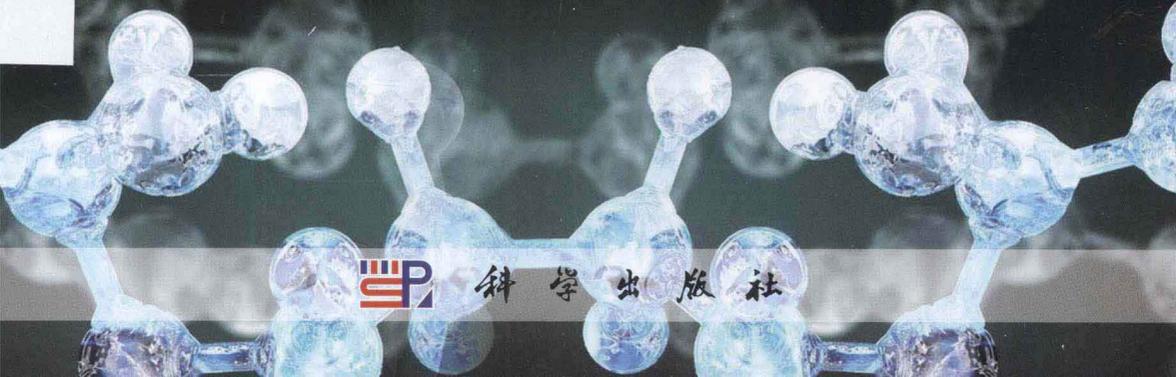




严六明 朱素华
编 著

分子动力学 模拟的理论与实践

Fenzi Donglixue
Moni de Lilun yu Shijian



科学出版社

分子动力学模拟的理论与实践

严六明 朱素华 编著

科学出版社

北京

内 容 简 介

分子动力学模拟是近年来飞速发展的一种分子模拟方法,它以经典力学、量子力学、统计力学为基础,利用计算机数值求解分子体系运动方程的方法,模拟研究分子体系的结构与性质。作为继实验和理论两种研究方法之后,研究分子体系结构与性质的第三种科学研究方法,分子动力学模拟已经被广泛应用于化学化工、材料科学与工程、物理、生物医药等科学和技术领域,起到越来越重要的作用。

全书共 11 章:第 1 章为绪言;第 2~4 章为分子的经典力学模型,包括分子的物理模型、分子间相互作用以及常用分子力场;第 5~8 章为经典分子动力学模拟与第一性原理分子动力学模拟的基本原理与方法,包括分子体系的运动方程及其数值解、分子动力学模拟的技巧、MD 模拟的统计力学基础、第一性原理分子动力学模拟;第 9 章和第 10 章分别为分子动力学模拟的应用、分子力场的构建与 MD 模拟的应用实例;第 11 章为与 MD 模拟有关的其他分子模拟方法。

本书可作为化学化工、材料科学与工程、物理、生物医药等有关专业领域的高校教师、科研人员的参考书和研究生教材。

图书在版编目(CIP)数据

分子动力学模拟的理论与实践/严六明,朱素华编著. —北京:科学出版社,2013

ISBN 978-7-03-037359-5

I. ①分… II. ①严…②朱… III. ①分子-动力学 IV. ①0313

中国版本图书馆 CIP 数据核字(2013)第 083194 号

责任编辑:王艳丽 / 责任校对:宣 慧
责任印制:刘 学 / 封面设计:殷 靓

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>

江苏省句容市排印厂印刷

科学出版社发行 各地新华书店经销

*

2013 年 5 月第 一 版 开本:B5(720×1000)

2013 年 5 月第一次印刷 印张:16

字数:312 000

定价:85.00 元

(如有印装质量问题,我社负责调换)

序

试错法是研制新材料的重要方法之一。但是,利用试错法研制新材料需要进行大量的重复性实验,不但耗时长、人力和物力消耗大,而且失败的风险也较大,因而该方法已经无法满足人类对新材料开发时效性、低成本的迫切需求。利用现代科学知识,结合高性能计算机的强大计算能力,是加快新材料的研制过程,实现新材料研制模式根本转变的有效途径。与试错法相比,将集成计算和材料设计方法用于新材料的研制过程具有巨大的价值。通过集成计算和材料设计方法初选和优化实验方案,可以减少实验尝试次数,缩短研制周期,降低研制消耗。并且,集成计算是高度自动化的过程,只需要提高计算机系统的运算速度,不但可以缩短单个计算作业的运算时间,还可以在不增加或少量增加人员的基础上同时运行更多的计算作业,大大提高新材料的研制效率。

为此,美国在 2011 年 6 月底发布了《“材料基因组计划”白皮书》(Materials Genome Initiative, MGI),目标是把新材料的研制周期从目前的 10~20 年缩短至 5~10 年,并把研制成本降低一半。我国的有关部门也高度重视新材料研制模式的转变,专门组织香山科学会议进行研讨和部署。2012 年,上海大学“085”工程内涵建设项目“集成计算与材料设计创新平台”正式立项,标志着上海大学新材料研制模式正经历根本性的转变。

利用集成计算与材料设计方法加速新材料的研制过程的核心是建立材料的组成、结构、性能之间的关系。根据材料的组成预报材料的结构和性能,或根据人们对材料性能的要求,设计符合性能要求的新材料,可以缩短新材料的研制周期,控制研制风险,降低研制成本。目前,有三类方法可用于建立材料的组成、结构、性能之间的关系:根据量子力学或密度泛函理论,利用第一性原理计算的方法研究材料的结构与性能之间的关系;根据组成材料的原子、分子的物理模型,利用计算机模拟的方法研究原子、分子的行为与材料的结构和性质之间的关系;根据以往实验所积累的数据,利用统计学方法总结材料的组成、结构、性能之间关系的统计模型。

上述三类方法各有特点,互相补充,互不替代。利用第一性原理计算,虽可以得到材料的许多基本性质,但与材料的服役行为有关的大多数性质,往往不与任何量子力学算符对应,不能利用第一性原理计算直接得到。利用统计学方法,虽可以建立材料结构与任何性质之间的统计关系,但无法揭示它们之间的本质联系。分子动力学模拟利用组成材料的原子、分子的物理模型及其运动方程,模拟组成材料的原子、分子的运动行为,不但可以揭示原子、分子的运动行为与材料结构和性质

之间的关系,还可以揭示材料的制备过程中材料结构的形成过程,对优化材料的制备工艺具有重要意义。

该书内容涵盖了分子动力学模拟的多个方面,包括分子体系的常用经典力学模型,分子动力学模拟的基本原理、方法、及其统计力学基础,分子动力学模拟的应用与结果分析。此外,该书还介绍了第一性原理分子动力学模拟,及其他有关分子模拟方法。

该书在着力进行材料研制模式转变之际付梓出版,相信对从事相关领域的研究人员和工程技术人员有所帮助。



2013年5月2日于上海大学

前 言

分子动力学模拟是一种以分子的经典力学模型,即分子力场为基础,通过数值求解分子体系的运动方程,研究分子体系的结构与性质的计算机模拟方法。得益于现代计算机软硬件技术的发展和普及,分子动力学模拟已经被广泛应用于化学化工、材料科学与工程、物理、生物医药等领域,在建立研究对象和过程的微观机理和分子模型方面发挥越来越重要的作用。

特别是近年来研究生招生规模的不断扩大和毕业要求的提高,学习和应用分子动力学模拟的研究生日益增多。但是,研究生的专业水准、数理基础和外语水平等参差不齐,他们中的许多人难以利用原始文献学习分子动力学模拟,也难以理解分子动力学模拟的理论基础。因此,本书拟从以下几个方面帮助学习和应用分子动力学模拟的读者:第一,以通俗易懂的语言介绍分子动力学模拟的基本概念和常用算法,使读者迅速掌握分子动力学模拟。第二,通过整理大量的有关分子力场的原始文献,向读者系统介绍各种分子力场的特点,以及建立和完善分子力场的方法,提高读者把实际问题转化为合适的分子力场的能力。第三,介绍分子动力学模拟的应用方法,提高读者从模拟结果获取有用信息和知识的能力。第四,全面、系统地介绍分子动力学模拟的理论基础及最新进展,引导有志于研究分子动力学模拟方法、原理、算法的读者,迅速进入分子动力学模拟研究的前沿领域。

本书努力追求通俗易懂与数理推导的严谨之间的平衡。一方面,用通俗易懂的语言全面介绍分子动力学模拟方法,并将需要的数学基础限制在生化类微积分水平,物理基础限制在经典力学和简单的统计物理学水平。另一方面,深入浅出地介绍分子动力学模拟的严谨理论,特别是分子动力学模拟的经典力学、统计力学、量子力学、数值运算基础,并将需要的数学基础扩展到线性代数和微分方程的水平,物理基础扩展到经典力学、统计力学和量子力学水平(第7章的最后两节、第8章和第11章)。

本书的第10章由朱素华同志完成,其他章节均由严六明同志完成。

本书作者感谢课题组的研究生,如纪晓波、邵长乐、张冬芳、冯庆霞、陈晋、邸素

青、谢丽青、孙超、韩帅元、苏俊铭、刘慧婷、张叶沛等同学。特别感谢陈小红女士对作者生活上的帮助和照顾,精神上的鼓励,没有她的帮助,作者难以完成本书的写作工作。作者衷心感谢科学出版社的王艳丽老师和其他同志为本书的编辑付出的辛勤劳动。

感谢国家自然科学基金(项目编号:21073118),上海市教育委员会科研创新项目(项目编号:13ZZ078)和上海市教委“085 工程”项目的资助。

由于编写时间和水平所限,书中难免存在瑕疵,恳请各位专家学者和读者批评指正。

严六明

2013 年 1 月

于上海大学

目 录

序

前言

第 1 章	绪言	1
1.1	分子动力学模拟的发展历史	1
1.2	MD 模拟的应用与意义	6
1.3	MD 模拟的发展趋势	6
第 2 章	分子的物理模型	13
2.1	分子的物理模型在化学中的作用	13
2.2	原子、分子的几何模型	15
2.3	分子的经典力学模型	19
2.4	分子的量子力学模型	29
第 3 章	分子间相互作用	30
3.1	分子间相互作用与势函数	30
3.2	分子间特殊势函数	32
3.3	分子间相互作用的起源	34
3.4	氢键相互作用	39
3.5	常用分子间相互作用势函数	40
3.6	无机巨分子物质的势函数	43
3.7	金属势	45
3.8	近程相互作用和长程相互作用	46
第 4 章	常用分子力场	48
4.1	水分子力场	48
4.2	分子力场的种类	58
4.3	MM n 系列分子力场	60
4.4	全原子分子力场	66
4.5	联合原子分子力场	70
4.6	量子化学分子力场	76
4.7	通用力场	80
第 5 章	分子体系的运动方程及其数值解	83
5.1	分子体系的运动方程 (牛顿第二定律)	83
5.2	分子体系的运动方程 (哈密顿运动方程)	83
5.3	常微分方程的数值解	85
5.4	分子体系运动方程的求解思路	87
5.5	分子体系运动方程的数值解	88
5.6	刚体运动方程的解	90
5.7	约束动力学	94

第 6 章	分子动力学模拟的技巧	96
6.1	周期性边界条件	96
6.2	势函数的计算技巧	99
6.3	长程力的计算	104
6.4	温度和压力控制技术	106
第 7 章	MD 模拟的统计力学基础	111
7.1	热力学的基本概念	111
7.2	统计系综和可实现状态	115
7.3	各态历经假设与 MD 模拟	116
7.4	体系可实现状态数的估计	118
7.5	统计系综的概率分布	118
7.6	非 Hamilton 体系的统计理论	121
7.7	演化算符与差分格式	127
第 8 章	第一性原理分子动力学模拟	135
8.1	经典 MD 模拟的局限性	135
8.2	BOMD 和 CPMD 模拟方法	135
8.3	平面波基函数	142
8.4	赝势	147
8.5	PIMD 模拟	152
8.6	质子在甲醇分子和水分子之间迁移过程的 AIMD 模拟	152
第 9 章	分子动力学模拟的应用	157
9.1	MD 模拟的计算机软硬件	157
9.2	模拟体系分子模型的建立	160
9.3	MD 模拟的初始条件	163
9.4	MD 模拟技术参数的确定	164
9.5	MD 模拟的过程	164
9.6	MD 模拟结果的处理	165
第 10 章	分子力场的构建与 MD 模拟的应用实例	178
10.1	聚炔分子力场的构建	178
10.2	锂离子电池电解液的 MD 模拟	182
第 11 章	与 MD 模拟有关的其他分子模拟方法	199
11.1	非平衡分子动力学模拟	199
11.2	Brown 动力学模拟	207
11.3	耗散粒子动力学	217
11.4	Monte Carlo 模拟	223
参考文献	233
附录	术语约定	242
索引	244

第 1 章 绪 言

1.1 分子动力学模拟的发展历史

1.1.1 分子动力学模拟的概念

分子模拟 (molecular modeling 或 molecular simulation) 是一类通过计算机模拟来研究分子或分子体系结构与性质的重要研究方法, 包括分子力学 (molecular mechanics, MM)、Monte Carlo (MC) 模拟、分子动力学 (molecular dynamics, MD) 模拟等。这些方法均以分子或分子体系的经典力学模型为基础, 或通过优化单个分子总能量的方法得到分子的稳定构型 (MM); 或通过反复采样分子体系位形空间并计算其总能量的方法, 得到体系的最可几构型与热力学平衡性质 (MC); 或通过数值求解分子体系经典力学运动方程的方法得到体系的相轨迹, 并统计体系的结构特征与性质 (MD)。目前, 得益于分子模拟理论、方法及计算机技术的发展, 分子模拟已经成为继实验与理论手段之后, 从分子水平了解和认识世界的第三种手段。

1.1.2 MD 模拟的早期历史

最早的 MD 模拟在 1957 年就已实现。当时, Alder 和 Wainwright 通过计算机模拟的方法, 研究了从 32 个到 500 个刚性小球分子系统的运动。模拟开始时, 这些小球分子被置于有序分布的格点上, 具有大小相同的速度, 但速度方向随机分布。除相互间的完全弹性碰撞外, 刚性小球分子之间没有任何相互作用, 小球分子在碰撞间隙做匀速直线运动。在经过一段时间的模拟, 系统中的刚性小球分子速度达到 Maxwell-Boltzmann 分布后, 他们分别根据维力定理和径向分布函数计算了系统的压力, 发现两种方法得到的结果一致^[1]。1959 年, 他们提出可以把 MD 模拟方法推广到更复杂的具有方阱势的分子体系, 模拟研究分子体系的结构和性质^[2]。

1964 年, Rahman 模拟研究了具有 Lennard-Jones 势函数的 864 个 Ar 原子体系, 得到了与状态方程有关的性质、径向分布函数、速度自相关函数、均方位

移等^[3]。此后,分子模拟工作者广泛模拟研究了具有不同势函数参数的 Lennard-Jones 模型分子体系,得到了体系的结构及其各种热力学性质,探讨了 Lennard-Jones 势函数参数对体系结构与性质的影响,建立了 Lennard-Jones 势函数参数与模型分子体系结构及性质之间的关系。

1.1.3 分子体系运动方程的数值解

在刚性小球分子体系中,除碰撞瞬间外分子间没有任何相互作用,分子的运动轨迹由一系列的折线组成。因此,对刚性小球分子体系的 MD 模拟,算法的核心是计算刚性小球分子间的碰撞时间及其碰撞前后的运动方向和速度的变化,而不是直接数值求解刚性小球分子体系的运动方程。相反,对于 Lennard-Jones 分子,分子间一直存在相互作用,分子的运动轨迹复杂,必须通过求解体系的运动方程来确定,这就促进了运动方程数值方法的发展。

单原子分子没有内部结构,计算量小,容易实现 MD 模拟。相反,多原子分子具有复杂的内部结构,运动方程更加复杂,计算量大,难以实现 MD 模拟。因此,直到 20 世纪 70 年代,才实现水分子体系^[4,5]、正烷烃分子体系^[6]等多原子分子体系的 MD 模拟。

对于多原子分子体系,如果包括化学键伸缩势、键角弯曲势等所有分子内的相互作用势都由势函数描述,不存在所谓的约束,则体系运动方程的数值解与单原子分子相同,没有本质区别。常用的 MD 数值积分方法包括 Verlet 算法^[7,8]、预测-校正算法及其由 Verlet 算法衍生的蛙跳法^[9]、速度 Verlet 法^[10]等。但是,如果分子中的化学键等被约束在固定长度或整个分子被约束成刚体,则体系的运动方程将完全不同。根据经典力学,没有内部自由度的多原子分子可以被近似为刚体,运动状态可以分解为质心的平动和刚体的转动两种独立运动模式。其中,质心的平动与质点力学没有任何区别,不需要另外讨论。刚体的转动通常由 Euler 角随时间的演化描述,可以通过数值求解相应的运动方程得到。但是,以 Euler 角为变量描述刚体运动时,牵涉奇点问题,算法不稳定。为此,Evans 提出了以四元数描述刚体运动状态的方法,很好地克服了奇点问题,成为常用的 MD 模拟算法^[11,12]。

在室温附近,大多数原子、分子处于基态,分子中的化学键长和键角均在平衡位置附近以很小的幅度振动。因此,假设化学键长和键角在 MD 模拟过程中固定不变,不会对 MD 模拟结果产生显著影响。但是,除键长和键角以外的其他内部运动自由度,如二面角的旋转,仍被允许,分子不能被近似为质点或刚体。为了利用 MD 模拟研究具有固定键长的约束体系,可以采用 SHAKE 算法^[6]、RATTLE 算法^[13]等。约束体系的 MD 模拟,是 MD 模拟不可缺少的重要

方法, 至今仍是非常热门的研究课题^[14-16]。

目前, MD 模拟技术已经成熟, 不但可以模拟简单的不具内部自由度的单原子分子和刚性多原子分子, 也可以模拟具有内部自由度的多原子分子。即使对于蛋白质、DNA 这样复杂的生物大分子的模拟, 也已经没有任何算法上的限制, 只有计算机计算能力的限制。

1.1.4 温度与压力的调控与统计系综的实现

早期 MD 模拟体系被限制在一定的空间之内, 具有固定的体积。此外, 根据能量守恒定律, 模拟体系的总能量也被固定, 不允许任何波动。从统计力学角度分析, 这样的 MD 模拟体系是微正则系综或 NVE 系综。但许多实际的化学、生物等体系, 常与一恒温热源热接触, 具有固定温度, 但不具有固定的总能量, 这就是正则系综或 NVT 系综。此外, 大多数实际或人工体系, 除具有固定温度外, 也具有固定的压力, 但具有可变的总能量和体积, 这是 NPT 系综。

最简单的调控体系温度的方法是变标度 (scaling) 恒温法。当体系的温度, 即总动能偏离设定值时, 体系中所有原子或分子的速度被乘以称为标度因子的系数, 使体系的总动能回归设定动能^[17]。之后, Berendsen 改进了变标度恒温法, 使温度逐渐被调节到设定温度, 减小了温度的波动幅度^[18]。此外, Andersen 还提出了热浴法, 在模拟过程中让体系中的一个或若干个原子或分子与恒温热源中的分子发生随机碰撞, 达到调整体系总体温度并使之保持恒定的目的^[19]。但是, 由这样的方法模拟得到的相轨迹不连续, 不符合实际情况。

上述温度调控算法粗糙, 没有严格的理论基础。具有严格理论基础的温度调控算法是恒温扩展法 (extended system), 通过在模拟体系广义坐标和广义动量外引入额外的自由度与热浴耦合的方法, 达到调控温度的目的。理论上, 恒温扩展法可以通过改变或扩展实际模拟体系的哈密顿函数或拉格朗日函数实现, 如 Nosé 恒温扩展法^[20,21]及其经 Hoover 修正的 Nosé-Hoover 恒温算法等^[22]。

除了调控温度技术外, 压力的调控也是 MD 模拟的基本方法。与变标度恒温法相似, 最直接的压力调控算法是通过调整模拟体系的体积, 达到间接调控体系压力的目的。但是, 如果在各原子的位置坐标直接乘以同一系数, 将改变体系中分子内原子间距或键长, 得到错误的结果。因此, 调控压力的算法与变标度恒温法的实现方法不同, 需通过复杂的坐标变换, 在不改变分子内原子间相对距离的条件下改变分子的相对位置, 实现改变模拟体系体积的目的, 这就是标度变换恒压法。同样, 压力的调控也可通过扩展系统自由度, 采用恒压扩展法实现, 如 Andersen 恒压恒焓系综^[19]等。

扩展法的提出是 MD 理论和方法研究的一个里程碑。目前, 通过扩展模拟

体系自由度，在体系的哈密顿函数或拉格朗日函数上添加额外的项，达到调控体系的温度与压力的方法，仍然是一活跃的研究领域。

1.1.5 MD 模拟的理论发展与稳定算法

在 MD 模拟发展早期，研究重点是 MD 模拟算法与应用，但对 MD 模拟的理论基础重视不够^[23]。

早期的 MD 数值积分算法，重点是计算效率和精度，忽略了差分过程中误差的积累及其对模拟结果的影响。由于在 MD 模拟过程中误差的积累，可能导致差分算法的失稳，甚至严重影响模拟结果的可靠性。在各种差分算法中，辛算法的差分方法被认为是目前最稳定、高效的计算方法，适用于哈密顿体系。哈密顿体系的辛算法及其完整的理论框架由我国数学家冯康提出，是对世界科学的不朽贡献^[24]。

基于严格的经典力学 Liouville 方程和扩展哈密顿函数或拉格朗日函数的 MD 模拟理论，把温度和压力调控方法、差分算法等纳入统一的理论框架，为现代 MD 模拟的发展奠定了严格的理论基础^[25-27]。目前，满足辛对称性和时间反演可逆性的差分算法的设计，仍是 MD 模拟领域的重要研究内容。

1.1.6 分子力场的发展

最早的 MD 模拟对象是刚性小球分子体系。刚性小球分子体系具有非常特殊的势函数形式，当两个小球间的距离小于它们的半径和时，它们之间的相互作用势为无穷大；相反，当两个小球之间的距离大于它们的半径和时，它们之间没有任何相互作用，相互作用势为零；当两个小球之间的距离等于它们的半径和时，它们之间的相互作用势不连续。这样的相互作用势虽然代表了物质具有一定的体积、不能无限压缩这一最显著的性质；但是，由于刚性小球分子间没有任何吸引作用，刚性小球分子不能凝结为液体，也不存在气-液相变。对刚性小球模型的改进包括方阱势等，方阱势具有吸引力，分子可以凝结为液体。但是，方阱势函数不连续，与实际分子仍有较大的差异。

用于 MD 模拟的第一个连续势函数是 Lennard-Jones 势函数。在 MD 模拟发展的早期，曾广泛研究了 Lennard-Jones 分子体系的结构、状态方程、相变、热力学性质等。这些 MD 模拟结果与统计热力学理论计算结果的对比，对促进统计热力学的发展，具有重要的意义。同时，通过 MD 模拟研究 Lennard-Jones 分子，优化了稀有气体的 Lennard-Jones 势参数，加深了对稀有气体分子间相互作用势的认识。

分子间相互作用势函数的发展经历了从刚性小球模型、方阱势模型等不连续势函数模型，到 Lennard-Jones 相互作用势函数的过程。考虑到分子间相互作用的本质，分子间相互作用还应包括库仑相互作用、偶极或多极矩相互作用等。由于水是地球生态系统最重要的物质之一，对水的模拟一直吸引着研究者广泛的兴趣。在模拟研究水分子的过程中提出了大量不同种类的势函数，不但加深了对水分子间相互作用的认识，也丰富了分子间相互作用与物质性质之间的关系的认识^[28,29]。

除分子间相互作用外，多原子分子体系的分子力场还包括分子内相互作用。一般地，分子内相互作用包括键伸缩势、键角弯曲势、绕单键旋转势（或二面角扭曲势）、四点离面势等成键相互作用和分子内非键相互作用。分子内非键相互作用与分子间相互作用相同，但由于成键相互作用势远大于非键相互作用势，一般不计算具有成键相互作用的原子对间的非键相互作用，称为排除非键相互作用。而 1-4 原子的成键相互作用较弱，与非键相互作用处在同一数量级，有时只排除部分非键相互作用。早期的分子力场，得益于分子力学的发展，如 MM n 系列分子力场等^[30-32]。在 MD 模拟发展中起重要作用的分子力场，包括 AMBER 力场^[33]、CHARMM 力场^[34]、OPLS 力场^[35]等。

除无机或有机分子外，熔融盐、金属和合金、半导体、硅酸盐等也是 MD 模拟的重点研究对象，有关势函数有 Tosi-Fumi 势^[36-38]、金属势^[39-42]、半导体势^[43,44]、硅酸盐势^[45-49]等。所有这些完善了不同类型分子体系的分子力场模型，丰富了对分子内和分子间相互作用的了解，成为认识物质世界的一种新方法。

1.1.7 AIMD 模拟

虽然经典 MD 模拟已经被广泛应用于化学、生物、材料、物理等领域，但经典 MD 模拟以分子力场模型为基础，在模拟中必须输入力场模型是其不足之处。同时，大多数经典力场模型只能描述基态分子，无法描述远离平衡状态的化学反应过程等。基于量子力学理论的第一性原理分子动力学 (*ab initio* molecular dynamics, AIMD) 模拟，通过量子化学方法直接计算所有的分子内和分子间相互作用，不需要输入经验的力场模型，对于研究化学反应等具有重要的价值。目前，AIMD 模拟常被用于研究催化^[50]、质子传递^[51]、燃烧等过程。

AIMD 可分为 BOMD 和 CPMD 两种类型。其中，在 BOMD 模拟过程中，模拟的每步都需要利用 Schrödinger 方程或密度泛函理论计算体系的基态波函数，然后计算各原子的受力，并根据经典力学原理实现原子核位置的演化。相反，在 CPMD 模拟过程中，不需要每步都计算体系的基态波函数，而是令波函数与原子核坐标一样，按一定的规律演化，得到的不是体系的基态波函数^[52]。

但是, BOMD 和 CPMD 两种方法得到的原子轨迹没有显著区别^[53]。

除 BOMD 和 CPMD 外, AIMD 的第三种形式是 PIMD 模拟。与前两种 AIMD 不同, PIMD 不但利用量子化学方法计算原子间的相互作用, 还利用量子力学方法计算原子位置的演化, 可以计算 H 原子等的量子效应, 对研究质子在化学反应中的隧道效应等具有重要意义^[54]。

1.1.8 力的计算与节省时间算法

分子间和分子内相互作用力的计算, 是 MD 模拟中最耗时的部分, 也是最难并行化运算的部分。在 MD 模拟的发展过程中, 发展了 Lennard-Jones 势函数的截断与修正算法、库仑相互作用算法、Verlet 近邻列表算法、格子索引算法 (cell index method) 等, 对 MD 模拟的发展具有重要意义。

1.2 MD 模拟的应用与意义

MD 模拟是一种研究分子体系结构与性质的重要方法, 已被广泛用于化学化工、生物医药、材料科学与工程、物理等学科领域。MD 模拟最直接的研究结果是分子体系的结构特征, 包括溶液中的配位结构, 生物和合成高分子的构型与形貌, 生物和合成高分子与溶剂分子或其他小分子配体之间的相互作用, 分子在固体表面的吸附与分布, 分子在重力场、电磁场等外场中的取向与分布等。

除了分子体系的结构特征, MD 模拟方法还可以研究分子体系的各种热力学性质, 包括体系的动能、势能、焓、吉布斯自由能和亥姆霍兹自由能、热容等。通过 MD 模拟, 还可以得到与体系的状态方程有关的密度、压强、体积、温度等之间的关系。根据体系的能量和自由能, 还可以直接或间接地研究体系的相变与相平衡性质等。

此外, 利用 MD 模拟可以研究分子体系的速度自相关函数、速度互相关函数、均方位移等性质, 并由此计算体系的自扩散系数、互扩散系数、黏度系数等各种迁移性质。利用非平衡 MD 模拟, 还可以研究各种热力学流与热力学力之间的关系, 得到 Onsager 意义上的唯象系数。

最后, 利用反应性分子力场 MD 模拟或 AIMD 模拟, 还能得到化学键的断裂和生成等与化学反应有关的性质。

1.3 MD 模拟的发展趋势

总的来说, MD 模拟的发展趋势是以更高的效率、模拟更大的体系、实现更

长的演化时间、取得更精确的模拟结果为目的。为了实现这些目标，必须从计算技术、MD 模拟算法、分子模型等多方面进行广泛而深入的研究。

1.3.1 计算技术的发展方向

在经历了约半个世纪的指数式提高，计算机核心部件 CPU 的主频在 21 世纪初超过 3GHz 后，出现了停滞现象，失去了过去那种按 Moore 定律快速提高的趋势。但是，CPU 的制造技术并没有达到发展极限，出现了双核、四核、八核甚至十六核等多核 CPU。因此，Moore 定律继续有效，只是发展模式从不断提高 CPU 的主频，转化为提高单片 CPU 上集成的核芯数量。与 CPU 主频被不断提高的时代相比，这种新趋势对算法的开发和软件设计提出了新的挑战。

在 CPU 主频被不断提高的时代，一个因速度缓慢而性能不佳的计算程序，只要等待新一代具有更高主频的 CPU 的出现，就会有更出色的表现。现在，同样因速度缓慢而表现不佳的计算程序，在新一代主频几乎不变、但具有更多核芯的 CPU 上，其表现不一定会得到改善。事实上，为了改进计算程序的运算速度，必须改进程序的算法，提高其并行运算速度。不过，提高计算程序的并行运算速度，不是简单的工作，而是复杂的工程，必须发展适合并行运算的算法^[55]。

衡量一个算法并行运算效果的指标是加速比 (speedup)。当利用多个核芯进行并行运算时，一般只有算法的一部分能被并行加速，其他部分则不能被并行加速。因此，当用 N_p 个核芯进行并行运算时，运算时间一般不会缩短到单个核芯串行运算时间的 $1/N_p$ 。因此，并行运算的加速比，就是利用单个核芯进行串行运算所消耗的计算时间与利用多个核芯进行并行计算所消耗的计算时间之比。也就是说，如果一算法在单个核芯上的运算时间为 T_1 ，在 N_p 个核芯上的运算时间为 T_{N_p} ，则算法的加速比为 $S_{N_p} = T_1/T_{N_p}$ 。根据 Amdahl 定律，如果一个算法中能够被任意并行加速部分所占计算量为 α ，不能被并行加速部分所占计算量为 $1 - \alpha$ ，则利用 N_p 个核芯进行并行计算时的加速比为 $S_{N_p} = (1 - \alpha + \alpha/N_p)^{-1}$ 。当利用任意多个核芯进行并行计算时，得到算法的最大加速比 $S_{\max} = (1 - \alpha)^{-1}$ 。如果算法可以被并行加速部分的比例 α 未知，可以利用 N_p 个核芯并行运算时的实测加速比 S_{N_p} 估计， $\alpha = (S_{N_p}^{-1} - 1)/(N_p^{-1} - 1)$ 。与加速比相关的另一指标是并行计算的效率，定义为加速比与核芯数之比 $E_{N_p} = S_{N_p}/N_p$ 。此外，在进行并行运算时，各个进程的调度、进程之间的通信等，都需要消耗额外的时间，更降低了算法的加速比。

例如，某一作业，当用 1 个核芯进行计算时所消耗的计算时间为 100，用 2 个核芯进行计算时所消耗的计算时间为 60，则并行计算的加速比为 1.667，算法中可以被并行加速部分所占比例 $\alpha = 0.8$ 。当用 5 个核芯进行计算时，并行计算

加速比为 2.778，此时的并行效率为 55.6%。事实上，这个作业的最大理论加速比为 5，并行效果并不理想。

并行计算与三个和尚从山下往位于山顶的寺庙运水的故事相似。当方丈觉得一个僧人运水的速度太慢时，他有两种选择：增加人力，多派一些僧人运水；或训练运水的僧人，提高僧人的运水效率。前者相当于并行计算，通过利用更多的计算核芯，提高作业效率；后者相当于提高计算核芯的主频，通过提高单个计算核芯的运算速度来提高作业效率。方丈可以通过计算加速比，即在没有采取任何措施以前往山上运一桶水需要消耗的时间，与采取改进措施后运一桶水需要消耗的时间之比，来评估两种方案的实际效果。方丈肯定认为，由于受生理条件的限制，提高一个僧人运水的加速比肯定有限，更好的方案还是多派僧人运水。并且，如果派出 N_p 个僧人运水，可以将加速比提高到 N_p 。但是，故事的结局已经熟知，当方丈派出两个僧人运水时，运水作业的加速比没有增加；当方丈派出三个僧人运水后，加速比甚至降低为 0。

如果方丈了解并行计算技术，他在派出更多的运水僧人前会考察从山下往山上运水的环境是否适合并行作业，或者说他的算法是否具有可扩展性 (scalability)。曾经参观故事中寺庙的游客就会发现，僧人从山下往山上运水的山道，只能容纳一个人上下山，不适合并行作业，不具有可扩展性。因此，当方丈派出三个僧人一起进行运水作业时，山道阻塞，加速比降为 0。懂得并行计算技术的济公，考察了从山下往山上运水的山道，发现山道环境条件不适合并行作业，作业不具有可扩展性。如果不顾作业环境而进行并行作业，还会影响上山敬佛的香客，不利于吸引游客上山，发展旅游经济。因此，济公改变了算法，让僧人从井里往上提水。从井里往上提水的作业环境具有更大的空间，可以容纳多人并行作业，具有可扩展性。

虽然现代超级计算机可以模拟多达上千亿个原子，实现纳秒级的演化时间。例如，浮点运算峰值速度达 1×10^{15} 次/s 的超级计算机运行一天，可以实现的模拟量达到 $NT = 2.14$ 原子·秒 ($N = 2.14 \times 10^6$ 个原子，实现演化时间 $T = 1 \times 10^{-6}$ s)^[56]。但是，大多数 MD 模拟工作者，难以得到这样的超级计算机的计算服务，只能使用约每秒万亿次的中小型集群式计算系统。目前，除传统的 CPU 计算系统外，MD 模拟工作者的另一选项是 GPU (graphical processing units) 计算系统。GPU 计算系统的主要特点是并行性能优越，性能价格比远高于 CPU 计算系统。利用 GPU 计算系统，可以以小型集群式计算系统的成本，得到大型计算系统的浮点运算速度。例如，NVIDIA 的 M2090 GPU 运算卡包含 16 个多处理器，每个多处理器又包含 32 个计算核芯，总共多达 512 个计算核芯。该 GPU 运算卡的单精度浮点峰值运算速度达到每秒 1.331 万亿次以上，价格约 2 万元。因此，GPU 计算系统正吸引越来越多的 MD 模拟工作者的使用^[57-59]。