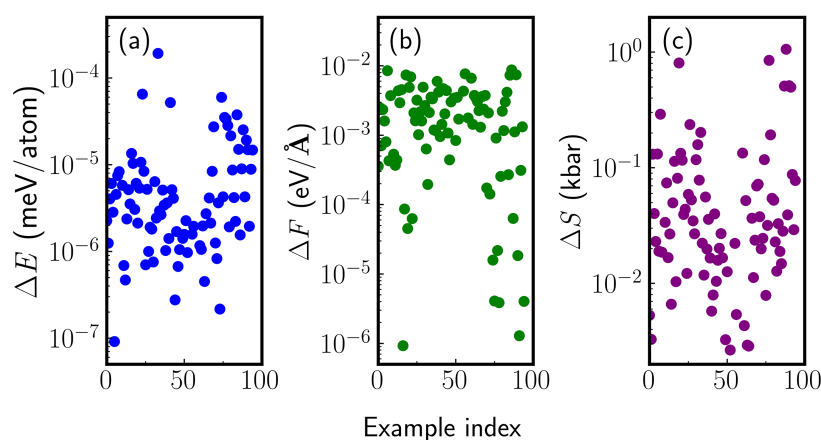


图4 使用ABACUS在CPU和DCU平台上的计算结果对比. 横轴表示不同的算例编号, 纵轴表示两个平台计算结果的偏差. (a) 平均每原子的能量偏差 ΔE (meV/atom); (b) 力所有分量的最大偏差 ΔF (eV/Å); (c) 应力所有分量的最大偏差 ΔS (kbar)

Figure 4 Comparison of computational results using ABACUS on CPU and DCU platforms. The horizontal axis represents the example index, and the vertical axis shows the deviation of results on two platforms. (a) The deviation of energy per atom (unit in meV/atom); (b) the maximum absolute deviation of force components (unit in eV/Å); (c) the maximum absolute deviation of stress components (unit in kbar)

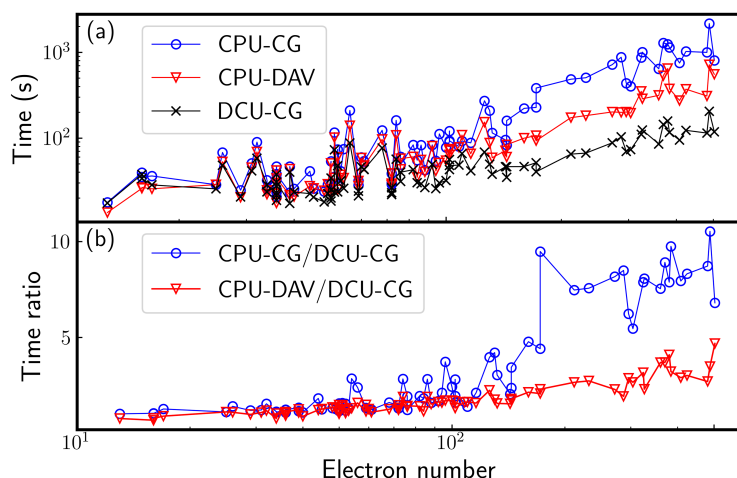


图5 使用ABACUS在CPU与DCU平台上, 对不同大小的95个体系的平均自洽场迭代计算耗时对比. CPU上的计算使用Davidson(DAV)和共轭梯度法(CG)对角化方法. DCU的计算使用CG对角化方法. 图(a)展示了分别使用CPU平台上CG/DAV方法和DCU平台上CG方法计算的平均每SCF电子步的耗时; 图(b)展示了CPU上CG和DAV方法与DCU上CG方法的耗时比值. 横轴表示测试体系的价层电子数量

Figure 5 Comparison of average self-consistent field (SCF) iteration computation time per step on CPU and DCU platforms for 95 different systems by ABACUS. The calculations on the CPU platform were performed employing the Davidson (DAV) and CG diagonalization methods. The DCU calculations use the CG diagonalization method. Figure (a) shows the cost time of the CG/DAV method on the CPU platform and the CG method on the DCU platform for systems of different sizes. Figure (b) shows the time ratio of CG and DAV methods on the CPU to CG methods on the DCU. The horizontal axis represents the number of valence electrons in the test system

对这些元素所采用的赝势需能合理预测材料的物理性质. 例如, 弹性模量是描述材料抵抗形变能力的物理量, 它对于理解材料在受力时的行为至关重要, 是合金材料最重要的性质之一. 我们利用APEX工具^[86]结合ABACUS, 对53种金属单质的弹性模量进行了计算(构型和参考结果: <https://github.com/deepmodeling-activity/abacus-test.github.io/tree/master/datasets/alloy-elastic>). 鉴于不同赝势对元素电子结构描述的差异性, 我们基于ABACUS赝势轨道库(ABACUS-pseudopotential-numerical atomic orbital-square, APNS)为每种元素筛选了多种赝势. 在APNS中, 我们对每种元素的多种赝势进行了状态方程(equation of state, EOS)的计算, 并与Wien2K^[83]全电子方法计算的EOS进行了比较, 通过计算两者之间的 Δ 值(式(10)), 筛选出了 Δ 值最小的5种赝势, 用于后续的弹性模量计算测试. 补充材料的表S1中同时列出了对应赝势的 Δ 值. 对于平面波截断能的设置, 依据APNS中的收敛性测试结果, 为每种元素确定了满足能量收敛性的最低平面波截断能阈值. 在大模型的训练中, 所有系统的计算模拟需要使用相同的计算参数, 以保证计算结果的统一. 在测试时, 对于收敛的能量截断值大于100

Ry的赝势, 我们额外进行了能量截断值为100 Ry的补充计算, 以验证这些元素在低于能量截断值收敛的参数设置下的性质计算结果. 例如, 对于铜(Cu)的SG15赝势, 在能量截断值为200 Ry的基础上, 额外计算了能量截断值为100 Ry时的弹性模量.

在计算中, 设定倒空间 k 点的最小采样间隔为0.08 bohr⁻¹, 以保证在大多数体系中能获得足够密的布里渊区 k 点采样区. 作为对比, 同时使用APEX结合VASP软件^[32](v5.4.4版本), 使用PAW方法和PBE泛函也计算了这些体系的弹性模量. 图6和图7展示了ABACUS和VASP预测的53种金属单质的体模量(BV)和剪切模量(GV). 图中展示的是每种元素所有赝势计算与VASP最接近的结果, 其余赝势计算的体模量和剪切模量以及其他弹性常数矩阵分量的结果列在表S1中. 结果表明, ABACUS的预测值与VASP的预测值显示出较好的一致性. 同时我们注意到, 使用不同价电子数的赝势进行弹性模量的预测时, 有可能会得到差别较大的结果. 例如, 对于Ce的面心立方晶体结构, ABACUS使用价电子数为11(4f的电子作为内层电子处理)的PD04模守恒赝势, 计算的BV和GV分别为32和17 GPa, 而使用价电子数为12(把4f的电子作为价层电

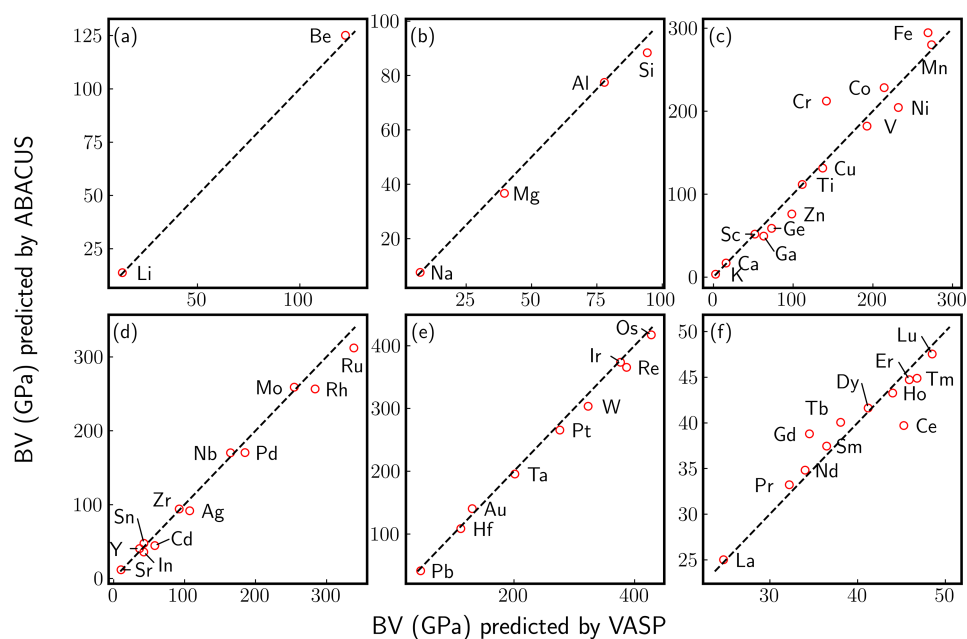


图 6 ABACUS和VASP软件计算得到的53种常见金属单质的体弹性模量(BV, 单位GPa). (a)~(e) 测试的第二至六周期元素的计算结果. (f) 所测试镧系元素的数据

Figure 6 The bulk modulus (BV, in GPa) of 53 common metallic elements predicted by using ABACUS and VASP software. (a)~(e) The predicted results for elements in the second, third, fourth, fifth, and sixth periods, respectively; (f) the data for the tested lanthanide elements

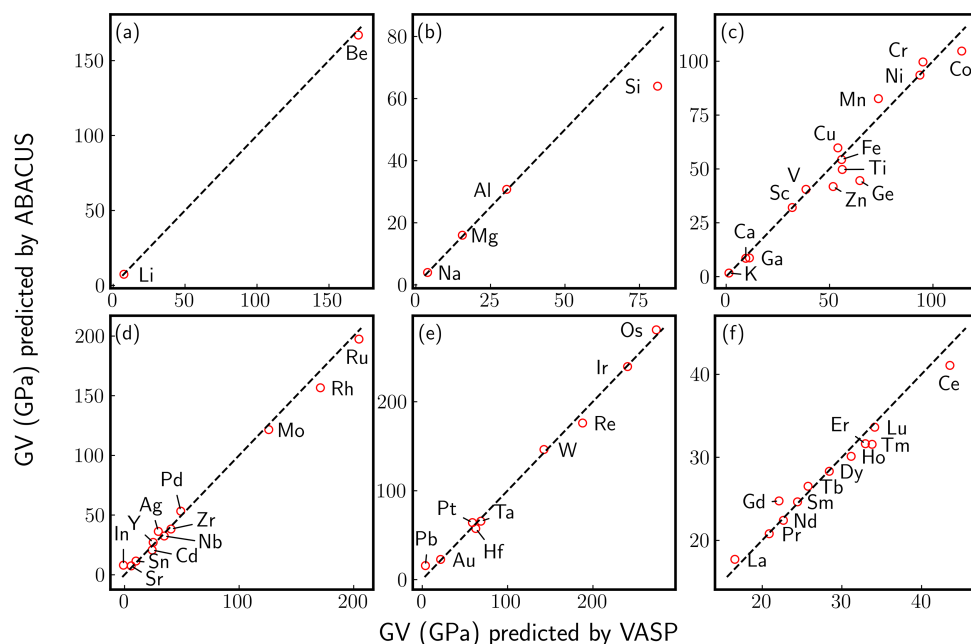


图 7 ABACUS和VASP软件计算得到的53种常见金属单质的剪切弹性模量(GV, 单位GPa). (a)~(e) 测试的第二至六周期元素的计算结果. (f) 镧系元素的测试数据

Figure 7 The shear modulus (GV, in GPa) of 53 common metallic elements predicted by using ABACUS and VASP software. (a)~(e) The predicted results for elements in the second, third, fourth, fifth, and sixth periods, respectively; (f) the data for the tested lanthanide elements

子)的PD04赝势, 计算得到的BV和GV分别是37和39 GPa. 相应地, 在VASP的计算中, 使用的是价电子数为12的Ce赝势, 其计算的BV和GV分别是45和44 GPa.

3.2 其他模型数据

钙钛矿氧化物因其结构和化学的灵活性, 成为探索不同自由度之间相互作用产生新现象的重要材料. 传统的分子动力学模拟依赖经典的力场, 这些力场通常被描述为参数化的解析函数, 它们对阐明包括钙钛矿氧化物在内的晶体固体的原子动力学和结构特性做出了重要贡献. 但是这些力场通常是针对特定材料或材料群定制的, 不能用于广泛固体材料的计算, 这阻碍了分子动力学在计算机辅助材料发现中的便捷部署. 来自西湖大学的Wu等人^[87]基于DPA-1模型开发了钙钛矿氧化物的通用原子间势能(UniPero). 该通用势函数覆盖了26种钙钛矿氧化物、200个组分以及14种金属元素. 所有密度泛函理论数据均来自ABACUS的数值原子轨道基组计算. 作者们首先使用DP-GEN获得三元系统(如 PbTiO_3 、 SrTiO_3)的DPA-1模型. 随后, 使用收敛的数据集作为起点, 通过DP-GEN改进四元体系(例如 $\text{Pb}_x\text{Sr}_{1-x}\text{TiO}_3$ 与 $\text{PbZr}_x\text{Ti}_{1-x}\text{O}_3$)的DPA-1模型. 通过不断迭代, DPA-1模型(UniPero)可以描述涉及14种金属元素及其任意组成的固溶体钙钛矿氧化物. 在模型的迭代训练过程中, 他们使用ABACUS用少于200000个CPU小时的成本, 获得了约20000个构型结果. 最后, 为了验证模型的准确性, 他们使用此模型结合分子动力学模拟, 准确地预测了几种不同铁电氧化物的相变序列. 例如, 包含6种元素的三元固溶体 $\text{Pb}(\text{In}_{1/2}\text{Nb}_{1/2})\text{O}_3$ - $\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3$ - PbTiO_3 (PIN-PMN-PT)体系. 该模型的成功建立体现了ABACUS的数值原子轨道基组方法的可靠性, 并且ABACUS较高的计算效率也大幅降低了训练数据的成本.

周期表中II族B到VIA族元素组成的化合物半导体因卓越的光电子特性, 在现代社会的光伏、光电子、热电等领域发挥着重要作用. Liu等人^[88]也基于ABACUS产生的密度泛函理论数据生成了一个基于DPA-1、面向19种半导体的预训练模型DPA-Semi. 该训练数据集包含了23万多个原子构型下的数据集, 全部采用ABACUS的数值原子轨道(NAO)基组进行密

度泛函理论计算. 这19种半导体材料包含了从IIB族至VIA族的Si, Ge, SiC, BAs, BN, AlN, AlP, AlAs, InP, InAs, InSb, GaN, GaP, GaAs, CdTe, InTe, CdSe, ZnS, CdS. 通过与密度泛函理论计算的结果进行比较, 作者们发现DPA-Semi模型在预测这些半导体材料的晶格常数和弹性模量方面具有较高的量子力学精度. 例如, 对于SiC, DPA-Semi模型能够准确区分不同异构体(如3C、2H、4H和6H)之间的能量差异, 并且与密度泛函理论计算的能量顺序关系一致, 即 $E_{2H} > E_{3C} > E_{6H} > E_{4H}$. 此外, 作者们采用DPA-Semi模型计算了半导体材料的声子谱, 以及进一步结合分子动力学方法模拟了半导体的熔点和非晶态结构, 通过对比前人数据和实验数据, 获得了合理的结果. 以上测试表明DPA-Semi模型不仅能以第一性原理的精度预测半导体材料的各种性质, 而且还能显著降低下游任务的计算成本, 为研究IIB族至VIA族半导体系统的科学问题提供了一个新的工具.

4 总结和展望

AI for Science领域的兴起为解决传统微观尺度算法和软件所遇到的精度和效率难以兼得的困境提供了新的可能. 在分子动力学领域, 基于机器学习的分子动力学方法如今已获得许多应用, 但是训练势函数的成本仍然过大, 无法满足日益增长的材料模拟需求. 目前, 国际上多个团队正在设计面向元素周期表的统一机器学习势函数模型, 希望能够通过该模型描述由元素周期表上的大部分元素所组成的材料体系. 这是一个相当宏伟的计划, 但要得到一个好的模型所需的海量第一性原理训练数据仍需从第一性原理材料模拟软件获得, 密度泛函理论方法是目前产生训练数据所使用的主要理论方法.

本文介绍了国产开源第一性原理材料计算软件ABACUS(原子算筹)助力原子势函数大模型计划OpenLAM的进展情况. 2021年, ABACUS开发团队与科学计算软件的开源社区DeepModeling展开深度合作, 致力于建设AI4S时代的新一代电子结构计算平台. ABACUS的主要功能是支持基于平面波或数值原子轨道的密度泛函理论计算. 过去几年, ABACUS软件秉承开源代码的精神进行快速迭代, 软件在易用性和稳定性上有了质的提升, 并且软件的发布

速度是保持每三个月发布一个大版本, 每两周发布一个小版本. 在此基础上, ABACUS不仅实现了一系列全新的电子结构算法, 还对求解Kohn-Sham方程等核心功能进行了高效的GPU/DCU异构加速, 并与国产曙光DCU硬件平台进行了算法的适配移植. ABA-CUS与国产硬件的结合极大提升了产生第一性原理数据的性价比, 使得ABACUS相比于国际其他主流密度泛函理论软件具备了性价比上的优势, 从而成为了为OpenLAM项目提供第一性原理训练数据的重要工具. 目前, 由ABACUS数据生成的机器学习势函数模型已应用于合金、半导体、高压氢化物等多个场景并取得了良好的效果. 在OpenLAM需求的推动下, 赝势库和相应的数值轨道基组库的建设也日趋完善, 方便用户

直接选取合适的赝势和数值原子轨道文件.

随着OpenLAM项目的不断深入, 其应用范围日益扩大, 涵盖了更多的实际场景. 在此过程中, ABA-CUS生成的大量数据被积极地应用于模型训练之中. 与此同时, 相关的模型、训练数据集以及工作流也陆续在AI Square平台(<https://www.aissquare.com/>)上进行不同程度的公开分享, 以促进知识的交流与合作. 除了助力大原子模型OpenLAM计划, ABACUS开源项目会持续对代码框架进行必要的调整和重构, 以适应多种高性能硬件架构和新的AI算法. 同时, ABACUS会持续地开发新的电子结构算法, 以及提升代码的稳定性和可靠性. ABACUS开源社区欢迎更多的开发者和用户关注.

参考文献

- 1 Hohenberg P, Kohn W. Inhomogeneous electron gas. *Phys Rev*, 1964, 136: B864
- 2 Kohn W, Sham L. Self-consistent equations including exchange and correlation effects. *Phys Rev*, 1965, 140: A1133
- 3 Blaha P, Schwarz K, Tran F, et al. WIEN2k: an APW+ lo program for calculating the properties of solids. *J Chem Phys*, 2020, 152: 074101
- 4 Zhang Y, Suo B, Wang Z, et al. BDF: a relativistic electronic structure program package. *J Chem Phys*, 2020, 152: 064113
- 5 Romero A, Allan D, Amadon B, et al. ABINIT: overview and focus on selected capabilities. *J Chem Phys*, 2020, 152: 124102
- 6 Nakata A, Baker J, Mujahed S, et al. Large scale and linear scaling DFT with the CONQUEST code. *J Chem Phys*, 2020, 152: 16112
- 7 KÜHne T, Iannuzzi M, Del Ben M, et al. CP2K: an electronic structure and molecular dynamics software package-Quickstep: efficient and accurate electronic structure calculations. *J Chem Phys*, 2020, 152: 194103
- 8 Tancogne-Dejean N, Oliveira M, Andrade X, et al. Octopus, a computational framework for exploring light-driven phenomena and quantum dynamics in extended and finite systems. *J Chem Phys*, 2020, 152: 124119
- 9 Prentice J, Aarons J, Womack J, et al. The ONETEP linear-scaling density functional theory program. *J Chem Phys*, 2020, 152: 174111
- 10 Sun Q, Zhang X, Banerjee S, et al. Recent developments in the PySCF program package. *J Chem Phys*, 2020, 153: 024109
- 11 Giannozzi P, Barone A, Bonfá P, et al. Quantum ESPRESSO toward the exascale. *J Chem Phys*, 2020, 152: 154105
- 12 García A, Papior N, Akhtar A, et al. Siesta: recent developments and applications. *J Chem Phys*, 2020, 152: 204108
- 13 Kresse G, Joubert D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B*, 1999, 59: 1758
- 14 Zhang L, Han J, Wang H, et al. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys Rev Lett*, 2018, 120: 143001
- 15 Batatia I, Kovacs D, Simm G, et al. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv Neural Inf Proc Syst*, 2022, 35: 11423–11436
- 16 Batatia I, Benner P, Chiang Y, et al. A foundation model for atomistic materials chemistry. 2023, arXiv: 2401.00096
- 17 Chen C, Ong S. A universal graph deep learning interatomic potential for the periodic table. *Nat Comput Sci*, 2022, 2: 718–728
- 18 Qi J, Ko T, Wood B, et al. Robust training of machine learning interatomic potentials with dimensionality reduction and stratified sampling. *npj Comput Mater*, 2024, 10: 43
- 19 Choudhary K, Decost B, Major L, et al. Unified graph neural network force-field for the periodic table: solid state applications. *Digital Disc*, 2023, 2: 346–355
- 20 Choudhary K, Decost B. Atomistic line graph neural network for improved materials property predictions. *npj Comput Mater*, 2021, 7: 185
- 21 Deng B, Zhong P, Jun K, et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat Mach Intell*, 2023, 5: 1031–1041
- 22 Merchant A, Batzner S, Schoenholz S, et al. Scaling deep learning for materials discovery. *Nature*, 2023, 624: 80–85
- 23 Zhang L, Wang H. Simulating the microscopic world: from the Schrödinger equation to the large atomic model (in Chinese). *Physics*, 2024, 53: 431–441 [张林峰, 王涵. 模拟微观世界: 从薛定谔方程到大原子模型. *物理*, 2024, 53: 431–441]

- 24 Zhang D, Bi H, Dai F, et al. Pretraining of attention-based deep learning potential model for molecular simulation. *npj Comput Mater*, 2024, 10: 94
- 25 Zhang D, Liu X, Zhang X, et al. DPA-2: towards a universal large atomic model for molecular and material simulation. 2023, arXiv:2312.15492
- 26 He F, Ren X, Jiang J, et al. Real-time, time-dependent density functional theory study on photoinduced isomerizations of azobenzene under a light field. *J Phys Chem Lett*, 2022, 13: 427–432
- 27 Lin P, Ren X, He L. Efficient hybrid density functional calculations for large periodic systems using numerical atomic orbitals. *J Chem Theory Comput*, 2020, 17: 222–239
- 28 Liu Q, Chen M. Plane-wave-based stochastic-deterministic density functional theory for extended systems. *Phys Rev B*, 2022, 106: 125132
- 29 Sun L, Chen M. Machine learning based nonlocal kinetic energy density functional for simple metals and alloys. *Phys Rev B*, 2024, 109: 115135
- 30 Chen Y, Zhang L, Wang H, et al. DeePKS-kit: a package for developing machine learning-based chemically accurate energy and density functional models. *Comput Phys Commun*, 2023, 282: 108520
- 31 Li W, Ou Q, Chen Y, et al. DeePKS+ ABACUS as a bridge between expensive quantum mechanical models and machine learning potentials. *J Phys Chem A*, 2022, 126: 9154–9164
- 32 Kresse G, Furthmüller J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput Mater Sci*, 1996, 6: 15–50
- 33 Giannozzi P, Baroni S, Bonini N, et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J Phys Condens Matter*, 2009, 21: 395502
- 34 Clark S, Segall M, Pickard C, et al. First principles methods using CASTEP. *Z kristall*, 2005, 220: 567–570
- 35 Hutter J, Iannuzzi M, Schiffmann F, et al. cp2k: atomistic simulations of condensed matter systems. *Wiley Interdiscip Rev: Comput Mol Sci*, 2014, 4: 15–25
- 36 Boker S, Neale M, Maes H, et al. OpenMx: an open source extended structural equation modeling framework. *Psychometrika*, 2011, 76: 306–317
- 37 Soler J, Artacho E, Gale J, et al. The SIESTA method for *ab initio* order-N materials simulation. *J Phys Condens Matter*, 2002, 14: 2745
- 38 Blum V, Gehrke R, Hanke F, et al. *Ab initio* molecular simulations with numeric atom-centered orbitals. *Comput Phys Commun*, 2009, 180: 2175–2196
- 39 Wang H, Zhang L, Han J, et al. DeePMD-kit: a deep learning package for many-body potential energy representation and molecular dynamics. *Comput Phys Commun*, 2018, 228: 178–184
- 40 Zeng J, Zhang D, Lu D, et al. DeePMD-kit v2: a software package for deep potential models. *J Chem Phys*, 2023, 159: 054801
- 41 Zhang Y, Wang H, Chen W, et al. DP-GEN: a concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput Phys Commun*, 2020, 253: 107206
- 42 Li H, Wang Z, Zou N, et al. Deep-learning density functional theory Hamiltonian for efficient *ab initio* electronic-structure calculation. *Nat Comput Sci*, 2022, 2: 367–377
- 43 Gu Q, Zhouyin Z H, Pandey S K, et al. DeePTB: A deep learning-based tight-binding approach with *abinitio* accuracy. 2023, arXiv:2307.04638
- 44 Zheng Q, Chu W, Zhao C, et al. *Ab initio* nonadiabatic molecular dynamics investigations on the excited carriers in condensed matter systems. *Wiley Interdiscip Rev: Comput Mol Sci*, 2019, 9: e1411
- 45 Wang X, Wang Z, Gao P, et al. Data-driven prediction of complex crystal structures of dense lithium. *Nat Commun*, 2023, 14: 2924
- 46 Li W, Carrete J, Katcho N, et al. ShengBTE: a solver of the Boltzmann transport equation for phonons. *Comput Phys Commun*, 2014, 185: 1747–1758
- 47 Mostofi A, Yates J, Lee Y, et al. wannier90: a tool for obtaining maximally-localised Wannier functions. *Comput Phys Commun*, 2008, 178: 685–699
- 48 Larsen A, Mortensen J, Blomqvist J, et al. The atomic simulation environment—a Python library for working with atoms. *J Phys: Condens Matter*, 2017, 29: 273002
- 49 Jin G, Pang H, Ji Y, et al. Pyatb: an efficient python package for electronic structure calculations using *ab initio* tight-binding model. *Comput Phys Commun*, 2023, 291: 108844
- 50 He X, Helbig N, Verstraete M, et al. TB2J: A python package for computing magnetic interaction parameters. *Comput Phys Commun*, 2021, 264: 107938
- 51 Blum V, Gehrke R, Hanke F, et al. *Ab initio* molecular simulations with numeric atom-centered orbitals. *Comp Phys Comm*, 2009, 180: 2175–2196
- 52 Talman J. Numerical Fourier and Bessel transforms in logarithmic variables. *J Comput Phys*, 1978, 29: 35–48
- 53 Havu V, Blum V, Havu P, et al. Efficient O (N) integration for all-electron electronic structure calculation using numeric basis functions. *J Comput Phys*, 2009, 228: 8367–8379
- 54 Sankey O, Niklewski D. *Ab initio* multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Phys Rev B Condens Matter*, 1989, 40: 3979–3995
- 55 Porezag D, Frauenheim G G, Köhler G G, et al. Construction of tight-binding-like potentials on the basis of density-functional theory: application to carbon. *Phys Rev B*, 1995, 51: 12947–12957

- 56 Horsfield A. Efficient *ab initio* tight binding. Phys Rev B, 1997, 56: 6594–6602
- 57 Junquera J, PazÓ, Sánchez-Portal D, et al. Numerical atomic orbitals for linear-scaling calculations. Phys Rev B, 2001, 64: 235111
- 58 Ozaki T. Variationally optimized atomic orbitals for large-scale electronic structures. Phys Rev B, 2003, 67: 155108
- 59 Ozaki T, Kino H. Numerical atomic basis orbitals from H to Kr. Phys Rev B, 2004, 69: 195113
- 60 Sánchez-Portal D, Artacho E, Soler J. Projection of plane-wave calculations into atomic orbitals. Solid State Commun, 1995, 95: 685–690
- 61 Sánchez-Portal D, Artacho E, Soler J. Analysis of atomic orbital basis sets from the projection of plane-wave results. J Phys: Condens Matter, 1996, 8: 3859
- 62 Chen M, Guo G, He L. Systematically improvable optimized atomic basis sets for *ab initio* calculations. J Phys: Condens Matter, 2010, 22: 445501
- 63 Haynes P, Payne M. Localised spherical-wave basis set for $O(N)$ total-energy pseudopotential calculations. Comput Phys Commun, 1997, 102: 17–27
- 64 Lin P, Ren X, He L. Strategy for constructing compact numerical atomic orbital basis sets by incorporating the gradients of reference wavefunctions. Phys Rev B, 2021, 103: 235131
- 65 Schlipf M, Gygi F. Optimization algorithm for the generation of ONCV pseudopotentials. Comput Phys Commun, 2015, 196: 36–44
- 66 Schwerdtfeger P. The pseudopotential approximation in electronic structure theory. ChemPhysChem, 2011, 12: 3143–3155
- 67 Hamann D, Schlüter M, Chiang C. Norm-conserving pseudopotentials. Phys Rev Lett, 1979, 43: 1494
- 68 Van Setten M, Giantomassi M, Bousquet E, et al. The PseudoDojo: training and grading a 85 element optimized norm-conserving pseudopotential table. Comput Phys Commun, 2018, 226: 39–54
- 69 Dal Corso A. Pseudopotentials periodic table: from H to Pu. Comput Mater Sci, 2014, 95: 337–350
- 70 Vanderbilt D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Phys Rev B, 1990, 41: 7892
- 71 Troullier N, Martins J. Efficient pseudopotentials for plane-wave calculations. Phys Rev B, 1991, 43: 1993
- 72 Rappe A, Rabe K, Kaxiras E, et al. Optimized pseudopotentials. Phys Rev B, 1990, 41: 1227
- 73 Krack M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. Theor Chem Acc, 2005, 114: 145–152
- 74 Garrity K, Bennett J, Rabe K, et al. Pseudopotentials for high-throughput DFT calculations. Comput Mater Sci, 2014, 81: 446–452
- 75 Jollet F, Torrent M, Holzwarth N. Generation of projector augmented-wave atomic data: a 71 element validated table in the XML format. Comput Phys Commun, 2014, 185: 1246–1254
- 76 Pizzi G, Cepellotti A, Sabatini R, et al. AiiDA: automated interactive infrastructure and database for computational science. Comput Mater Sci, 2016, 111: 218–230
- 77 Talirz L, Kumbhar S, Passaro E, et al. Materials Cloud, a platform for open computational science. Sci data, 2020, 7: 299
- 78 Prandini G, Marrazzo A, Castelli I E, et al. Precision and efficiency in solid-state pseudopotential calculations. npj Comput Mater, 2018, 4: 72
- 79 Lejaeghere K, Van Speybroeck V, Van Oost G, et al. Error estimates for solid-state density-functional theory predictions: an overview by means of the ground-state elemental crystals. Crit Rev Solid State Mater Sci, 2014, 39: 1–24
- 80 Birch F. Finite elastic strain of cubic crystals. Phys Rev, 1947, 71: 809
- 81 Kresse G, Furthmüller J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. Phys Rev B, 1996, 54: 11169
- 82 Mohr S, Ratcliff L, Genovese L, et al. Accurate and efficient linear scaling DFT calculations with universal applicability. Phys Chem Chem Phys, 2015, 17: 31360–31370
- 83 Bosoni E, Beal L, Bercx M, et al. How to verify the precision of density-functional-theory implementations via reproducible and universal workflows. Nat Rev Phys, 2024, 6: 45–58
- 84 Jain A, Ong S, Hautier G, et al. The materials project: a materials genome approach to accelerating materials innovation. APL Mater, 2013, 1: 011002
- 85 GraZULIS S, Chateigner D, Downs R, et al. Crystallography open database—an open-access collection of crystal structures. J Appl Crystallogr, 2009, 42: 726–729
- 86 Li Z, Wen T, Zhang Y, et al. An extendable cloud-native alloy property explorer. 2024, arXiv:2404.17330
- 87 Wu J, Yang J, Liu Y, et al. Universal interatomic potential for perovskite oxides. Phys Rev B, 2023, 108: L180104
- 88 Liu J, Zhang X, Chen T, et al. Machine-learning-based interatomic potentials for group IIB to VIA semiconductors: toward a universal model. J Chem Theory Comput, 2024, 20: 5717–5731

Summary for “开源密度泛函理论软件ABACUS: 助力大原子模型发展”

Open-source density functional theory software ABACUS: promoting the development of large atomic-scale machine learning pre-trained models

Xingliang Peng¹, Weiqing Zhou¹, Daye Zheng¹, Zuxin Jin¹, Yike Huang¹, Denghui Lu², Zhuoyuan Li^{3,4}, Tongqi Wen^{3,4} & Mohan Chen^{1,2*}

¹ AI for Science Institute, Beijing 100080, China

² Center for Applied Physics and Technology, Peking University, Beijing 100871, China

³ Center for Structural Materials, Department of Mechanical Engineering, The University of Hong Kong, Hong Kong 999077, China

⁴ Materials Innovation Institute for Life Sciences and Energy (MILES), The University of Hong Kong, Shenzhen 518000, China

* Corresponding author, E-mail: mohanchen@pku.edu.cn

Machine-learning-enriched interatomic potential frameworks are notable catalysts that transform the realm of computational materials science. By skillfully combining quantum mechanical accuracy with the computational efficiency of classical empirical potentials, these methods significantly enhance the capabilities of molecular dynamics simulations and related investigative fields. The integration of these frameworks provides researchers with the ability to analyze the complex dynamics of material systems with an unprecedented combination of accuracy and speed. The increasing application of machine learning potential function methods and the accumulating data lay a solid foundation for the construction of a large atomic model (LAM) that comprehensively covers the elements of the periodic table, turning the vision into a feasible reality. Within this paradigm, the open large atomic model (OpenLAM) project aims to build a high-precision, high-efficiency pre-trained model suitable for complex material systems. The efficient and accurate interatomic potential models can archive through further downstream fine-tuning and distillation. A key factor in this endeavor is the comprehensive collection of advanced training data, which relies on the accuracy provided by first-principles software. ABACUS, a leading open-source tool in the density functional theory landscape, plays a pivotal role in advancing the OpenLAM project. With support for plane waves and numerical atomic orbital basis sets, ABACUS enables researchers to select the most suitable basis set for their specific research needs. Additionally, ABACUS provides an accessible pseudopotential and numerical atomic orbital library, and enables users to conveniently select pseudopotentials and orbits. ABACUS also fosters a collaborative framework for first-principles computation, creating a community-oriented research environment. And at present, ABACUS has supported lots of algorithms that combine DFT and AI, such as DeepKS, DeepH, etc. ABACUS's hardware versatility, demonstrated by its successful adaptation across various platforms, significantly enhances its utility by leveraging the computational power of diverse architectures. The successful adaptation on domestic deep computing unit (DCU) hardware makes ABACUS be a cost-effective choice. This flexibility has been instrumental in enabling ABACUS to contribute a substantial amount of density functional theory data to the OpenLAM project, covering a wide range of material systems including alloys, semiconductors, and perovskites. The use of ABACUS for generating large amounts of data in these fields has validated the stability and reliability of ABACUS, with the potential to unlock novel functionalities and applications in modern technology. The evolution of ABACUS is closely aligned with the rapidly advancing field of AI for Science. As artificial intelligence continues to integrate into scientific research, ABACUS is well-positioned to incorporate emerging AI-driven methodologies and algorithms. Moving forward, ABACUS will remain a cornerstone of the OpenLAM project and a catalyst for new discoveries, driving materials science into a new era of AI-enhanced research and development.

density functional theory, ABACUS, machine learning potentials, large atomic model

doi: [10.1360/TB-2024-1243](https://doi.org/10.1360/TB-2024-1243)