arXiv:2303.08169v1 [cs.DC] 14 Mar 2023

Allegro-Legato：通过尖锐度感知最小化实现可扩展、快速且稳健的神经网络量子分子动力学

Hikaru Ibayashi[1],Taufeq Mohammed Razakh[1], Liqiu Yang[1], Thomas Linker[1], Marco Olguin[2],Shinnosuke Hattori[3], Ye Luo[4], Rajiv K. Kalia[1], Aiichiro Nakano[1], Ken-ichi Nomura[1], 和 Priya Vashishta[1]

[1] 南加州大学高级计算与模拟协作中心，美国加利福尼亚州洛杉矶，邮编 90089[2] 南加州大学高级研究计算中心，美国加利福尼亚州洛杉矶，邮编 90089[3] 索尼集团公司研发中心高级研究实验室，日本神奈川县厚木市旭町 4-14-1，邮编 243-0014[4] 阿贡国家实验室阿贡领导级计算设施，美国伊利诺伊州莱蒙特，邮编 60439

摘要。 基于机器学习的神经网络量子分子动力学（NNQMD）模拟正在革新材料的原子级模拟，既能提供量子力学精度，又能将计算速度提升数个数量级，2020 年 ACM 戈登·贝尔奖 [11]及 2021 年决赛入围奖 [17]便是明证。基于群论构建、具备旋转等变性和局部描述符的最先进 NNQMD 模型，准确度和速度都远超此前模型，因此被命名为 Allegro（意为“快速”）。然而，在大规模并行超级计算机上，该模型面临保真度缩放问题：随着原子间作用力预测中非物理结果的数量增加，涉及更多原子、持续更长时间的模拟将难以进行。在此，我们将 Allegro 模型与尖锐度感知最小化（SAM）相结合，通过改善损失景观的平滑性来增强模型的鲁棒性，从而解决这一问题。由此得到的 Allegro-Legato（意为“快速且平滑”）模型能够延长失效时间t_{failure}，同时不牺牲计算速度或准确度。具体而言，与最先进的 Allegro 模型 $\left(t_{\mathrm{failure}} \propto N^{-0.29}\right)$、 *i.e.*相比，Allegro-Legato 的失效时间对问题规模的依赖性弱得多 $t_{\mathrm{failure}} \propto N^{-0.14}$ （N 为原子数），并能系统性地延缓失效，因此可开展规模更大、时长更长且不会失效的 NNQMD 模拟。该模型在阿贡领导力计算设施的 Polaris 超级计算机上还展现出出色的计算可扩展性和图形处理器加速能力。这类兼具可扩展性、准确度、速度和鲁棒性的 NNQMD 模型，有望广泛应用于新兴百亿亿次浮点运算计算机上的 NNQMD 模拟；其中一个具体应用是考虑氨分子动力学中的核量子效应，为推动可持续发展的绿色氨技术奠定基础。

这项工作发表于国际超级计算大会 2023

关键词： 分子动力学 · 等变神经网络 · 尖锐度感知最小化。

1 引言

基于机器学习的神经网络量子分子动力学（NNQMD）模拟，通过以大幅降低的计算成本，按照量子力学精度追踪所有原子的轨迹，正在革新材料的原子级建模 [2]。NNQMD 不仅能准确预测原子间作用力，还能捕捉电子极化 [13]和电子激发 [14], 等量子性质，这也正是 NNQMD 中“Q”的含义。NNQMD 是当前高性能超级计算机上最具可扩展性的科学应用之一，获奖者获得了 2020 [11] 年的 ACM 戈登·贝尔奖，并入围了 2021 [17] 年的决赛，便是明证。NNQMD 领域近期的一项突破是显著提高了力预测的准确度 [1]，超越了以往模型；这一进展得益于基于张量场群论表述的旋转等变神经网络 [25]。如今，最先进（SOTA）的准确度与创纪录的速度相结合，最新的 NNQMD 模型 Allegro（意为“快速”）便是基于空间局域描述符实现了这一点 [16]。

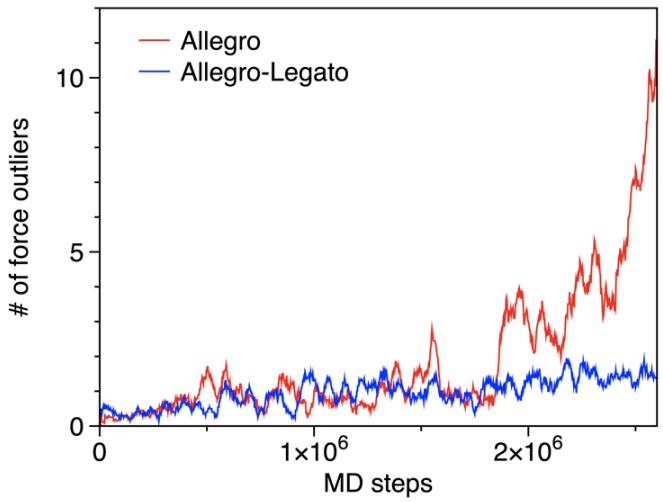

图1：NNQMD 模拟过程中原子力推断的离群值数量：随着模拟的推进，原始Allegro模型预测出的非物理超大力值（超过 5σ）数量不断增加，导致原子动力学变得不稳定，最终在经过 2.6×10^6个分子动力学步后模拟失败（红色）。相比之下，所提出的模型（Allegro-Legato）维持了近乎恒定的离群值数量，使模拟保持稳定（蓝色）。

尽管具有卓越的计算可扩展性，大规模并行 NNQMD 模拟仍面临一个尚未解决的重大问题，即保真度缩放 [20]。在大规模 NNQMD 模拟中，微小的预测误差可能不断累积，导致产生非物理的原子力，从而使原子轨迹的准确度随时间降低。这些力的离群值甚至可能导致模拟意外终止（图 1）。随着模拟在空间尺度上不断扩大、在时间上不断延长，非物理力预测的数量预计也会成比例增加，这可能严重限制 NNQMD 模拟在新一代百亿亿次级超级计算平台上的保真度，尤其是对最令人期待的远离平衡态应用而言 [14,15]。

本文借鉴了机器学习领域近期的一项进展，解决了保真度-可扩展性问题。解决这一问题需要提升 NNQMD 模型的鲁棒性，即当模拟轨迹遇到训练数据集之外的原子构型时，减少非物理力预测离群值的数量。研究发现，尖锐度感知最小化（SAM） [7] 能够增强神经网络模型的鲁棒性——这是一种训练算法，可在优化模型训练损失的同时，对模型的尖锐度（即损失曲面的曲率）进行正则化。因此，我们采用 SAM 训练快速的 Allegro 模型，使其损失景观更加平滑，从而提升模型的鲁棒性。结果表明，所得 Allegro-Legato（意为快速且“平滑”）模型能够延长失效时间 t_{failure}，即在微正则系综下，NNQMD 模拟能够运行的分子动力学步数，同时保持相同的推理速度和几乎相同的准确度。具体而言，与最先进的 Allegro 模型 $\left(t_{\mathrm{failure}} \propto N^{-0.29}\right)$相比，Allegro-Legato 的失效时间对问题规模 $t_{\mathrm{failure}} \propto N^{-0.14}$ （N 为原子数）的依赖性弱得多，因此能够在不发生失效的情况下进行规模更大、时长更长的 NNQMD 模拟。除了这一主要贡献外，我们还发现，NNQMD 模型的保真度-可扩展性与模型尖锐度的相关性，比与模型参数数量的相关性更强。 [5]

快速且稳健的 Allegro-Legato 模型已在我们可扩展的并行神经网络量子分子动力学代码 RXMD-NN 中实现。在 Argonne 领导计算设施（ALCF）的 Polaris 超级计算机上，我们使用 480 个计算节点实现了 0.91 的弱扩展并行效率；每个节点配备一颗 AMD EPYC 中央处理器（中央处理器）和四块 NVIDIA A100 图形处理器（图形处理器）。此外，与 Polaris 上单颗 32 核中央处理器相比，该代码使用四块图形处理器实现了 7.6 倍的单节点性能加速。

Allegro-Legato 能够实现远超其他方法的时空尺度更大的神经网络量子分子动力学模拟。与常用于对分子构型进行“有效”长时间采样的热浴分子动力学模拟不同（例如，蛋白质折叠），这类模拟会扰乱动力学轨迹；Allegro-Legato 则能实现“真实”的长时间哈密顿动力学，可直接与高分辨率光谱实验中观测到的精细振动模式进行比较。具体而言，我们现在能够满足基于路径积分分子动力学研究氨的动力学时核算细微核量子效应这一极高的计算需求——

[5] 代码可在 github.com/ibayashi-hikaru/allegro-legato 获取

这对于解开橡树岭国家实验室近期一项高分辨率中子散射实验观测中的谜团至关重要。最先进的中子实验与领先水平的 NNQMD 模拟相结合，为以绿色氨为基础的燃料技术奠定了基础，助力实现可持续社会。

2 方法创新

本节首先概述：(1) NNQMD 模拟方法及最先进的 Allegro 模型；(2) 用于稳健神经网络模型训练的 SAM。随后，我们介绍 SAM 增强的 Allegro 模型 Allegro-Legato 的关键方法创新，并进一步介绍其可扩展的并行实现。

2.1 神经网络量子分子动力学概述

分子动力学（MD）模拟追踪原子位置随时间的演化$\{\mathbf{r}_i \mid i = 1, \ldots\ldots, N\}$（即，轨迹），其中原子为 N，

$$m_i \frac{d^2}{dt^2}\mathbf{r}_i = \mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} E\left(\{\mathbf{r}_i\}\right) \tag{1}$$

其中，m_i 和$\mathbf{f}_i$ 分别为第 i 个原子的质量及作用于其上的力，而E 是由量子力学（QM）决定的原子间势能。在 NNQMD 中，神经网络经过训练，以复现一组原子构型 $\{\{\mathbf{r}_i\}_t \mid t = 1, \ldots\ldots, N_{\text{training}}\}$（$N_{\text{training}}$ 为训练构型数量）[2,13,14,11,17]对应的真实量子力学值 $E(\{\mathbf{r}_i\}_t)$。在最先进的 Allegro 模型中，能量 E 由有限截断距离内原子对（i, j）之间的成对嵌入能 E_{ij}构成，以保持数据局部性 [16]。Allegro 能够实现高准确度，关键在于所有能量项对于旋转、反演和平移都满足群论等变性，即相对于欧几里得群 $E(3)$ [1,25]具有等变性。这是通过使用张量及其不可约表示，将能量表示为最高秩为 ℓ 的张量和张量积来实现的。简言之，Allegro 通过群论等变性实现高准确度，并通过数据局部性提升计算速度。

2.2 尖锐度感知最小化概述

神经网络通过最小化损失函数 $L(\mathbf{w})$ 进行训练，其中 $\mathbf{w}$ 表示神经网络的权重参数。优化方法的选择在机器学习中至关重要，因为它会影响收敛速度和泛化性能等多个因素 [22]。尤其是，易受对抗攻击影响是神经网络特有的问题 [9]，在计算机视觉 [24]和自然语言处理 [11]等领域已得到广泛研究。近期研究表明，神经网络量子分子动力学（NNQMD）中的保真度-可扩展性也可以看作是对“大规模模拟期间对抗

攻击”的鲁棒性 [23,5], : 在大规模模拟期间，原子轨迹会受到累积的非物理预测的“攻击”，即在长时间、大规模模拟过程中遭遇“对抗扰动”。因此，我们自然会预期，用于应对对抗攻击的优化方法能够提升神经网络量子分子动力学（NNQMD）的保真度-可扩展性。尖锐度感知最小化（SAM）是计算机视觉领域提出的一种增强鲁棒性的方法 [7]。SAM 的核心在于最小化模型的“尖锐度”，其定义为

$$\max_{\|\epsilon\|_2 \leq \rho} \{L(\mathbf{w}+\epsilon) - L(\mathbf{w})\}, \tag{2}$$

其中， ρ （邻域大小）是用于定义尖锐度的超参数。虽然无法直接计算尖锐度，但已有研究表明，可以通过以下更新规则实现对 $L(\mathbf{w}) + \max_{\|\epsilon\|_2 \leq \rho}\{L(\mathbf{w}+\epsilon) - L(\mathbf{w})\}$（训练损失与尖锐度之和）的最小化：

$$\mathbf{w} = \mathbf{w} - \eta \nabla_{\mathbf{w}'} L(\mathbf{w}') \big|_{\mathbf{w}'=\mathbf{w}+\rho\frac{\nabla_{\mathbf{w}} L(\mathbf{w})}{\|\nabla_{\mathbf{w}} L(\mathbf{w})\|}} \quad (\eta : \text{learning rate}), \tag{3}$$

该方法利用一阶导数，即 也就是说， $\nabla_{\mathbf{w}} L(\mathbf{w})$。这样便无需计算开销巨大的二阶导数，也能优化尖锐度。

2.3 关键创新：Allegro-Legato：SAM增强的Allegro

如上所述，我们的假设是，SAM平滑损失景观能够提升神经网络量子分子动力学的保真度缩放。为定量检验这一假设，我们将SAM融入Allegro神经网络量子分子动力学模型 [16], 的训练中，该模型兼具最先进的准确率和计算速度。我们将由此得到的SAM增强Allegro模型称为Allegro-Legato。 [6]

为了确定合适的尖锐度正则化强度，我们调节 SAM 的超参数 ρ，以获得最稳健的模型。结果发现，在我们的设置中， $\rho = 0.005$ 能带来最长的 t_{failure}。在小规模模拟测试中，我们使用了 LAMMPS，这是一款广泛使用的开源分子动力学模拟软件（lammps.org）。详细的训练设置见第 4.5 节。

ρ	0.0	0.001	0.0025	**0.005**	0.01	0.025	0.05
$t_{\text{失效}}$	4020	4030	6420	**8480**	4760	4210	3780

表 1：**SAM 强度 ρ 与失效时间 t_{failure}：** 我们通过在 0.001 到 0.05 的范围内进行网格搜索来调整 ρ。在小规模模拟（$N = 432$）中， $\rho = 0.005$ 的模型取得了最大的 t_{failure}。

[6] 在音乐中，Legato 意为音符之间平滑衔接，没有突然的断开。

2.4 RXMD-NN：Allegro-Legato NNQMD 的可扩展并行实现

为测试大规模计算与保真度缩放，我们在 RXMD-NN 软件 [14,20],中实现了所提出的 Allegro-Legato NNQMD 模型。RXMD-NN 是我们开发的可扩展并行反应性分子动力学软件 RXMD [18]的扩展版本。RXMD-NN 采用分层分而治之方案，实现“全局可扩展、局部快速”（或“全局稀疏、局部稠密”）的并行化 [19]：(1) 全局可扩展的空间分解，最适用于大规模并行计算平台；(2) 高效的局部链表分解及后续邻居列表构建，从而达到 $O(N)$ 计算复杂度。进程间通信通过消息传递接口（MPI）库的非阻塞应用程序编程接口（API）实现，通信模式采用无锁设计，并将节点间数据交换降至最低。尽管这是大规模分子动力学应用中最广泛采用的策略之一，但它尤其适合 NNQMD 算法利用现代高性能计算（HPC）架构的优势。在这种架构中，少数性能强大的图形处理器卡通过加速计算密集型内核承担主要计算任务，同时由众多核中央处理器并行执行随机内存访问和乱序数据处理。在 RXMD-NN 中，中央处理器负责并行构建邻接表。构建完成的邻接表与原子位置及类型信息一同转换为 PyTorch张量对象，以便在图形处理器上进行力推断。RXMD-NN 支持控制计算粒度，例如每个域中的原子数和每个节点中的域数，从而找到水平扩展性与垂直扩展性之间的理想平衡，充分利用可用硬件资源。

由于 PyTorch 提供了用于复杂模型架构的 API，能够在图形处理器上高效训练和推理，因此它已成为机器学习社区的标准 Python 库。然而，高性能计算集群、移动设备和边缘节点等生产平台通常需要一系列 Python 并非为之设计的功能，例如多线程、低延迟计算以及大规模并行分布式架构。Allegro 模型的图形处理器卸载通过 TorchScript 实现。TorchScript 是一种静态类型中间表示，可用于创建经过序列化且可优化的机器学习模型。序列化后的模型可以从其他编程语言（例如 C++）加载，从而部署在 Python 代码难以运行的环境中，同时不牺牲多线程和优化机会。

3 结果

我们在阿贡领导级计算设施（ALCF）的领导级规模计算平台 Polaris 上，测试了所提出的 Allegro-Legato 神经网络量子分子动力学模型在 RXMD-NN 代码中的实现的保真度和计算可扩展性。

3.1 实验平台

我们在 ALCF 的 Polaris超级计算机上进行了数值实验。该系统基于慧与科技（HPE）Apollo 6500 Gen $10+$ 构建，每个机箱包含两个计算节点，每个机架有七个机箱，共有 40 个机架，总计 560 个节点。每个 Polaris 节点配备一颗 2.8GHz AMD EPYC Milan 7543P 32 核中央处理器和 512 GB DDR4 内存、四个每个配有 40GB HBM2 内存的 NVIDIA A100 图形处理器、两块组成 RAID0 的 1.6 TB 固态硬盘，以及两个 Slingshot 网络端点。Polaris 使用 NVIDIA A 100 HGX 平台，通过 NVLink 连接全部 4 个图形处理器，图形处理器互连带宽为 600 GB/s。由 Cray 设计的 Slingshot 互连基于高基数 64 端口交换机，按蜻蜓拓扑排列；它通过为应用分配流量类别，提供自适应路由、拥塞控制和带宽保障。Polaris 的生产峰值性能为 44 千万亿次浮点运算，节点级双精度性能为 78 万亿次浮点运算。

3.2 保真度缩放结果

在保真度缩放测试中，我们使用 $\ell = 1$训练了 Allegro 和 Allegro-Legato，并从 t_{failure}方面考察了它们的鲁棒性，即 t_{failure}越大，鲁棒性越强。测试所用的分子动力学模拟参数经过精心选择，以确保每次分子动力学模拟都能在合理的时间内失败，但不会立即失败。基于 Nose-Hoover 恒温器的恒温系综方法（即，NVT 系综）用于研究热平衡性质，但它可能通过外部恒温器将原子相连，从而抑制并掩盖模型的不合理预测。微正则系综（NVE）方法通过在没有外部控制的情况下直接积分运动方程，对模型鲁棒性进行最严格的测试（同时也更适用于非平衡过程）。在每次模拟中，首先使用 NVT 系综将液氨系统在 200 K 下热化 1,000 个时间步。随后，我们将系综切换为 NVE，并继续模拟，直至其失败，以确定 t_{failure}。在整个鲁棒性测试中，时间步长 Δt 均设为 2 飞秒（fs）。对于每种系统尺寸，我们对十次以上独立模拟的结果取平均，以测量 t_{failure}。

为量化保真度缩放，我们通过以下缩放关系定义保真度缩放指数 $N^{-\beta}$：

$$t_{\text{failure}} = \alpha N^{-\beta}, \tag{4}$$

其中 α 是一个前因子。β 值越小（即，保真度缩放越弱），失效时间就越长，因此能够在大规模并行计算机上准确研究时空尺度更大的过程。Allegro-Legato 模型显著改善了保真度缩放，超出统计不确定性的幅度达到 $\beta_{\text{Allegro-Legato}} = 0.14 < \beta_{\text{Allegro}}$ =0。29（见图 2 中的误差条），从而系统性地延长了失效时间。

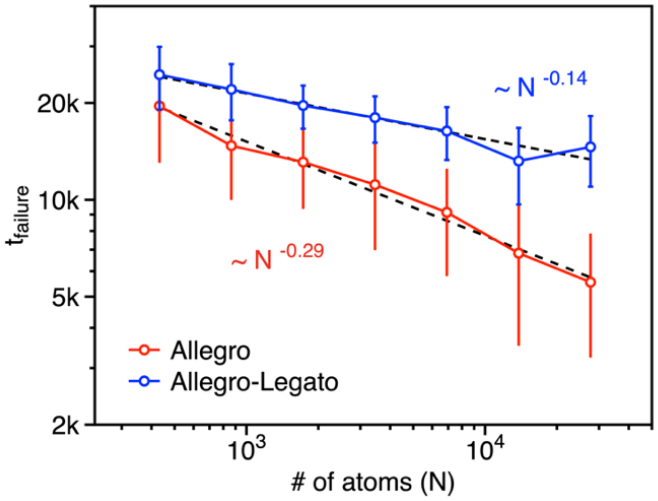

图 2：**神经网络量子分子动力学模拟的保真度缩放：** 此处，t_{failure} 是在采用 2fs 时间步长的 NVE 系综下测得的。即使在最小的系统规模下，也能观察到统计上有所改善的 t_{failure}；随着系统规模增大，这一现象更加显著。幂律拟合的指数表明，使用 Allegro-Legato 模型后，该指数几乎降低了一半。

3.3 计算扩展性结果

我们测量了工作负载扩展时每个分子动力学步的墙钟时间，测试系统为含 6,912 P 个原子的氨系统，运行于 P 个分子动力学域上。在该测试中，每个分子动力学域包含 6,912 个原子，并将其卸载到单个图形处理器上。除了力推断，执行时间还包括邻接表构建、主机与图形处理器内存之间的数据传输，以及通过网络结构进行的节点间通信。图 3 展示了墙钟时间随 P 变化的情况。随着问题规模与图形处理器数量线性增长，运行时间仅略有增加，表明其具有出色的可扩展性。

此处，我们将 NNQMD 算法的速度定义为原子总数乘以每秒执行的分子动力学步数，并据此量化并行效率。等粒度加速比是指在 P个分子动力学域上的速度相对于单个域基准速度的比值。因此，弱可扩展性的并行效率等于等粒度加速比除以 P。当每个域的粒度为 6,912 个原子时，我们在 1,920 个 A100 GPU 上实现了出色的弱可扩展性效率：对于多达 13,271,040 个原子，效率达到 0.91。尽管每个域的粒度相对较大，为 6,912 个原子，我们仍实现了较短的求解时间，每个分子动力学步仅需 3.46 秒，从而支持生产运行每天执行 25,000 个分子动力学步。

图 4 展示了单个 Polaris 节点上 NNQMD 算法的 GPU 加速效果。直方图显示了每个分子动力学步的墙钟时间缩短量（取 10 个分子动力学步数的平均值，基

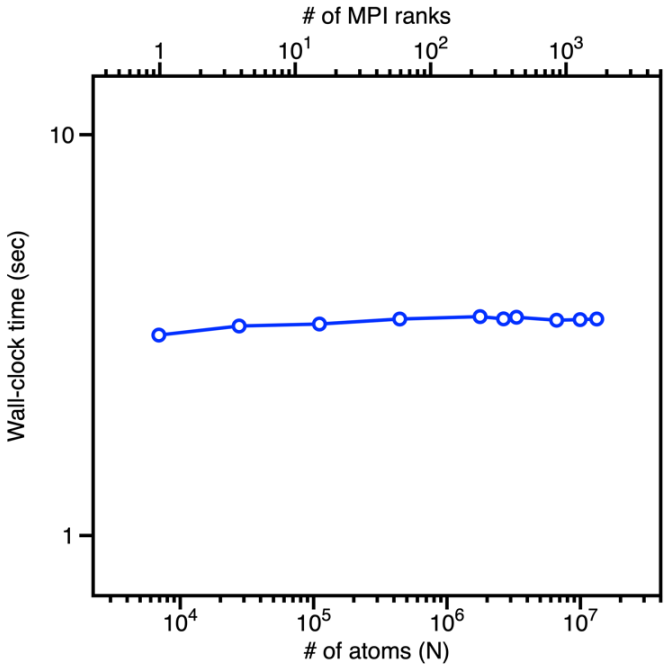

图 3：在采用不同规模的工作负载（6、912 P 个原子）模拟液氨时，RXMD-NN 代码每个分子动力学步的墙钟时间，使用 P 个 A100 GPU（P =1、.....、1,920）。

准为仅使用中央处理器时的运行时间（32 个核心、32 个线程））。我们考察了：(1) 三种系统规模：N =1、728、6,912 和 13,824 个氨原子；以及 (2) 三种域分解方式，即单个、双倍和四倍子域。每个域中的原子分配给一个图形处理器。对于 N =1、728 原子的系统，我们观察到图形处理器加速效果有限，加速比最高为 1.24 倍；系统规模增大后，加速效果显著提升。对于包含 N =13、824 个原子的系统，采用四个子域时，我们实现了 7.6 倍的加速比。

4 讨论

如上一节所示，SAM 增强的 Allegro 模型 Allegro-Legato 相较于最先进的 Allegro 模型具有更强的鲁棒性。本节将讨论 SAM 训练对准确度和计算速度等其他方面的影响。

4.1 模拟时间

首先，SAM 不会影响分子动力学模拟时间，因为 SAM 仅应用于分子动力学模拟的训练阶段，而不应用于推理阶段。表2比较了

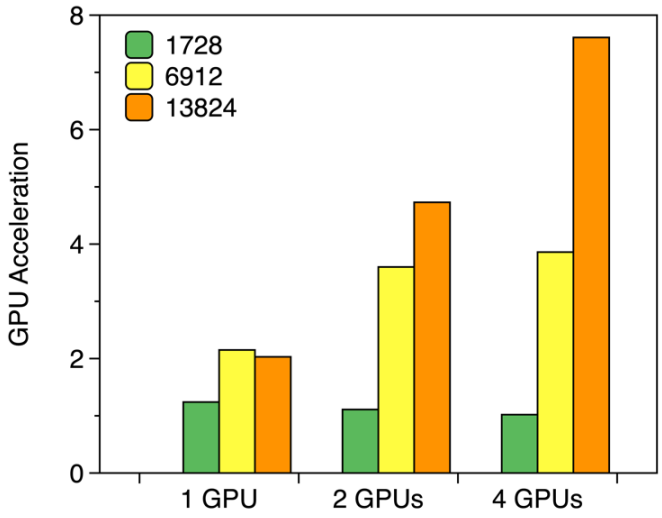

图4：**NNQMD 算法的 GPU 加速：**我们考察了三种系统规模：$N = 1728$、6912 和 13,824 个原子。直方图展示了每个分子动力学步的墙钟时间相较于以不使用 GPU、采用 32 个中央处理器核心运行时为基准的缩短幅度。基准测试平台以及 GPU 和 CPU 架构的详细信息见正文。使用四个 GPU 处理 N =13,824 个原子时，我们实现了 7.6 倍的加速比。

基准 Allegro模型和所提出的 Allegro-Legato 模型每个分子动力学时间步长的模拟时间。下文中，我们采用默认值 $\ell = 1$ 作为最大张量秩，因此两个模型的参数数量相同。由于实验平台并非专用，测量不确定性范围内两个模型的模拟时间相同。

作为对比，表2还展示了具有另外两种张量秩（ $\ell = 0$ 和 2）的基线 Allegro模型。较大的 ℓ 会生成更准确但规模更大的模型（即参数数量更多），因此模拟时间也更长。根据准确度与计算成本之间的权衡，使用 Allegro模型进行生产级神经网络量子分子动力学模拟时通常采用 $\ell = 1$。

4.2 训练时间

如第 2.2 节所述，SAM 的缺点是比基础优化器需要更多计算时间，因为每个训练轮次都必须计算两次一阶梯度。然而，在我们的设置中，SAM 比默认优化器收敛更快，因此总训练时间没有受到显著影响（表3）。作为参考，我们还测量了最大张量秩分别为 $\ell = 0$ 和 2 的 Allegro 模型的训练时间，并观察到较大的 ℓ 会使训练成本大幅增加。总之，Allegro-Legato 提升了 Allegro 的鲁棒性，同时没有增加训练成本。

模型	参数数量	每步时间（毫秒）
Allegro	133,544	916
Allegro-Legato	133,544	898
参考模型		
Allegro ($\ell = 0$)	95,656	395
Allegro ($\ell = 2$)	183,720	2,580

表2：**模拟时间比较：**由于 SAM 仅应用于训练阶段，且不会改变架构的规模，因此不会影响模拟的计算成本。

模型	总时间（小时）	每轮时间（秒）	训练轮数
Allegro	11.1	248	161
Allegro-Legato	13.6	433	113
参考模型			
Allegro ($\ell = 0$)	4.4	127	127
Allegro ($\ell = 2$)	19.6	636	111

表3：**训练时间比较：**尽管 SAM 每轮训练所需时间较长，但其收敛更快，因此对总训练时间的影响并不显著。与 Allegro 模型各种变体的参考训练时间相比，Allegro-Legato 增加的训练成本可以忽略不计。

4.3 模型准确率

虽然准确复现系统能量对于正确指导模型训练至关重要，但对于分子动力学模拟而言，最关键的是准确预测力。使用 Allegro-Legato（$\ell = 1$）模型得到的原子力验证误差为 15.9（均方根误差，RMSE）和 11.6（平均绝对误差，MAE）；使用原始 Allegro 模型（$\ell = 1$）得到的误差分别为 14.7（RMSE）和 10.7（MAE）。所有误差值的单位均为毫电子伏/$\mathring{A}$。Chmiela 等人最近提出，可靠进行分子动力学模拟所需的 MAE 应为 1 千卡/摩尔/$\mathring{A}$，相当于 43。4毫电子伏/$\mathring{A}$ [4]。与原始 Allegro 模型相比，Allegro-Legato 的力预测误差略有增加（在液氨数据集中约为 8%），但所得力误差约为可靠进行分子动力学模拟所需误差上限的四分之一。也就是说，Allegro-Legato 提升了鲁棒性，同时没有牺牲准确度。

4.4 Allegro 中的隐式尖锐度正则化

虽然我们提出显式控制模型的尖锐度，但我们发现，基线 Allegro 模型中的一个控制参数（即，用于表示特征的张量的最大秩）会隐式调节模型的尖锐度。在表 4 中，除了我们的 Allegro-Legato 模型具有较低的尖锐度外，Allegro $\ell = 1$，2 模型的尖锐度也显著低于 Allegro $\ell = 0$ 模型，且 t_{failure} 更高。也就是说，具有更高 ℓ 的 Allegro 会隐式地对尖锐度进行正则化，从而提高鲁棒性（即，t_{failure} 更大），但计算成本也会增加。Allegro-Legato（$\ell = 1$）模型以更短的计算时间达到了与 Allegro（$\ell = 2$）模型相同的尖锐度水平；见表 2 和表 3。

模型	Allegro ($\ell = 0$)	Allegro ($\ell = 1$)	Allegro ($\ell = 2$)	Allegro-Legato ($\ell = 1$)
尖锐度	5.0×10^{-4}	3.2×10^{-4}	9.8×10^{-5}	1.2×10^{-4}

表 4：**Allegro 的隐式尖锐度正则化：**尽管我们的 Allegro-Legato 模型的尖锐度低于 Allegro，但在 Allegro 模型中，随着 ℓ 增大，尖锐度会逐渐降低。这里，我们通过在每个极小值周围的 0.05 邻域内独立随机采样 1,000 次，取最大值来测量尖锐度：$\max_{\|\epsilon\|_2 \leq \rho} \{L(\mathbf{w} + \epsilon) - L(\mathbf{w})\}$。

图 5 展示了 Allegro（ℓ=0、1 和 2）和 Allegro-Legato（$\ell = 1$）模型的损失曲面。图中证实：（1）在 Allegro 模型中，随着 ℓ 增大，曲面逐渐变得平滑（即尖锐度降低），这是通过准确度实现的隐式正则化所致，但计算成本也随之增加；（2）在 ℓ 相同的情况下，SAM 可使 Allegro-Legato 相较于 Allegro 实现显式平滑，且不会增加额外的计算成本。

4.5 训练细节

最后，为了完整起见，我们提供了详细的训练配置（表 5）。为确保公平比较，我们使用了作为最先进模型发布的默认超参数，而 SAM 训练则使用默认优化器作为其基础优化器。

5 应用

所提出的 Allegro-Legato 模型在保持 Allegro 最先进的准确率和计算速度的同时，提升了鲁棒性，使其能够在领导级计算机上进行大规模时空尺度的神经网络量子分子动力学模拟。一个颇具说服力的例子是研究氨的振动性质。开发能够准确再现分子

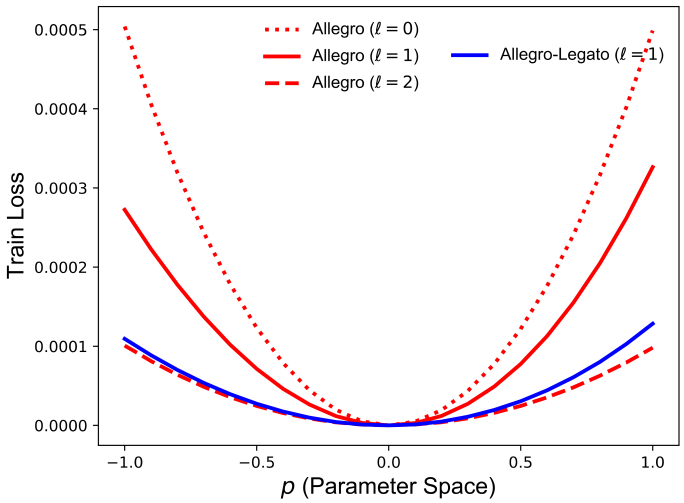

图 5：**损失曲面可视化：**各模型损失曲面的一维可视化。根据尖锐度的定义（公式 2），我们随机采样一个向量 $\mathbf{d}$，该向量给出尖锐度方向，以计算 $L(\mathbf{w} + p\mathbf{d})$，用于$p \in [-1, 1]$。

晶体和液体振动光谱的动力学模型，对于预测其热力学行为至关重要，而这对它们在能源、生物和制药系统中的应用十分关键 [10]。特别是，过去几年中，为实现可持续社会而开发绿色氨基燃料技术的趋势日益显著。氨（NH_3）的能量密度甚至高于液氢，但与液氢相比，氨的储存所需能耗低得多，储存温度为 -33℃，而液氢为 -253℃；此外，得益于氨在农业中的百年应用历史，现已建成庞大的氨基础设施 [3]。每年生产的氨超过 1.8 亿公吨，且有 120 个港口配备了氨码头 [3]。氨相关技术的发展，将取决于我们理解和建模那些产生其独特性质的复杂物理和化学相互作用的能力。

在开发能够准确描述振动性质的计算框架时，需要仔细考虑多个复杂因素，例如核量子效应（NQEs）及其与振动非谐性的耦合 [10]。用于研究振动性质的标准第一性原理计算只对电子进行量子力学处理；振动性质可通过对晶胞 Hessian 矩阵进行傅里叶变换和矩阵对角化来确定，其规模至多只有几百个元素 [26]。评估 NQEs 及其与振动非谐性的耦合作用，通常采用所谓的路径积分分子动力学（PIMD）方法，该方法对整个量子系统的量子配分函数进行采样 [6,21]。这需要对大量副本进行长时间模拟，这些副本属于大型分子动力学系统，并且通过谐性相互作用耦合成

材料类型	液态 NH_3
每个构型中的原子数	432
训练样本数（N_{training}）	4,500
验证样本数量	500
$\mathbf{r}_{\max}$ 对于截断	6.0
最大张量秩 (ℓ)	1
批量大小	4
峰值学习率	$2e-3$
学习率衰减	ReduceLROnPlateau
学习率调度器耐心值	50
学习率调度器因子	0.5
（基础）优化器	Adam
Adam 的（β_1，β_2）	(0.9, 0.999)
损失函数	每原子均方误差
损失系数（力、总能量）	(1。0，1。0)
停止准则	$\Delta L_{\text{验证}} \leq 3e-3$ 经过 100 训练轮数

表 5：**详细训练设置：**除非另有说明，本文所有训练方案均采用这些参数。

彼此作用的环状聚合物，尤其是在低温下 [6,21]。图6a背景展示了典型的第一性原理模拟：原子按经典方式处理，电子电荷密度则按量子力学方式处理，以计算原子力；图中以蓝色等值面示意电子电荷密度。在前景中，我们突出显示了同一原子构型的 PIMD模拟中的一个 NH_3分子，其中每个原子都有 32 个副本，并通过谐性相互作用彼此耦合。副本模拟的计算具有极高的并行性，只需进行固定的最近邻副本通信；主要计算开销在于计算每个副本模拟中各原子的能源和力，这通常通过第一性原理方法完成。然而，我们的 Allegro-Legato 模型具有更强的鲁棒性，能够以接近量子精度的准确度稳定地进行长时间分子动力学模拟，因此可以替代 PIMD模拟中耗费巨大的第一性原理计算；否则，要准确评估氨的低能分子间振动模将难以实现。

我们利用 Allegro-Legato 模型进行了大规模并行 PIMD模拟，在每个副本模拟中计算能源和力，以评估氨的分子间模式的声子谱。Allegro-

Legato模型经证实能够产生预期的有限温度下高能模式软化现象，并纳入核量子效应；与谐近似下的标准矩阵对角化结果相比，如图6b所示。具体而言，30 – 90毫电子伏范围内振动模式的能源降低，与作者去年夏天在橡树岭国家实验室开展的高端中子实验所得振动谱结果一致（这些结果将在其他地方发表）。

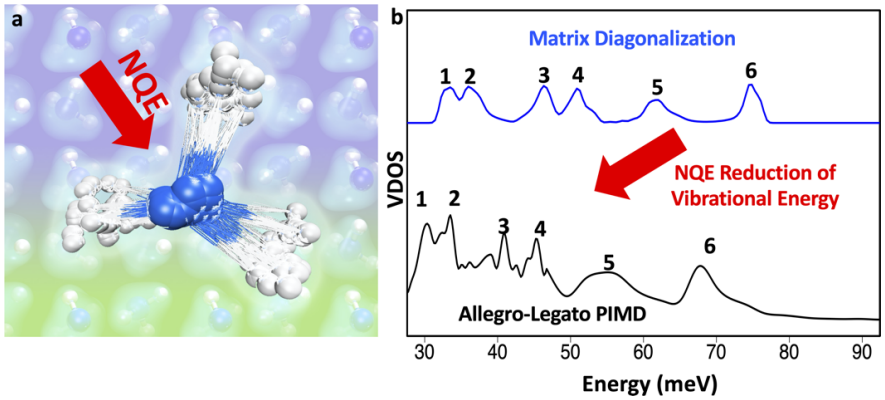

图6：**氨的计算振动光谱：**(a) 典型的第一性原理模拟将原子视为经典粒子，将电子视为量子力学对象；而 PIMD模拟则使用每个原子的多个副本来模拟核量子效应（NQE）。(b) 上方曲线显示的是在零温度下、不考虑 NQE 时计算得到的振动谱；下方曲线则显示了在有限温度下通过 Allegro-Legato PIMD模拟得到的振动谱。纳入 NQE 后，Allegro-Legato PIMD 能正确展现有限温度下预期出现的高能分子间模式软化现象，并解释高能端的中子散射观测结果。

6 相关工作

NNQMD 模拟及其可扩展并行实现 [2,13,14,1,16] [11,17]的发展与应用呈爆发式增长。另一方面，直到最近，研究人员才开始用失效时间 t_{failure}[8] 来量化 NNQMD 的鲁棒性，并指出其失效时间会随着问题规模增大而不断缩短（即保真度缩放问题） [20]。本研究首次：（1）通过引入保真度缩放指数 β ，正式量化保真度缩放，其中 $t_{\mathrm{failure}} \propto N^{-\beta}$ （N 为原子数）；（2）提出利用尖锐度感知最小化解决保真度缩放问题。

抵御对抗攻击的鲁棒性是机器学习领域的核心议题，受到广泛研究 [11,9,24]。与典型的对抗攻击相比，为 NNQMD 生成对抗性扰动并非易事。这是因为我们考虑的攻击不仅针对模型的准确度，还针对模型的失效时间（t_{failure}）；而失效时间只能通过长时间模拟来确定 [23,5]。生成对抗网络（GAN）是在在线学习场景中采样分子构型的一种可行方法 [12]。然而，我们要指出，分子动力学模拟的真正优势在于能够计算动态关联，从而直接解释高分辨率光谱实验；这需要一条长时间不间断的哈密顿轨迹，而对抗网络通常不适用于此。因此，在这一领域，Allegro-Legato 提供了独特的解决方案。

7 结论

我们将所提出的、基于 SAM 的保真度缩放问题解决方案引入 Allegro神经网络量子分子动力学模型 [16] 中，该模型代表了最先进的准确度和速度。由此得到的 Allegro-Legato 模型通过显著降低指数 $\beta_{\mathrm{Allegro\text{-}Legato}}$=0.14< $\beta_{\mathrm{Allegro}} = 0.29$, 大幅改善了保真度缩放，从而系统性地延长了失效时间。改善保真度缩放对于确保从领导级并行计算机上的大规模模拟中提取有意义的科学知识至关重要。我们对 Allegro-Legato 的可扩展并行实现具有出色的计算扩展性和 GPU 加速能力，兼具准确度、速度、鲁棒性和可扩展性，因此能够在百亿亿次级计算平台上，针对具有挑战性的应用开展实用的大规模时空尺度神经网络量子分子动力学模拟。

8 致谢

本研究作为美国能源部科学办公室基础能源科学资助的计算材料科学计划的一部分开展，资助编号为 DE-SC0014607。H.I. 和 K.N. 获得了美国国家科学基金会项目 OAC-2118061 的部分资助。模拟在阿贡领导级计算设施通过美国能源部 INCITE 计划完成，可扩展代码的开发则由 Aurora ESP计划支持。作者感谢南加州大学高级研究计算中心提供计算资源，这些资源为本文报告的研究成果作出了贡献。我们感谢 Makiko Hirata博士就 Allegro-Legato 提供了宝贵的讨论。

参考文献

1. Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J.P., Kornbluth, M., Molinari, N., Smidt, T.E., Kozinsky, B.：用于

数据高效且准确的原子间势。《自然·通讯》**13**, 2453 2021 10 1038 41467 022 29939 ( ). https://doi.org/ . /s - - -5 2. Behler, J.：高维神经网络势的构建：教程式综述。《国际量子化学杂志》**115**(16), 1032–1050 (2015). https://doi.org/ 10 1002 24890 < > 000357606000005 . /qua. , 3 GotoISI ://WOS: . Chehade, G., Dincer, I.：绿色氨生产作为潜在无碳燃料的进展。《燃料》**299**, 120845 (2021). https://doi.org/10.1016/j.fuel.2021.120845, https://www.sciencedirect.com/science/article/pii/S 4 0016236121007225 .

Chmiela, S., Sauceda, H.E., Muller, K.R., Tkatchenko, A.：借助机器学习力场实现精确分子动力学模拟。《自然·通讯》**9**(1), 3887 (2018). https://doi.org/10.1038/s41467-018-06169-2, https: //doi.org/10 2 5 1038 41467 018 06169 . /s - - - . Cubuk, E.D., Schoenholz, S.S.：物理模型的对抗力。NeurIPS机器学习与物理科学研讨会论文集 (2020) 6. Feynman, R.P., Hibbs, A.R.：《量子力学与路径积分》。McGraw-Hill，纽约，NY (1965) 7. Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B.：尖锐度感知最小化：高效提升泛化能力。国际学习表征会议论文集，ICLR p. 1839 (2021). https://doi.org/10 01412 8 48550 2010 . /arXiv. . . Fu, X., Wu, Z., Wang, W., Xie, T., Keten, S., Gomez-Bombarelli, R., Jaakkola, T.：仅有力还不够：通过分子模拟对机器学习力场进行基准测试和批判性评估。arXiv p. 2210.07237 (2022). https://doi.org/10 07237 9 48550 2210 . /arXiv. . . Goodfellow, I.J., Shlens, J., Szegedy, C.：解释并利用对抗样本。国际学习表征会议论文集，ICLR (2015 6572 10 10 48550 1412). https://doi.org/ . /arXiv. . . Hoja, J., Reilly, A.M., Tkatchenko, A.：分子晶体的第一性原理建模：结构与稳定性、温度与压力。WIREs计算分子科学**7**(1), e1294 (2017). https://doi.org/10.1002/wcms.1294, https://doi.org/10 1294 11 1002. /wcms. . Jia, W., Wang, H., Chen, M., Lu, D., Lin, L., Car, R., E, W., Zhang, L.：借助机器学习将具有从头算精度的分子动力学规模推至一亿个原子。超级计算会议论文集 p. 5 (ACM/IEEE, 2020). https://doi.org/10 3433707 12 5555 3433701 . / . . Jinnouchi, R., Miwa, K., Karsai, F., Kresse, G., Asahi, R.：大规模原子模拟中原子间势的在线主动学习。《物理化学快报》**11**(17), 6946–6955 (2020). https://doi.org/10.1021/acs.jpclett.0c01061, https://doi.org/10.1021/acs. jpclett.0 01061 13 c . Krishnamoorthy, A., Nomura, K., Baradwaj, N., Shimamura, K., Rajak, P., Mishra, A., Fukushima, S., Shimojo, F., Kalia, R., Nakano, A., Vashishta, P.：利用神经网络量子分子动力学测定液态水的介电常数。《物理评论快报》**126**(21), 216403 (2021). https://doi.org/10.1103/PhysRevLett.126.216403, <GotoISI>:// WOS:000655929800008 14. Linker, T., Nomura, K., Aditya, A., Fukshima, S., Kalia, R.K., Krishnamoor- thy, A., Nakano, A., Rajak, P., Shimmura, K., Shimojo, F., Vashishta, P.：利用激发态神经网络量子分子动力学，探索铁电氧化物中远离平衡态的超快极化控制。《科学进展》**8**(12), eabk2625 (2022). https://doi.org/10.1126/sciadv.abk2625

15. Misawa, M., Fukushima, S., Koura, A., Shimamura, K., Shimojo, F., Tiwari, S.C., Nomura, K., Kalia, R.K., Nakano, A., Vashishta, P.：将基于第一性原理的人工神经网络势应用于固体材料的多尺度冲击动力学模拟。《物理化学快报》**11**, 4536–4541 (2020). https://doi.org/10 00637 16 1021 0 . /acs.jpclett. c . Musaelian, A., Batzner, S., ohansson, A., Sun, L., Owen, C.J., Kornbluth, M., Kozinsky, B.：学习用于大规模原子动力学的局部等变表示。arXiv p. 2204 05249 17 05249 2022 10 48550 2204 . ( ). https://doi.org/ . /arXiv. . . Nguyen-Cong, K., Willman, J.T., Moore, S.G., Belonoshko, A.B., Gayatri, R., Weinberg, E., Wood, M.A., Thompson, A.P., Oleynik, I.I.：在极端条件下，以实验时间和长度尺度进行的十亿原子碳分子动力学模拟。《超级计算会议论文集》p. 4 (IEEE/ACM, 2021). https://doi.org/10 3487400 18 1145 3458817 . / . . Nomura, K., Kalia, R.K., Nakano, A., Rajak, P., Vashishta, P.：Rxmd：一种可扩展的反应分子动力学模拟器，可优化求解时间。SoftwareX **11**, 100389 100389 19 2020 10 1016 2019 ( ). https://doi.org/ . /j.softx. . . Nomura, K., Kalia, R.K., Nakano, A., Vashishta, P., Shimamura, K., Shimojo, F., Kunaseth, M., Messina, P.C., Romero, N.A.：按需氢的超可扩展量子分子动力学模拟。《超级计算会议论文集》，SC14 59 20 661 673 2014 10 1109 2014 pp. – (IEEE/ACM,). https://doi.org/ . /SC. . . Rajak, P., Aditya, A., Fukushima, S., Kalia, R.K., T. Linker, T., Liu, K., Luo, Y., Nakano, A., Nomura, K., Shimamura, K., Shimojo, F., Vashishta, P.：Ex-nnqmd：极大规模神经网络量子分子动力学。《国际并行与分布式处理研讨会论文集》研讨会，IPDPSW pp. 943 00145 21 946 2021 – (IEEE,). https://doi.org/ . /IPDPSW . . . Rossi, M., Ceriotti, M., Manolopoulos, D.E.：如何消除环聚合物分子动力学中的伪共振。《化学物理学杂志》**140**(23), 234116 (2014). https://doi.org/10.1063/1.4883861, https://doi.org/10.1063/1.4883861 22. Schmidt, R.M., Schneider, F., Hennig, P.：穿越拥挤山谷——对深度学习优化器进行基准测试。《国际机器学习会议论文集》，ICML **139**, 9367–9376 (2021). https://doi.org/10.48550/arXiv.2007.01547, https://proceedings.mlr.press/v139/schmidt21a.html 23. Schwalbe-Koda, D., Tan, A.R., G´omez-Bombarelli, R.：利用基于不确定性的对抗攻击对分子几何构型进行可微采样。《自然通讯》**12**(1), 5104 (2021). https://doi.org/10.1038/s41467-021-25342-8, https: //doi.org/10 8 24 1038 41467 021 25342 . /s - - - . Shafahi, A., Najibi, M., Ghiasi, M.A., Xu, Z., Dickerson, J., Studer, C., Davis, L.S., Taylor, G., Goldstein, T.：免费的对抗训练！《神经信息处理系统进展》论文集，NeurIPS **32** (2019). https://doi.org/10 12843 25 48550 1904 . / arXiv. . . Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., Riley, P.：张量场网络：面向三维点云的旋转和平移等变神经网络。arXiv p. 1802 08219 26 08219 2018 10 48550 1802 . ( ). https://doi.org/ . /arXiv. . . Togo, A., Tanaka, I.：材料科学中的第一性原理声子计算。《材料学报》**108**, 1–5 (2015). https://doi.org/10.1016/j.scriptamat.2015.07.021, https://www.sciencedirect.com/science/article/pii/S1359646215003127