



南京大学材料科学与工程系列丛书

第一性原理材料计算基础

周 健 梁奇锋◎编著



科学出版社

南京大学材料科学与工程系列丛书

第一性原理材料计算基础

周 健 梁奇锋 编著

研 字 出 版 社

北 京

内 容 简 介

本书介绍了计算材料和计算凝聚态物理学中常用的密度泛函理论、程序及应用实例,主要包括材料计算背景介绍;晶体结构和晶体对称性;能带理论和紧束缚近似;密度泛函理论基础;VASP 程序基本功能、参数和应用;材料拓扑性质理论和计算实例。

全书分为六章。第1章为绪论,主要介绍材料设计的基本概念、材料数据库的建立和应用、高性能计算和 Linux 操作系统。第2章为晶体结构和晶体对称性,主要包括晶体点阵、元胞、对称操作、点群、晶系、原子坐标和倒易空间等内容。第3章为电子能带结构,包括布洛赫定理、玻恩-冯·卡门边界条件、本征方程、紧束缚近似及一些简单材料的算例。第4章为密度泛函理论,主要介绍了 Hartree 方程、Hartree-Fock 方程、密度泛函理论基础、Kohn-Sham 方程、基组、赝势以及交换关联势等内容。第5章为密度泛函计算程序 VASP,主要介绍 VASP 程序的基本功能和常见参数,并列举了几个常见的计算实例。第6章为拓扑材料计算实例,专题介绍了材料拓扑性质的基本理论,并列举了若干使用密度泛函理论研究材料拓扑性质的计算实例。

本书可作为计算凝聚态物理、计算材料专业及其他相关专业研究生和高年级本科生的教材,也可供自学者参考。

图书在版编目(CIP)数据

第一性原理材料计算基础/周健,梁奇锋编著. —北京:科学出版社,2019.10
(南京大学材料科学与工程系列丛书)

ISBN 978-7-03-062610-3

I. ①第… II. ①周…②梁… III. ①材料科学-计算 IV. ①TB3

中国版本图书馆 CIP 数据核字 (2019) 第 225261 号

责任编辑:张析高微/责任校对:杜子昂

责任印制:吴兆东/封面设计:王浩

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京建宏印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2019年10月第一版 开本:720×1000 B5

2019年10月第一次印刷 印张:13 1/4

字数:267 000

定价:78.00 元

(如有印装质量问题,我社负责调换)

前 言

近三十年来,随着计算机技术的飞速发展,现代计算机的计算能力得到了空前的提高,为基于量子力学的材料计算提供了坚实的基础。在计算材料科学和计算凝聚态物理中,使用最为广泛的方法是密度泛函理论(density functional theory, DFT)。密度泛函理论由美国加州大学沃尔特·科恩(Walter Kohn)教授在20世纪60年代建立,目前已经在物理、材料、化学甚至生物等领域得到了广泛的应用。科恩也因此荣获1998年的诺贝尔化学奖。

目前国内许多高校和研究所在密度泛函理论计算(通常也称为第一性原理计算)方面开展了大量科研工作,每年有许多研究生在学习相关的理论知识和计算方法。初学者往往觉得密度泛函理论计算难以入门,除了密度泛函程序比较复杂之外,更主要的原因可能是缺乏对其背后理论知识的理解。对初学者来说,密度泛函程序类似一个黑盒子,用户输入材料的结构参数,然后用程序输出计算结果。很多学生虽然学会了程序的使用流程,却往往知其然而不知其所以然。例如,在计算中最重要的参数之一是平面波截断能,如果用户不了解截断能的物理含义,就往往不知如何设置该参数。另外,即使学习了固体物理和能带理论等相关知识,但由于这些内容过于理论化,学生往往不能将它们和具体的密度泛函计算过程及计算结果联系起来。例如,尽管在固体物理教材上都会详细讲解常见的晶体结构和布里渊区,但很多学生仍然不能正确写出实际材料中原子的分数坐标和布里渊区高对称点的位置。

造成以上问题的原因可能在于,目前大部分的相关教材往往只涉及某一方面的内容。如固体物理教材虽然会涉及一些能带理论,但大多是基于简单模型的讨论,很少详细介绍密度泛函理论;密度泛函理论方面的教材则对密度泛函理论进行详细介绍,对固体物理基础和能带的基本概念却涉及很少,一般也没有实际程序讲解和算例。目前国内有关密度泛函理论的中文教材较少,而关于密度泛函程序介绍和使用实例的教材更为稀缺,初学者往往只能从网上搜索一些碎片化信息进行学习。密度泛函程序的使用手册虽然内容完整,但对初学者而言往往难以理解,而且这些手册很少详细介绍基本理论背景。本书正是基于以上现状,试图以初学者为目标,系统介绍固体物理和密度泛函的基础理论知识(如晶体结构、能带理论、密度泛函理论等),程序功能和参数(以使用最广泛的VASP程序为例),计算实例。本书从最基本的晶体元胞概念出发,一直到材料拓扑性质的第一性原理计算实例,包括了初学者所需了解的大部分内容。囿于作者能力和篇幅限制,本书只介绍一些最

基本和最重要的理论和方法,不追求对每一个部分做完整和深入的介绍,如在固体物理部分集中介绍晶体结构和对称性,但不介绍晶体结合类型、X 射线衍射、声子等其他常见内容。在密度泛函理论部分,也避免一些过于理论化和专门化的描述,如不同基组下哈密顿矩阵元的详细解析表达式、各种交换关联势的具体表达式等。在程序计算实例中,也以电子能带、态密度等基础功能为主,许多其他高级功能并未涉及,如第一性原理分子动力学、光学性质、过渡态计算等。另外本书也不涉及密度泛函理论在程序中的具体实现和程序编写。期望通过本书的学习,初学者(高年级本科生和研究生)可以初步了解密度泛函理论概况,并可以从“零”开始,直至逐步掌握程序的基本使用。

本书分为三大部分:第 1 章为背景介绍,主要介绍了密度泛函理论在材料设计中的作用和发展现状。第 2~4 章为基本理论,包括晶体结构和晶体对称性、电子能带结构、紧束缚近似和密度泛函基本理论以及基本实现。第 5~6 章为 VASP 程序基本功能和参数介绍、基本算例和材料拓扑性质基本理论及计算。

本书前四章由周健编写,第 6 章由梁奇锋编写,双方共同编写了第 5 章,并修订了全书。在本书编写过程中,南京大学王聪同学参与了部分图片的绘制工作。作者特别感谢南京大学现代工程与应用科学学院的支持。

由于作者水平有限,书中不可避免会存在一些疏漏和不足,敬请广大读者批评指正。

周 健 (南京大学)

梁奇锋 (绍兴文理学院)

2019 年 7 月

目 录

前言

第 1 章 绪论	1
1.1 材料设计简介	1
1.1.1 背景介绍	1
1.1.2 材料数据库	2
1.1.3 材料数据库的应用	5
1.1.4 存在的问题	5
1.1.5 展望和总结	7
1.2 材料计算简介	8
1.2.1 材料计算的基本内容	8
1.2.2 晶体的微观结构和宏观性质	10
1.3 高性能计算和 Linux 系统	10
1.3.1 高性能计算	10
1.3.2 Linux 基础知识	12
第 2 章 晶体结构和晶体对称性	13
2.1 常见材料的晶体结构	13
2.1.1 平移周期性	13
2.1.2 三维晶体	14
2.1.3 二维晶体	22
2.1.4 一维晶体	24
2.1.5 零维材料	25
2.2 点阵和元胞	26
2.2.1 基元、结点和点阵	26
2.2.2 元胞的取法	31
2.2.3 常见三维点阵的元胞	32
2.3 对称操作和点群	34
2.3.1 对称操作	34
2.3.2 分子和晶体中的对称性	35
2.3.3 变换矩阵	38
2.3.4 对称操作的集合	40

2.3.5	点群和空间群	41
2.3.6	点群和空间群的命名	42
2.4	晶系和点阵	44
2.4.1	七大晶系	44
2.4.2	14 种点阵	45
2.4.3	32 个点群	46
2.5	原子坐标	48
2.5.1	分数坐标和直角坐标	48
2.5.2	分数坐标和直角坐标的转换	49
2.5.3	Wyckoff 位置	50
2.6	晶体的倒易空间	51
2.6.1	倒易空间和倒易点阵	51
2.6.2	体心立方和面心立方的倒易点阵	53
2.6.3	布里渊区	54
第 3 章	电子能带结构	57
3.1	引言	57
3.2	布洛赫定理	59
3.2.1	布洛赫定理的证明	59
3.2.2	玻恩-冯·卡门边界条件	61
3.3	本征方程	63
3.3.1	本征方程的推导	63
3.3.2	能量本征值的对称性	65
3.4	紧束缚近似	68
3.4.1	紧束缚近似方法	68
3.4.2	一维聚乙炔的能带	72
3.4.3	二维石墨烯的能带	75
第 4 章	密度泛函理论	79
4.1	波函数方法	79
4.1.1	多粒子哈密顿	79
4.1.2	Hartree 方程	79
4.1.3	Hartree-Fock 方法	81
4.2	密度泛函理论基础	86
4.2.1	Thomas-Fermi-Dirac 近似	86
4.2.2	Hohenberg-Kohn 定理	87
4.2.3	Kohn-Sham 方程	88

4.3	基函数	93
4.3.1	平面波基组	93
4.3.2	数值原子轨道基组	98
4.3.3	缀加波方法	99
4.4	赝势方法	104
4.4.1	正交化平面波	104
4.4.2	赝势	105
4.4.3	模守恒赝势和超软赝势	107
4.4.4	PAW 方法	109
4.5	交换关联势	111
第 5 章	密度泛函计算程序 VASP	114
5.1	VASP 程序简介	114
5.2	四个重要输入文件	115
5.2.1	POSCAR	116
5.2.2	KPOINTS	118
5.2.3	POTCAR	120
5.2.4	INCAR	121
5.3	其他输入输出文件介绍	121
5.4	INCAR 文件介绍	123
5.5	常见功能设置	128
5.6	几个实例	130
5.6.1	非磁性材料计算 ——BaTiO ₃ 的电子结构	130
5.6.2	磁性材料计算 ——CrCl ₃ 的电子结构	134
5.6.3	杂化密度泛函计算 ——MoS ₂ 单层的带隙计算	137
5.6.4	硅的 Γ 点声子频率计算	140
5.6.5	硅的声子能带和态密度	142
第 6 章	拓扑材料计算实例	145
6.1	拓扑材料简介	145
6.1.1	拓扑量子物态	145
6.1.2	Berry 相位与拓扑物态模型	148
6.1.3	最大局域化 Wannier 函数方法	150
6.2	二维量子自旋霍尔效应体系 Bi@SiC 的电子结构计算	151
6.2.1	Bi@SiC 的晶体结构	152
6.2.2	Bi-Bi-H@Si(111) 的能带结构	153
6.2.3	Bi-Bi-H@Si(111) 的量子自旋霍尔效应	154

6.2.4	Bi-Vac-Vac@Si(111) 的量子自旋霍尔效应	156
6.3	$K_{0.5}RhO_2$ 中量子反常霍尔效应的第一性原理计算	159
6.3.1	$K_{0.5}RhO_2$ 的晶体结构	159
6.3.2	$K_{0.5}RhO_2$ 的非共面反铁磁基态	160
6.3.3	$K_{0.5}RhO_2$ 的能带	164
6.4	三维拓扑绝缘体 Bi_2Se_3 的第一性原理计算	167
6.4.1	Bi_2Se_3 的晶体结构	167
6.4.2	Bi_2Se_3 的体相能带结构	168
6.5	展望	171
参考文献		172
附录一 泛函及其导数		180
附录二 元胞和布里渊区的标准取法		181

第 1 章 绪 论

人类研究自然的传统方法有实验和理论两种，随着现代计算机的飞速发展，数值模拟越来越重要，已经成为与实验、理论相并列的第三种重要方法。在材料科学的研究中，数值模拟也扮演着重要的角色。材料数值模拟计算是材料设计的重要内容，可以辅助实验加速新材料的研发。本章概述了材料设计和材料数据库的基本概念、发展、现状及其存在的问题，同时也对材料计算所需的软硬件环境给出了简要的介绍。

1.1 材料设计简介

1.1.1 背景介绍

材料是“人类用于制造物品、器件、构件、机器或者其他产品的那些物质”，是人类赖以生存和发展的物质基础^[1]。一种新材料的出现和应用，都伴随着现代科学技术的巨大飞跃。特别是从 20 世纪初开始，新材料不断出现，如超导材料、半导体材料、新能源材料等，直接促成了第三次科技革命，极大地推动了人类社会、政治、经济和文化等领域的变革，使人类从电气时代进入了信息时代。20 世纪 70 年代，信息、材料和能源被称为当代文明的三大支柱。虽然新材料的发现有一定的偶然性，如超导材料和高温超导材料的发现都具有不可预知性，但不可否认的是，20 世纪许多新材料的出现和应用，都是基于量子力学的发展。使用量子力学理论去解决固体材料中的某些问题，逐渐产生了固体物理和能带理论这些新的衍生理论和学科。这些理论把材料的微观结构（原子类型及其空间排布方式）和宏观性质（如磁性、导电性、导热性等）联系起来。借助于现代计算机和数值计算程序，人们已经可以基本不依赖经验参数，从晶体的微观结构预测许多宏观的物理性质。例如，只需要几分钟的时间，就可以从理论上获得半导体硅的电子能带、载流子有效质量等信息。面对 21 世纪激烈的全球竞争，国家对新材料的需求更加迫切。航空航天、新能源和信息等战略性行业中对于具有超常性能材料的需求越来越强烈，而传统的以经验积累和简单循环试错的材料研发方式已显得效率低下。新材料的研发必须从以经验为主过渡到以科学设计为主。“材料设计”的概念也由此被提出：根据所需要的性能来设计材料的组分、结构和生产工艺，或者说，通过理论设计来开发具有特定性能的新材料^[2]。

材料设计包含正向和逆向两个问题,如图 1.1 所示。通过材料结构来预测性能是一个正向问题,相对比较成熟,目前有各种理论方案。例如,在原子和电子的微观层次上,可以使用量子力学、分子动力学和蒙特卡罗模拟等方法;而在介观尺度,可以使用粗粒化方法和相场方法等;在更大的尺度上,还可以使用有限元等方法。反之,通过物理性质来设计或预测材料的结构和组分却非常困难,因为这是一个反向问题,并没有直接的理论可以通过性质来推测材料的结构。

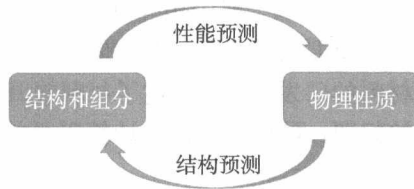


图 1.1 材料设计中的正向 (性能预测) 和反向 (结构预测) 问题

但是,对于材料性能的预测仍有助于加速新材料的研发,其基本思想是借助相对廉价和快速的数值计算,对所有已知稳定的材料进行高通量计算 (highthroughput calculation),把得到的物理性质等各种信息整理和存储,形成一个大规模的材料数据库。在此基础上,根据物理性质反向搜索可能的晶体结构,从而加速新材料的研发。美国前总统奥巴马在 2011 年率先提出“材料基因组计划”,明确指出该计划的总目标是“将先进材料的发现、开发、制造和使用的速度提高一倍”。“材料基因组计划”包括了开发高通量的理论计算方法,高通量的实验生长和测试方法,以及建立一套庞大的数据信息平台,利用这些数据,可以加速新材料的筛选和发现速度^[3]。

1.1.2 材料数据库

初看起来,要计算所有已知晶体的物理性质是一件不可能完成的任务。但随着超级计算机的飞速发展,这项任务已变得切实可行。事实上,目前国外已经建立了几个大规模的材料数据库。材料数据库的建立基于以下四个条件。

(1) 高性能计算机的飞速发展: 2018 年,全球最快的超级计算机是美国能源部橡树岭国家实验室的“Summit”,每秒钟浮点运算超过十亿亿次 (10^{17}),而我国的“天河二号”和“太湖之光”超级计算机也曾先后位列全球最快计算机之首。这为高通量材料计算提供了硬件的保证。

(2) 密度泛函理论的建立: 量子力学的基本方程是薛定谔方程,它原则上严格描述了晶体中电子的运动。但是直接使用薛定谔方程计算实际材料的性质仍然超出了当今计算机的能力范围。目前,最为成熟且使用最为广泛的方法是基于电子密度的密度泛函理论。密度泛函理论在多粒子薛定谔方程的基础上,采用了绝热近

似、单电子近似等方法,把多粒子问题等效成一个无相互作用的单电子问题,从而使得用量子力学研究实际材料变得切实可行。密度泛函理论基本不依赖于经验参数(所以也被称为第一性原理计算,或者从头计算方法),只需要晶体结构就可以计算,非常适合实际材料的研究。另外,密度泛函理论虽然不是最精确的方法,但是它在计算精度、速度和通用性方面取得了很好的平衡,从而被广泛应用于物理、材料,甚至是化学、生物医学等领域。目前,全世界每年发表的与“密度泛函理论”相关的论文接近两万篇^①。

(3) 密度泛函程序的发展和成熟:密度泛函理论虽然在 20 世纪 60 年代就提出了,但受限于当时的计算机速度,并未获得大量的应用。直到 20 世纪 80~90 年代,计算机逐渐普及,许多专业软件开始出现。事实上,现在一些流行的密度泛函程序,都是从那时开始编写的。目前,各种密度泛函理论程序层出不穷,各有特色。其中,VASP^[4]程序使用最为广泛,现有的几个著名材料数据库都是使用 VASP 作为计算程序。本书也将介绍 VASP 程序的使用,并给出相关计算实例。

(4) 无机晶体学数据库:材料数据库的基础是现有的大量天然或者人工合成的晶体材料。最为著名的晶体结构数据库为无机晶体结构数据库(Inorganic Crystal Structure Database, ICSD)^[5],它包括从 1913 年以来绝大部分已知的无机晶体结构。目前为止,ICSD 已经包含了近 20 万种晶体的结构。除此以外,还有一些其他的晶体数据库,如开放晶体学数据库(Crystallography Open Database, COD)^[6],日本 NIMS 的 Inorganic Material Database^[7]等。

基于以上条件,再通过编写程序,对大量晶体材料进行自动化的高通量计算,并最终得到材料数据库。主要的流程如图 1.2 所示:

(1) 从晶体数据库中获得晶体结构文件(通常是 cif 格式文件),并做初步的筛选。

(2) 从晶体结构自动生成密度泛函程序所需的输入文件,包括自动转换晶体结构文件,设置 k 点、截断能、收敛精度、赝势、Hubbard U 等参数。

(3) 设定所需计算的流程,如结构优化、能带和态密度计算。

(4) 向服务器递交任务,开始计算。

(5) 自动监控计算过程,在出现错误时能够及时修正或者终止计算。

(6) 在确定计算完成后处理输出文件,并从中收集所需信息。

(7) 把所有信息写入数据库。

(8) 制作用户界面,方便用户查询和使用。

目前国外已经有许多包含材料电子结构信息的材料数据库,如 Materials Project^[8], Automatic FLOW Repository (AFLOW)^[9], NREL Materials Database^[10],

^① 通过 Web of Science 搜索含有“density functional theory”关键词的论文得到的数据。

Computational Electronic Structure Database (CompES-X)^[11], Novel Materials Discovery Repository (NOMAD)^[12], Open Materials Database^[13], Materials Mine^[14], Materials Cloud^[15], Open Quantum Materials Database^[16], Computational Materials Repository^[17]。这些数据库都是免费的,其中 Materials Project 和 AFLOW 这两个数据库包含的数据最多,例如在 2018 年底,Materials Project 中包含 80 000 多种无机化合物、50 000 多个电子能带结构、20 000 多个分子、10 000 多个弹性张量、2000 多个压电张量、近 20 000 种电极材料,同时还包括 50 000 张 XANES 谱、1500 张声子谱。其他数据库并没有如此大量的数据,但也各有特色,如有的数据库特别关注二维材料,有的数据库关注材料的热力学性质等。

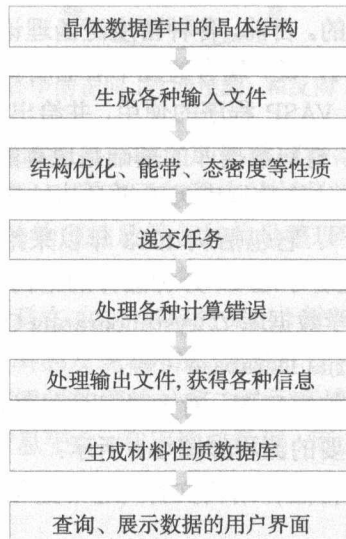


图 1.2 建立材料数据库的流程图

现有的数据库都是基于已有的材料进行计算。事实上,许多材料可能在几十年前就被合成了,但是某些性能并未被发现。例如, Bi_2Se_3 是一种常见的热电材料,已经被研究了几十年,但直到 2009 年才被证明是一种拓扑绝缘体^[18]。因此,发掘现有材料的性质仍大有可为。

另外,理论预测全新的材料结构也是可能的。现有的全局优化算法可以根据元素符号和化学配比,从零开始预测各种可能的稳定结构,甚至找到基态结构,如 CALYPSO^[19] 和 USPEX^[20] 程序都可以利用第一性原理计算对材料结构做全局优化。但这些程序和算法通常局限于少量的原子和元素,而且全局优化算法是根据材料的能量,而不能根据材料性能来预测新的结构。目前更为可行的方案是基于特定的晶体结构和化学式,通过穷举法来搜索新材料。例如,丹麦 Jóhannesson 等考虑

32 种过渡金属元素, 构造出 192 016 种可能的面心立方和体心立方结构的四元合金, 通过高通量计算它们的生成焓 (enthalpy of formation), 获得了一些已知的和全新的稳定合金材料^[21]。这也是一种材料设计的思路, 可以大大减少实验尝试的次数, 加速新材料的研发。

1.1.3 材料数据库的应用

数据库的建立虽然需要大量的工作, 但最大的挑战其实来自对数据库中数据的利用和信息的获取。一般来说, 密度泛函理论可以计算得到许多物理性质, 包括电子能带、态密度、结合能、能隙、弹性常数等。但是很多重要的材料宏观性质并不能直接从计算中得到, 如载流子迁移率、电导率和热导率、热电材料的热电优值 (ZT)、太阳能电池材料的效率等。这是因为: 一方面有的物理量原则上可以计算, 但计算量太大, 并不适合高通量计算。例如, 材料的晶格热导率可以通过计算声子的非谐效应来获得, 但是使用密度泛函理论直接计算三阶或四阶力常数非常耗时, 目前只适用于小体系的计算。另一方面, 实际材料中一般包含缺陷和杂质, 它们会影响甚至主导材料的物理性能。例如, 材料在低温下的晶格热导率主要来自缺陷对声子的散射, 而不是声子之间的散射。这种情况下, 理论计算显然没有办法预测低温下的晶格热导率, 因为不同样品的缺陷浓度通常是不同的。

因此, 寄希望于简单搜索材料数据库来研发新材料是不太现实的, 通常要具体问题具体分析。一般可以从物理上考虑, 根据所需要的性质定义一个特征量, 综合考虑数据库中材料的多种相关性质, 最后根据这个特征量的大小来判断材料性能的好坏。

例如, 利用高通量计算来搜寻高效的光伏 (photovoltaic) 材料, 但高效的光伏材料应该具有什么性质呢? 这并没有一个简单的物理量来衡量, 而是需要考虑材料的多方面性能。一般而言, 光伏材料需要有适当的能隙 (1.3 eV 左右) 和高的光吸收系数。Yu 和 Zunger 提出了一个针对光伏材料的特征量, 称为光谱极限最大效率 (spectroscopic limited maximum efficiency, SLME), 它包括材料的能隙、光吸收谱的形状以及复合损耗的大小^[22]。通过对 ICSD 数据库的搜索, 作者找到了一系列高 SLME 的材料, 其中包括目前已知的在太阳能电池中使用的光吸收材料, 如 CuInSe_2 、 CuGaSe_2 和 CuInS_2 等。同时也挖掘出一些新型光伏材料, 如 Cu_7TlS_4 、 Cu_3TlS_2 和 Cu_3TlSe_2 等, 当然 Tl 元素有毒, 不适合用于实际的太阳能电池, 但是如果能用其他无毒元素替代 Tl, 仍然有可能找到新的高效光伏材料。

1.1.4 存在的问题

虽然材料数据库已经包括了大量的材料和物理性质, 但目前来看仍然不能满足新材料研发的需求。从计算方面来看, 很多物理性质的计算量过大, 不适用于高

通量计算。同时缺陷和杂质等因素难以考虑，因此数据库的性质一般适合于晶体材料，而对于非晶体（如陶瓷和纳米材料）则无能为力。目前材料数据库的数据过于简单，对以下物理性质的计算仍存在较大问题：

(1) 材料在有限温度下的稳定性：一般密度泛函只计算零温 (0 K) 的结果，所以数据库预测合金的稳定性也局限于零温或者低温。为了获得高温的稳定性，必须考虑熵的贡献，需要计算材料的构型数、振动或者磁性等，而目前这方面的高通量计算还未涉及。

(2) 输运性质：不管是电子输运还是声子输运，都会涉及载流子寿命的计算。在电子输运的计算中，通常采用常数弛豫时间近似 (constant relaxation time approximation)，这显然是一个非常粗糙的近似。关于声子输运，可以计算三阶力常数，然后求解玻尔兹曼方程，但整个计算量非常大，并不适合高通量计算。除此以外，两者都会涉及电子/声子与杂质、缺陷或者界面的散射，这方面的计算一般只能依赖一些经验或实验参数，所以也不适用于高通量计算。

(3) 强关联材料：目前的高通量计算都是基于常规的密度泛函理论计算，一般采用局域密度近似 (LDA) 或者广义梯度近似 (GGA)。这些方法对于含有 d 电子或者 f 电子过渡金属元素的材料往往不够准确，甚至会得到完全错误的结论。现在最常见的方法是采用 LDA+ U 或者 GGA+ U 的方案，这种方案计算速度快，但结果依赖经验参数 U 的大小，即便如此，采用加 U 方法也不能完全处理所有的关联体系。目前对于强关联系统的研究有一些更好的办法，如动力学平均场理论、量子蒙特卡罗等，但是这些方法计算量非常大，目前不适用于实际材料的高通量计算。

(4) 激发态问题：常规密度泛函理论的一个重大问题是会严重低估半导体或者绝缘体材料的能隙。例如，对于半导体硅，实验能隙是 1.12 eV，而通过 GGA 计算得到的能隙只有 0.6 eV 左右。能隙的大小对于某一些应用是至关重要的，如光伏材料的研究。目前有一些方法可以很大程度上对能隙进行修正，包括杂化泛函或者 GW 方法。例如，采用 HSE06 杂化泛函可以计算得到硅的能隙为 1.17 eV，已经非常接近实验值。但是这些方法的缺点是计算量过大，并不适用于高通量计算。最近几年，一些 meta-GGA 泛函，如 mBJ 泛函，在略微增加计算量的情况下，可以很大程度上修正能隙，这为高通量计算准确预测材料能隙提供了可行的方法。

(5) 磁性问题：磁性材料通常含有 d 电子或者 f 电子，因此会有前面提到的电子关联的问题。除此以外，还有一个问题也颇为棘手，即目前的高通量计算往往只计算铁磁态，而不能自动考虑反铁磁、亚铁磁或者非共线磁性的情况。这是因为这些复杂的磁性结构通常需要扩大原始的晶体元胞，但到底如何扩大元胞则与磁结构的类型有关。对于各种具体材料，还没有很好的办法可以自动寻找各种可能的磁结构。所以磁性材料的高通量计算目前仍然有很大的局限性。

(6) 自旋轨道耦合 (spin-orbit coupling, SOC)：一些重元素组成的材料，自旋

轨道耦合效应往往是不可忽略的。但是自旋轨道耦合计算也会极大增加计算量,因此在目前的材料数据库中也鲜见考虑自旋轨道耦合作用。虽然有的软件用微扰来处理自旋轨道耦合(如 WIEN2k),可大大减少计算量,但在材料数据库中常用的 VASP 程序并不支持该方法。

(7) 大规模计算和跨尺度模拟:一般密度泛函程序的计算时间与系统尺寸的三次方成正比,目前的程序可以方便地处理元胞中有 100 个原子左右的材料。但是随着原子数增加,计算量很快会超出计算机的处理能力。一些纳米材料或者低掺杂的实际材料,原子数很容易超过几千甚至几万个。如何处理这些材料也是目前的一个挑战。一些基于局域轨道基组的密度泛函程序,可以实现计算时间随系统尺度线性增加,即线性标度 (order-N) 计算,但目前还很少用于高通量计算。另外,基于混合量子力学/分子力学 (hybrid quantum mechanics/molecular mechanics) 的多尺度模拟也是处理大系统的方法之一,该方法对系统中的重要部分(如分子的反应活性部位)采用量子力学模拟,而其余部分采用经验势场(经典力学)模拟,从而具有非常高的效率。但很显然,如何选取量子力学和分子力学的区域,并如何处理两个区域的相互作用,增加了这种方法的复杂程度。另外经验势场依赖经验参数,不能做到通用性,因此这种方法也未见于高通量计算。

除此以外,在材料数据库的建设中,还会有一些其他问题,如晶体数据库中有许多占位无序的材料(如合金材料)等,而目前常规的密度泛函程序很难处理这些体系。再如,对于许多层状材料或者分子晶体,范德瓦耳斯相互作用对晶格常数的计算十分重要。虽然现在的密度泛函程序可以考虑该相互作用,但需要额外增加计算参数,而在高通量计算中,很难自动判别哪个是层状材料,从而也无法有针对性地对其增加范德瓦耳斯相互作用修正。此外,在第一性原理计算程序中增加范德瓦耳斯相互作用有多种方案,在高通量计算中,如何选取最准确的修正方案也将是一个问题。

1.1.5 展望和总结

基于上述讨论,特别是针对目前材料数据中存在的问题,我们可以看到一些新的趋势和方向:

(1) 发展新的用于强关联系统、激发态性质、输运性质的理论方法,要求在计算速度和方便性上有大的突破,从而可以用于高通量计算。只有如此,才可以更好地发挥材料数据库的优势。

(2) 现代超级计算机广泛借助于图形处理器 (GPU) 的计算能力,一颗 GPU 的浮点运算能力至少是一个计算机中央处理器 (CPU) 的十倍以上。在美国能源部最新的超级计算机“Summit”上, GPU 提供了整个机器 95% 以上的浮点运算能力。将现有程序移植到 GPU 上是一个迫切的问题。因此,在不改变任何理论方法的情

况下,通过重新设计或者修改程序,也可以大大加快整个计算过程。事实上,这方面的工作已经取得了不少的进展,如 VASP 程序的部分功能已经可以借助 GPU 加速计算。

(3) 对数据库现有信息的挖掘,针对特定物理问题选取合适的特征量至关重要,但很多时候我们很难找到合理的特征量来帮助发现新材料。例如,对于热电材料的研究,热电优值取决于材料的 Seebeck 系数、热导率和电导率,但后两者都涉及输运计算,目前并没有高效快速而且准确的方法。当今机器学习方兴未艾,机器学习可以从大量数据中找到输入和输出之间的关联,这种关联不依赖于物理公式,而是依赖它们内在的因果关系。从物理上考虑,热电优值必然与材料的元素类型、晶体结构、电子和声子能带等有关,但并没有严格的物理公式可以直接关联它们。如果采用机器学习方法,则可能从数学上建立起两者的联系,从而避免物理上的困难。当然,机器学习需要大量的实验数据,如何获得大量材料实验的热电优值,把什么物理量作为机器学习的输入特征,依然是一个很大的挑战。但不管如何,机器学习在材料设计中可能有颠覆性的突破,而且目前已经有不少工作把机器学习应用于材料设计中。例如, Xie 和 Grossman 通过卷积神经网络来预测晶体的许多物理性质,包括结合能、能隙、费米能、弹性模量等^[23]。Balachandran 等通过机器学习和密度泛函理论来预测新的钙钛矿结构材料,并找到了 87 种可能稳定的新材料^[24]。

总体而言,材料数据库在新材料的研究和发现中已经取得了一些成果,但是目前仍未达到一个理想的状态。许多实验关心的物理性质并不能通过现有理论直接获得,或者不能快速地获得,因此发展适合高通量计算的新理论显得十分重要。另外,利用现有数据,采用机器学习的方法来预测材料性能是另外一个值得努力的方向。正如前面所讨论的,材料的结构和最终的物理性能的关系非常复杂,如果完全从物理上考虑,具有很大的难度,但是机器学习的好处是完全不依赖于物理规律,直接通过大量可调参数把两者联系起来,可能会取得意想不到的效果。

1.2 材料计算简介

1.2.1 材料计算的基本内容

对材料进行高通量计算和建设材料数据库的核心理论是密度泛函理论,而密度泛函理论只适用于晶体材料的计算。下面对材料的分类、材料计算的研究内容和适用范围做简要的介绍。

常见的材料一般都是固体,相比于液体和气体,固体原子之间有较强的相互作用,从而可以维持比较固定的体积和形状。如图 1.3 所示,固体按其中原子排列的