计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(7): 1743−1762, 2026

机器学习原子间势的综述:模型、训练及评估

胡思宇[1] 谭光明[1,2] 贾伟乐[1,2]
[1](处理器芯片全国重点实验室(中国科学院计算技术研究所) 北京 100190)
[2](中国科学院大学 北京 101408)
(husiyu@ict.ac.cn)

Survey of Machine Learning Interatomic Potentials: Models, Training, and Evaluation

Hu Siyu[1], Tan Guangming[1,2], and Jia Weile[1,2]
[1](*State Key Laboratory of Processors* (*Institute of Computing Technology*, *Chinese Academy of Sciences*), *Beijing* 100190)
[2](*University of Chinese Academy of Sciences*, *Beijing* 101408)

Abstract Machine learning interatomic potentials (MLIPs) bridge the gap between microscopic quantum mechanics and macroscopic property simulation by fitting high-dimensional potential energy surfaces. While maintaining near-first-principles accuracy, MLIPs extend molecular dynamics simulations to significantly larger spatial and temporal scales. Driven by the synergy of "model, training, and evaluation", where diverse architectures need tailored training strategies, efficient optimizations can accelerate model development, and comprehensive task evaluation ensures the MLIPs' reliability. This paper provides a systematic review of the recent representative advances in MLIPs' architectures, training optimizations, and evaluation benchmarks. For model architecture, MLIPs can be categorized into invariant, equivariant, and unconstrained models based on their symmetry-keeping approach. We then analyze the evolution from "specialized models" to "pretrained models" and introduce several representative models. On the training strategies, we discuss key techniques such as framework-free implementations and kernel fusion for specialized MLIPs, alongside load balancing and two-stage training strategies for pretrained MLIPs. Finally, we list the evaluation metrics of modern benchmarks such as Matbench Discovery and FAIR Chemistry Leaderboard (OMol25).

Key words machine learning interatomic potentials (MLIPs); model architecture; training optimization; model evaluation; molecular dynamics simulation

摘 要 机器学习原子间势(machine learning interatomic potentials, MLIPs)通过拟合势能面,在接近第一性原理精度的同时,将分子动力学模拟推进到更大时空尺度,成为衔接微观量子力学与宏观物性模拟的有效途径。围绕模型架构、训练优化与评估体系协同演进的现状展开:多样化的模型架构要求设计特定的训练策略、高效的训练优化可提升模型的迭代速度,而严谨的多任务评估体系是度量模型可靠性的关键。系统综述了 MLIPs 在模型架构、训练优化及模型评估方面的代表性成果与最近进展。在模型架构层面,从对称性保持角度 MLIPs 可分为不变、等变及无约束 3 类,同时梳理了"专用模型"向"预训练模型"的演进历程和代表性模型。在训练优化层面,介绍了专用模型上无框架实现、优化器算法、算子融合等技

收稿日期:2026-01-09;**修回日期**:2026-05-27
基金项目:爆炸科学与安全防护全国重点实验室(北京理工大学)开放课题(KFJJ26-09M);国家自然科学基金项目(92270206, T2125013, 62372435)
This work was supported by the Opening Project of State Key Laboratory of Explosion Science and Safety Protection (Beijing Institute of Technology) (KFJJ26-09M) and the National Natural Science Foundation of China (92270206, T2125013, 62372435).

术,以及预训练模型上常用的负载均衡、两阶段训练等常用优化策略。在模型评估层面,分析了 Matbench Discovery 与 FAIR Chemistry Leaderboard(OMol25)等较为权威评测榜单的多个评价指标。

关键词 机器学习原子间势;模型架构;训练优化;模型评估;分子动力学模拟

中图法分类号 TP311.13; TP309

DOI: 10.7544/issn1000-1239.202660042 **CSTR:** 32373.14.issn1000-1239.202660042

机器学习原子间势(machine learning interatomic potentials, MLIPs)[1]通过学习基于密度泛函理论(density functional theory, DFT)的高精度数据,拟合势能面(potential energy surface, PES)。在保持接近量子力学精度的同时[2],可将分子动力学(molecular dynamics, MD)模拟推进到更大时空尺度[3-4]。基于密度泛函理论的第一性原理的分子动力学模拟,虽精度高但受限于时空尺度,其计算代价极其昂贵[5]。采用经验力场进行分子动力学模拟,效率高但精度受限[6]。MLIPs 的发展基于能量分解局部性原理[1],该原理通常将体系总能量分解为各原子局域环境的能量贡献之和,并假定每个原子的能量主要由截断半径内的邻域原子决定。因此可将高维 PES 的学习转化为可扩展的局域映射学习问题。该原理使得分子动力学模拟在原子规模上可线性扩展。目前,MLIPs 已在相变与成核过程模拟、离子扩散、固态电解质筛选、高通量材料发现、晶体结构弛豫、声子谱与热导率预测等场景上应用[7-11],成为计算材料领域的强大工具。

从模型对称性保持的角度来看,MLIPs 可分为:不变的(invariant)、等变(equivariant)的和无对称性约束的 MLIPs[12]。此外也存在多种对称性组合的混合 MLIPs[13],在本文中不展开介绍。其中不变 MLIPs 通过将原子坐标转换为标量特征(如键长、键角)作为输入,从根本上保证了旋转不变性,然而全标量化的建模可能会损失几何结构中的方向信息。等变 MLIPs 则在模型内部保留矢量或高阶张量特征,通常利用群论中的张量积和球谐函数确保特征随坐标系旋转而对应变换,从而保留角动量信息。无约束 MLIPs 放宽了架构上的硬性约束,通过数据增强或正则化约束让模型从数据中学习对称性,通常具有更快的推理速度和更灵活的架构设计,然而缺乏严格的物理一致性[12]。

在对称性分类的维度之外,MLIPs 的发展正经历着从专用(specialized)MLIPs 到预训练(pretrained)MLIPs 的范式转移。专用 MLIPs 在特定体系或有限类别体系的 DFT 数据下进行训练,训练后在特定的体系下可达到较高的精度(例如在小分子上误差可小于 1 kcal/mol),但体系依赖性强[14]。预训练 MLIPs 期望达到“一次训练,多次使用”的效果,因此预训练 MLIPs 尝试在覆盖更广化学空间的 DFT 数据上学习 PES 的通用表示,以达到在新体系上实现零样本或少样本泛化的目标[15-17]。

随着算力的提升和预训练数据集的积累,预训练 MLIPs 进入快速发展阶段。其中一个标志性事件是开源 MPtrj 数据集的出现,MPtrj 数据集包含元素周期表中 89 种元素组成的约 158 万个弛豫轨迹下的样本[15]。在 MPtrj 数据集出现之前,DFT 数据相对分散。以 DPA2 为代表的预训练 MLIPs 采用多任务的方式,通过组合构建多种不同 DFT 设置的数据集,进行 MLIPs 的预训练[18]。在 MPtrj 数据集出现后,预训练 MLIPs 的发表与公开数量快速增加。以 Matbench Discovery 榜单为例,截至 2025 年 12 月,预训练 MLIPs 的公开数量与评测参与度显著上升[19]。与此同时,诸如 OMat24, OMol25 等大规模开源数据集也进一步促进了预训练 MLIPs 的发展[20-21]。

由于专用 MLIPs 与预训练 MLIPs 两类模型的目标不同,专用 MLIPs 追求面向特定目标体系的极致精度[22],预训练 MLIPs 更关注泛化能力与模型的可迁移性,因此 2 类模型的设计、参数量也存在差异。专用 MLIPs 的可训练参数一般在数千至百万级别,通常在特定下游体系上从头开始训练。而预训练 MLIPs 为了在多种元素种类上具有更强的泛化性,通常在更大规模、多样化的开源 DFT 数据集上训练,其可训练参数一般在数百万乃至数十亿级别[16]。在下游任务上通常直接预测或经过微调后预测。

相应地,专用和预训练 MLIPs 在训练策略上也呈现不同侧重点。在专用 MLIPs 的发展过程中,优化算法[23]、主动学习(通过迭代采样不断补齐势能面覆盖范围)[24]等扮演着重要角色。预训练 MLIPs 的训练则更依赖大规模分布式训练策略、面向数据的定制化负载均衡、两阶段训练、参数微调等方法,以支撑大量 DFT 数据的高效训练[16]。诸如算子融合、编译优化等技术在专用和预训练 MLIPs 两类模型的训练过程中均被大量使用。

另一方面，由于 PES 本身的复杂性和 MLIPs 机器学习模型的黑盒性质，MLIPs 的评估成为关键问题之一[25]。能量与力的预测精度是 MLIPs 重要评价指标之一，但仅依赖能量与力的精度指标并不足以支撑稳健评估[26]。许多 MLIPs 在平均误差指标（如 RMSE、MAE）上可能表现良好，即便报告的平均误差显示模型接近第一性原理精度，然而由于误差在 MD 模拟过程中的传播效应，模型在实际模拟中仍可能无法复现实质性的物理现象，最终产生错误预测。为了能更客观地评价 MLIPs，目前出现了一系列的评测基准（benchmarks）[27-28]。本文也介绍了几个较新的测评榜单，例如 Matbench Discovery[19] 等。

近年来，已有多篇综述从不同角度总结 MLIPs 及其发展。例如，Behler[29-31] 系统梳理了高维神经网络势和原子环境描述符的构造方法。Unke 等人系统介绍了机器学习力场的基本概念、构建流程、应用案例和挑战[32]。Friederich 等人[33] 强调 MLIPs 在下一代物质模拟中的作用，重点讨论其在跨越时间和空间尺度方面相对 DFT 与经验势的优势[33]。Ko 与 Ong[34] 概述了 MLIPs 的最新进展、挑战和发展前景。此外，也有综述聚焦长程相互作用[35]、晶体缺陷[36] 和无机表面[37] 等特定问题。几何图神经网络相关综述则从图表示、等变性和群表示角度总结了三维原子体系建模方法[12,32,36]。也有工作从力场验证、分子模拟评估和下游物性预测角度讨论 MLIPs 的可靠性问题[25-26,28]。

与上述综述不同，本文重点从“模型架构—训练优化—模型评估”协同发展的视角展开，强调模型表达能力、训练效率和多任务评估之间的相互影响。三者的关联体现在：加速训练可以促进模型的快速发展。同时 MLIPs 模型经过重设计，又亟需多任务评估体系对模型可靠性进行判别。MLIPs 模型在多个任务上进行评估后，又可以指导 MLIPs 模型的架构改进。

1 预备知识

1.1 MLIPs 与能量分解局部性假设

MLIPs 采用机器学习模型拟合 DFT 数据，以近似 PES，使其在计算代价远低于 DFT 计算的前提下，准确预测体系总能量 E、原子受力 F 与应力等关键物理量。

2007 年，Behler 与 Parrinello 提出能量分解局部性假设（locality assumption）[1]，将高维势能面表示从“全局坐标到能量的黑盒映射”转化为可扩展的能量分解（energy decomposition）形式。如图 1 所示，输入数据经过周期性复制，针对输入数据里的每一个原子 i，在其截断半径 r_c 内构造邻居列表 R_i。然后基于邻域进行满足对称性条件的变换（图 1 中以 $R_1,\cdots,R_N$ 映射到 $D_1,\cdots,D_N$ 表示），通常通过合适的解析基函数/基组（如高斯基、径向/角向基等）、全连接神经网络结构、消息传递机制，以实现 R_i 到 D_i 的映射。R_i 到 D_i 映射方式成为不同 MLIPs 模型结构设计主要关注的内容。

每个结构的体系总能量被写成各原子能量贡献之和 $E=\sum_i E_i$，其中每个 E_i 仅由原子 i 的局域几何环境 D_i 决定，也即 E_i 可写为局域原子位置与原子序数的函数[29]。在实现上，相当于为每个原子构建共享权重的子网络。其中，力由能量对原子坐标的负梯度给出 $F=-\nabla_r E$。对于周期性边界的体系，应力可通过对晶格参数求导得到。在局部性原理的假设下，允许 MLIPs 与原子数呈线性可扩展的计算模式，这也奠定了后续几乎所有 MLIPs 架构的基本范式。

MLIPs 在学习上述映射关系，实现对总能量的拟

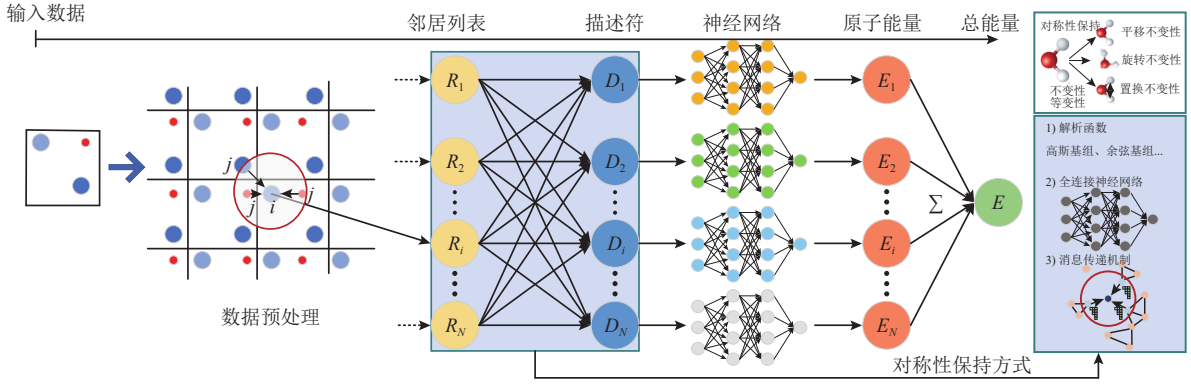


Fig. 1 Basic structure of MLIPs

图 1 MLIPs 的基本结构

合后,还应考虑物理一致性约束,例如能量守恒、势能面的可微性以及高阶导数的有界性,从而保障长时间 MD 模拟的数值稳定性与物性预测的可靠性[30-31]。

1.2 MLIPs 中的几何图表示与消息传递

经典图神经网络主要处理拓扑关系,满足节点置换不变性[38]。而原子体系除拓扑外还需具有几何结构信息,须显式编码欧氏空间坐标、距离、角度等几何量,我们称编码了这些几何信息的图神经网络为几何图神经网络(geometric graph neural network)。通常,几何图神经网络将原子作为节点、相互作用(化学键或在截断半径内的邻居与中心原子的交互关系)作为边,并要求模型在平移、旋转或反射等欧氏群变换下满足不变性或等变性[39]。

形式上,几何图神经网络$G=(V,E)$由节点集合V与边集合E组成。可将每个原子i视为一个节点,在第t层,原子i的节点特征记为$\boldsymbol{h}_i^t$,节点特征通常由原子类型等嵌入表示。在给定的截断半径r_c内的原子j被视为原子i的近邻,二者之间的相互作用可用一条边$\boldsymbol{e}_{ij}$表示,边特征通常包含几何信息(如用径向基函数展开原子间距r_{ij})。图神经网络在边上进行消息传递(message passing),并在节点上聚合(aggregation/pooling)邻域信息,以更新中心原子的表征。如式(1)所示,首先对节点i的邻居集合$j\in\mathcal{N}(i)$进行消息传递,得到更新后的消息向量$\boldsymbol{m}_{ij}^{t+1}$,其中$M_t$表示消息函数。再通过一个非线性更新函数$U_t$将聚合($\oplus$可表示求和、求均值等操作)的消息向量与中心节点特征融合,得到下一层的中心节点表示$\boldsymbol{h}_i^{t+1}$,如式(2)所示。几何图神经网络中的一个交互块(interaction block)包含消息计算(式(1))与特征更新(式(2))。多个交互块堆叠后,节点表征逐渐获得更长程的邻居原子的相互作用信息[2]。

$$\boldsymbol{m}_{ij}^{t+1}=M_t(\boldsymbol{e}_{ij},\boldsymbol{h}_i^t,\boldsymbol{h}_j^t)\,, \tag{1}$$

$$\boldsymbol{h}_i^{t+1}=U_t\left(\boldsymbol{h}_i^t,\oplus\,\boldsymbol{m}_{ij}^{t+1}\right)\,。 \tag{2}$$

1.3 群表示(Group Representation)和等变性

1)群表示。给定群G与向量空间$\mathcal{V}$,一个群表示是一个表示映射ρ把群元素作用到某个向量空间上的可逆线性映射:$\rho: G\to GL(\mathcal{V})$,满足对任意的$g$,$h\in G$,$\rho(g)\rho(h)=\rho(g\cdot h)$,其中$GL(\mathcal{V})$表示$\mathcal{V}$上所有可逆线性变换的集合[40]。

欧氏空间里常见的群包括:只含旋转与反射的正交群 O(n),只含旋转的特殊正交群 SO(n),含平移/旋转/反射的欧氏群 E(n),以及含旋转/平移的特殊欧氏群 SE(n);更一般地,许多连续对称性都可用李群(Lie Group)来描述[40]。

2)等变性[41]。模型Φ:$X\to Y$若关于群G等变,则对任意$g\in G$,$x\in X$满足交换关系:

$$\phi(\rho_X(g)\,x)=\rho_Y(g)\phi(x)\,,$$

其中ρ_X表示在输入空间X上的群作用$\rho_X: G\to GL(X)$,ρ_Y表示在输出空间Y上的群作用$\rho_Y: G\to GL(Y)$。直观地解释神经网络中的等变性为:先对输入进行变换,再经过神经网络,等价于输入先经过神经网络作用,再在输出向量上进行对应的变换。

不变性(invariance)是等变性的特例[42],如果输出对群变换完全不改变,即对任意$g\in G$,$\rho_Y(g)=\mathbb{1}$。此时等变性退化为不变性,即对任意$g\in G$,满足:

$$\phi(\rho_X(g)\,x)=\phi(x)\,。$$

2 MLIPs 模型的分类与介绍

从对称性视角,MLIPs 可分为:不变的、等变的与无约束的 MLIPs。此外也存在少量混合型 MLIPs,例如 Yan 等人提出的 HIENet[13],在本文中不展开介绍。总体来看,不变的 MLIPs 以距离、角度等标量作为输入,保证能量等标量的输出对平移/旋转具有不变性。等变的 MLIPs 通常在模型中显式引入矢量/张量表示与群表示理论算子,使力、应力等具有方向性的物理量在坐标变换下对应变换。无约束的 MLIPs 是指不在模型架构中编码几何对称性,仅依赖数据增强、正则化等策略让 MLIPs 模型在训练中学出数据本身具有的对称性[42]。随后针对上述 3 类 MLIPs,分别给出有代表性的专用 MLIPs(针对单一体系/数据集训练、强调特定数据集上的精度)与预训练 MLIPs(在大规模多体系数据上学习通用表示、支持迁移微调)的介绍。下文将分别围绕 3 类不同对称性约束下的 MLIPs 展开综述。

2.1 不变的 MLIPs

不变的 MLIPs 指模型的中间表示在对输入进行某种对称变换后保持不变。形式上,如果对原子坐标施加欧式群 E(3) 的变换(如平移/旋转/反射),一个不变的函数f应满足$f(g\cdot X)=f(X)$,即对于任意平移/旋转/反射变换g下$f(X)$均保持不变。例如一个原子系统的总势能只取决于原子间的相对距离,而不取决于原子所处的绝对坐标系。当整个系统平移或旋转时,总能量不应该改变。

不变的 MLIPs 又可分为 2 条路线:一类是面向特定体系或单一数据集的专用小模型,强调在有限

数据与固定化学空间内追求极致精度。另一类是在大规模多元素、多构型预训练材料数据集上训练得到的预训练 MLIPs，更关注跨体系泛化与迁移微调能力。下面将分别介绍这 2 类的代表方法与典型模型。

2.1.1 不变的专用 MLIPs 及有代表性模型介绍

专用 MLIPs 面向单一体系或相近体系，通常可分为 3 类：基于手工描述符的 MLIPs、基于全连接的端到端嵌入的 MLIPs、基于消息传递的端到端嵌入的 MLIPs。

1）基于手工描述符的 MLIPs。先将原子邻居列表 $\mathcal{R}_i$ 经人工指定的解析函数映射到描述符 $\mathcal{D}_i$，再用拟合网络（通常为全连接神经网络）预测能量/力。解析函数的构造往往基于物理对称性约束（平移、旋转、置换不变性），可解释性较强、计算高效，但描述符设计依赖领域专家经验。其中最有代表性的例子为 Behler 与 Parrinello 提出的 HDNNP，HDNNP 是基于手工描述符的 MLIP，是 MLIP 的启蒙工作[1]。HDNNP 采用两体/三体高斯对称函数作为手工描述符，两体项刻画径向信息，三体项引入角度信息，并在截断半径附近使用光滑截断函数以保证势能面的平滑。

HDNNP 模型里的中心原子 i 在两体和三体高斯基组的表示下，其描述符 G_i^{2B} 和 G_i^{3B} 的计算分别如式（3）（4）所示。两体高斯基组可建模邻居原子 j 对中心原子 i 的影响，用于建模径向信息。三体高斯基组可建模中心原子 i 与邻居原子 j 和 k 的角向相互作用，用于描述角度信息。

$$G_i^{2B}=\sum_{i\neq j}\mathrm{e}^{-\eta(r_{ij}-r_\mathrm{c})^2}f_\mathrm{c}(r_{ij}),\tag{3}$$

$$G_i^{3B}=2^{1-\zeta}\sum_{\substack{j,k\neq i\\ j<k}}(1+\lambda\cos\theta_{ijk})^{\zeta}\mathrm{e}^{-\eta\left(r_{ij}^2+r_{ik}^2+r_{jk}^2\right)}\cdot f_\mathrm{c}(r_{ij})f_\mathrm{c}(r_{ik})f_\mathrm{c}(r_{jk}),\tag{4}$$

其中，$\lambda,\eta,\zeta,r_\mathrm{c}$ 为用户指定的超参数。$f_\mathrm{c}(r_{ij})$ 表示平滑函数，其计算如式（5）所示。平滑函数可以刻画邻居原子与中心原子的笛卡尔距离 r_{ij} 对基组的影响。当 r_{ij} 接近 0 时，平滑函数接近 1；当 r_{ij} 趋于截断半径 r_c 时，按余弦函数逐渐衰减到 0；当 r_{ij} 大于截断半径 r_c 时，平滑函数为 0。$\cos\theta_{ijk}$ 为向量夹角余弦，其计算如式（6）所示。把中心原子 i 周围的邻居键角编码为一个无量纲标量，再与三体高斯基组中的其他项结合，用以区分不同角度分布的邻居原子对中心原子的影响，以此建模多体的表示能力。在实际使用中，两体高斯基组和三体高斯基组通常拼接起来作为原子描述符。

$$f_\mathrm{c}(r_{ij})=\begin{cases}0.5\times\left[\cos\left(\dfrac{\pi r_{ij}}{r_\mathrm{c}}\right)+1\right], & r_{ij}\leqslant r_\mathrm{c}\\ 0, & r_{ij}>r_\mathrm{c}\end{cases},\tag{5}$$

$$\cos\theta_{ijk}=\frac{r_{ij}\cdot r_{ik}}{|r_{ij}|\cdot|r_{ik}|}。\tag{6}$$

其他一些基于手工描述符的 MLIPs 包括 NEP，其利用切比雪夫多项式和勒让德多项式构造径向与角向不变描述符，并以全连接神经网络预测原子能，其 GPUMD 实现具有较高的分子动力学计算效率[43]。此外，两体和三体余弦基组（2Body, 3Body Cosine Basis, 2B3B-Cosine）[44]、矩张量势（moment tensor potential, MTP）[45]、谱邻域分析势（spectral neighbor analysis potential, SNAP）[46] 等也属于基于手工描述符的 MLIPs。诸如上述基于手工描述符的 MLIPs 需精心设计解析表达式。描述符需要能够编码与问题相关的物理相互作用。随着 MLIPs 的发展，描述符逐渐从手工设计转向端到端的表征方式，而端到端嵌入表示的模型又可分为基于全连接神经网络和基于消息传递机制 2 类。

2）基于全连接的端到端嵌入的 MLIPs。端到端的 MLIPs 不需显式地设计解析表达式。一个有代表性的基于全连接神经网络的端到端嵌入的专用 MLIPs 模型为深度势能模型（deep potential model, DP）[47]，DP 的实现如下。

DP 由嵌入网络和拟合网络组成。其中嵌入网络把局域径向信息映射为可学习的嵌入矩阵 $\boldsymbol{G}_i$；再根据对称性描述符，进一步构造满足置换不变性的原子环境表征 $\boldsymbol{D}_i$，并作为后续拟合网络的输入；最后由拟合网络得到原子能 E_i。形式化表示为：首先根据邻居列表计算原子的笛卡尔距离 r_{ij}，经过一个连续可微的平滑函数 s 变换得到的输出记为 $S_{ij}\triangleq s(r_{ij})$。经过全连接嵌入神经网络 $G(S_{ij})$ 的作用，得到嵌入矩阵 $\boldsymbol{G}_i\in\mathbb{R}^{N_m\times M}$，其中 N_m 为每个原子的最大邻居数，M 为嵌入网络最后一层的神经网络节点数。然后取 $\boldsymbol{G}_i$ 的前 M' 列得到的截断后嵌入矩阵 $\boldsymbol{G}_i'$，通常 $M'\leqslant M$。按照式（7）所示生成 $\boldsymbol{D}_i$，$\boldsymbol{D}_i\in\mathbb{R}^{M\times M'}$，其中 $\tilde{\boldsymbol{R}}_i\triangleq s(r_{ij})(1,\vec{r}_{ij}/\|\vec{r}_{ij}\|)\in\mathbb{R}^{N_m\times 4}$，而 $\vec{r}_{ij}=(x_j-x_i,y_j-y_i,z_j-z_i)$，$(x_i,y_i,z_i)$ 为原子 i 的位置，(x_j,y_j,z_j) 为原子 j 的位置。式（7）保证了 $\boldsymbol{D}_i$ 对原子体系构型的平移、旋转以及交换不变性。最后 $\boldsymbol{D}_i$ 经过全连接拟合神经网络 $F(\boldsymbol{D}_i)$，即可得到 E_i。

$$\boldsymbol{D}_i=\boldsymbol{G}_i^{\mathrm{T}}\tilde{\boldsymbol{R}}_i(\tilde{\boldsymbol{R}}_i)^{\mathrm{T}}\boldsymbol{G}_i',\tag{7}$$

3）基于消息传递的端到端嵌入的 MLIPs。端到端嵌入的专用 MLIPs 中，除了以全连接神经网络为骨架的 MLIPs，还有基于消息传递的端到端嵌入的

MLIPs。消息传递机制中每一个交互块通过计算信息、聚合再更新节点表示，经过 k 个交互块迭代地构造原子表征，使得节点特征在 k 层后包含 k-hop 邻居的信息。

SchNet[48] 堆叠若干个交互块以获得多跳邻居原子的影响。SchNet 把从邻居 j 发送到中心原子 i 的消息表示为 $\boldsymbol{m}_{ij}=\sigma_2(r_{ij})\sigma_1(\boldsymbol{h}_j)$，其中 $\boldsymbol{h}_j$ 是邻居原子 j 的嵌入/特征表示，$\sigma_1(\cdot)$ 对邻居特征做一次可学习的非线性变换，而 $\sigma_2(r_{ij})$ 则根据笛卡尔距离 r_{ij} 生成连续卷积滤波器，可理解为随距离变化的一组通道权重，连续滤波卷积是 SchNet 的关键算子。总体来看，消息 $\boldsymbol{m}_{ij}$ 的计算本质上是距离条件滤波器与邻居特征嵌入的逐通道乘法耦合。为了使滤波器与特征变换都具备足够的表达能力，σ_1 和 σ_2 为全连接操作，以在连续空间中刻画复杂的原子相互作用规律。

其他基于消息传递的端到端嵌入的 MLIPs 还包括：DimeNet 引入了两键之间的夹角，提出定向消息传递机制以同时利用距离和角度计算消息[49]。GemNet 在 DimeNet 基础上加入二面角（扭转角）信息，将消息传递扩展到两跳原子四元组，捕捉三键构型的扭转角关系[50]。DimeNet++在 DimeNet 的基础上简化了交互块结构并优化了基函数计算，在保持距离和角度等几何信息建模能力的同时显著降低了计算开销，提高了训练和推理效率[51]。SphereNet 利用球面坐标系整合原子间距离、角度和扭转角等 3D 信息，提出球面消息传递以高效建模大分子[52]。ComENet 提出的消息传递机制实现了全局完备的 3D 信息建模，避免显式计算扭转角[53]。

2.1.2 不变的预训练 MLIPs 及有代表性模型介绍

预训练 MLIPs 通过在多样化、大规模的 DFT 数据集上进行学习，得到具有泛化能力的 PES，使得模型不再局限于单一材料/化学体系。预训练 MLIPs 通过学习跨越更大化学空间的可迁移表征，从而在下游任务上可以零样本（zero-shot）推理或少量样本微调后推理。不变的预训练 MLIPs 往往具有实现简洁、数值稳定、易于大规模训练等优势。然而由于几何信息的标量化处理，其方向信息可能弱化，目前出现以下思路：通过更高阶（如角、二面角）的结构或更强的表示学习机制来弥补表达能力。绝大多数预训练模型针对 1 个或多个预训练数据集进行训练，一个有代表性的不变的预训练 MLIP 为 Materials Representation and Interaction Simulation（简称 MatRIS）。部分预训练模型利用不同 DFT 设置、不同领域的数据集，实现联合训练。另一个有代表性的不变的预训练 MLIP 为 Deep Potential Attention v3（简称 DPA3）。

MatRIS 模型[54] 中包括原子图和键图，其中原子图以原子为节点，在给定截断半径内的原子建立连接的边，用于描述二体相互作用。键图将原子图中的边（原子对）抽象为节点，中心原子 i 与邻居原子 j,k 若能形成键角，则映射为键图中的边，键图用于刻画三体相互作用。MatRIS 在消息传递模块中引入基于注意力的更新机制，并把键图中更新的特征回传到原子图中。模型采用源-目标双通道的注意力策略，分别计算中心原子作用于其邻居（源侧）和邻居作用于中心原子（目标侧）的注意力权重，并在特征通道的维度上分别执行 Softmax 归一化操作，以此保证注意力机制线性标度的计算量。

DPA3[55] 把数据集/任务作为条件信息，在不同 DFT 设置、不同领域数据间实现联合训练。利用此方式可以从工程实现上对任务数与模型参数量进行解耦合，使模型规模基本不随数据集数量增长。更早期的 DPA2 处理多任务的方式是对每个数据集/任务设置一个输出头[18]。当新增一个数据集/任务时，只需对应增加一个该任务的输出头，整体网络骨架无需改变[56]。DPA3 不仅在原子图 $G^{(1)}$ 上做消息传递，同时构造一系列高阶线图（line-graph series, LiGS）$\{G^{(2)},G^{(3)},\cdots,G^{(K)}\}$。高阶图的节点对应更高体的几何实体，例如 $G^{(2)}$ 的节点表示键、$G^{(3)}$ 的节点表示角、$G^{(4)}$ 的节点表示二面角等。每一个交互层 l 中，DPA3 在 $G^{(K)}$ 上更新顶点特征 $\boldsymbol{v}^{l,K}$ 与边特征 $\boldsymbol{e}^{l,K}$，并把 $G^{(K)}$ 上的顶点特征回传给 $G^{(K-1)}$ 的边特征，依此类推得到原子图上的节点信息。

其他不变的预训练 MLIPs 如 CHGNet 也构建高阶图，以捕获多体相互作用[15]。再如基于图原子簇展开（graph atomic cluster expansion）的 GRACE 也是不变的 MLIPs[57]。CHGNet 基于 MPtrj 数据集训练得到 CHGNet-v0.3.0 检查点。GRACE 在 OMat24 数据集上预训练得到 GRACE-1L 与 GRACE-2L 两个不同层数的检查点。DPA3 基于 MPtrj 数据集训练得到 DPA-3.1-MPtrj 检查点，基于 OpenLAM-v1 大规模多学科数据集（包含约 1.63 亿个训练样本）训练并进行微调得到 DPA-3.1-3M-FT 检查点。

2.2 等变的 MLIPs

等变的 MLIPs 显式编码几何特征，采用向量或更高阶张量（tensorial representations）在网络中传播信息，使得模型交互块的输出随输入的旋转和平移而对应变化。等变性通常对力预测、各向异性相互作用建模更有利，等变的预训练 MLIPs 目前在多个基

准测试中仍然保持着高精度[17,19-21,27,58]。另一方面，等变性也意味着更复杂的张量运算与更高的计算开销。下面分别介绍有代表性的专用的和预训练的等变模型。

2.2.1 等变的专用 MLIPs 及有代表性模型介绍

等变的 MLIPs 主要分为 2 类：基于向量的等变模型和基于高阶张量的等变模型[12,42]。基于向量的等变模型是笛卡尔坐标系下的低阶等变模型，仅利用坐标差和简单的向量运算来传递方向信息，也叫标量化的 MLIPs（Scalarization-Based MLIPs）。基于高阶张量的等变模型是指高阶球面张量等变模型运用完整的群表示工具处理任意阶的特征，也叫高阶可导向的 MLIPs（High-Degree Steerable MLIPs）。

1）基于向量的等变专用模型。基于向量的模型在消息传递时除了对标量信息进行变换，同时还维护向量表示，一般来说，通过简单的操作（如直接利用原子间相对位移$\vec{r}_{ij}$）更新向量特征。基于向量的等变模型等效于考虑$l=0$（标量）和$l=1$（矢量）两种表示，利用向量的加、减、点积、叉积等操作保持等变性。由于基于向量的模型无需分解高阶张量，其计算开销接近于不变模型。低阶等变模型能够在一定程度上提高方向相关变量的捕获能力，在一些中小规模分子和材料数据集上取得了优于不变模型的精度，同时保持了较高的计算效率。总体来说基于向量的等变模型的优点是实现简单、计算代价低，无需昂贵的球谐变换即可保证输出的等变。EGNN[59]是有代表性的标量化的等变专用模型。

EGNN 在消息传递中先用相对距离r_{ij}构造不变性的标量消息，如式（8）所示。

$$\boldsymbol{m}_{ij}=\sigma_1(\boldsymbol{h}_i,\boldsymbol{h}_j,\|\vec{r}_{ij}\|^2,\boldsymbol{e}_{ij}), \tag{8}$$

其中$\boldsymbol{h}_i,\boldsymbol{h}_j$为节点特征，$\boldsymbol{e}_{ij}$为边特征，$\|\vec{r}_{ij}\|^2$为平方距离（对旋转/平移不变），即$\|\vec{r}_{ij}\|^2=(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2$，$(x_i,y_i,z_i)$为原子 i 的笛卡尔坐标，(x_j,y_j,z_j)为原子 j 的笛卡尔坐标。式（8）得到的$\boldsymbol{m}_{ij}$经过σ_2函数作用，得到标量$\sigma_2(\boldsymbol{m}_{ij})$。$\sigma_2(\boldsymbol{m}_{ij})$用于更新原子的位置信息，如式（9）所示，

$$\begin{aligned}x_i&=x_i+c\sum_j(x_i-x_j)\sigma_2(\boldsymbol{m}_{ij}),\\ y_i&=y_i+c\sum_j(y_i-y_j)\sigma_2(\boldsymbol{m}_{ij}),\\ z_i&=z_i+c\sum_j(z_i-z_j)\sigma_2(\boldsymbol{m}_{ij})。\end{aligned} \tag{9}$$

EGHN[60]为 EGNN 的变体。其他基于向量的专用 MLIPs 包括 GMN，GMN 将 EGNN 拓展为多通道形式，允许每个节点携带多个几何向量（如位置、速度等）[61]。PaiNN 通过耦合旋转不变的标量特征与旋转等变的向量特征，并利用原子间相对方向传递几何信息，实现旋转等变消息传递[62]。ClofNet 将邻域原子的相对位置矢量正交化得到不变的局部坐标基，在该基下提取标量消息并投影回原坐标系产生等变更新[63]。其他基于向量的等变模型还包括 FastEGNN[64]、ETNN[65]、VisNet[66]等。

2）基于高阶张量的等变专用模型。为了充分发挥等变性的表达潜力（建模更高阶的角动量信息），一些 MLIPs 采用了球坐标系和球谐张量的方法，直接在 SE(3)/E(3) 群表示的框架下进行卷积和特征更新[67]。NequIP 是近年面向原子力场的高阶等变 MLIP 的代表模型之一。它采用 E(3) 等变的神经网络（基于球谐张量的消息传递），在多个分子/材料数据集上实现了较高精度和高数据效率。且 NequIP 在小分子数据集（如 MD17）上在给定极少训练数据的情况下仍能准确拟合复杂的 PES，展现出较高的数据效率。NequIP 的实验表明严格嵌入等变性的 MLIP 可以降低训练数据量的需求，提高模型外推能力。下面介绍 NequIP 的模型结构。

NequIP[68]分为 3 个模块：嵌入块、交互块、输出块。其中嵌入块对原子种类$(Z_1,\cdots,Z_n)$进行编码。交互块的输入为编码后的原子种类信息和原子距离$(\vec{r}_1,\cdots,\vec{r}_n)$，得到的原子表征如式（10）所示。NequIP 的输出块由原子表征得到原子能E_i。

$$S_m^l(\vec{r}_{ij})=R(r_{ij})Y_m^l(\vec{r}_{ij}), \tag{10}$$

其中$R(r_{ij})$是由基函数$B(r_{ij})$经过全连接得到，如式（11）所示。基函数$B(r_{ij})$的计算如式（12）所示，$f_c(r_{ij})$为多项式函数。式（10）中的$Y_m^l(\vec{r}_{ij})$为球谐函数。

$$R(r_{ij})=\sigma(B(r_{ij})), \tag{11}$$

$$B(r_{ij})=\frac{2}{r_c}\frac{\sin\left(\frac{b\pi}{r_c}r_{ij}\right)}{r_{ij}}f_c(r_{ij})。 \tag{12}$$

其他高阶可导向的等变专用模型包括：Tensor Field Network（TFN）是高阶可导向的等变专用模型的开创性工作之一。TFN 利用球谐函数将邻居原子的相对位置映射到特定阶的张量表示，并通过 Clebsch-Gordan（CG）系数（张量积耦合规则）来实现等变卷积[69]。SE(3)-Transformer 则将等变思想与注意力机制结合，引入了矢量式的查询、键和值，使注意力权重计算和消息聚合都满足 SE(3) 等变性。通过注意力框架，SE(3)-Transformer 在保证对称性的同时提高了长程建模能力[70]。基于 Transformer 架构的等

变 Transformer (TorchMD-Net)利用自注意力机制全局建模粒子间相互作用,并在注意力计算中融入等变约束[71]。TorchMD-Net 在多个分子基准(如 MD17、QM9、ANI-1)上取得了高精度预测。MACE 将原子簇展开(ACE)方法融入消息传递网络,在消息传递层中通过高阶张量积构造二体、三体甚至四体等变特征[72]。PACE 与 MACE 类似,是另一个基于高阶张量的 SE(3) 等变专用 MLIP。PACE 增加边增强模块(edge booster),通过将边增强模块与 ACE 结合,可捕获更复杂的原子几何关系和高阶多体相互作用,同时保持模型对平移、旋转和同类原子置换的对称性[73]。

2.2.2 等变的预训练 MLIPs 及有代表性模型介绍

2.2.1 中的基于高阶张量的等变专用模型中介绍的 MACE[72] 和 NequIP[68] 于 2022 年分别由 Batatia 等人和 Batzner 等人提出。两者最初主要在小分子数据集上进行验证。随着大规模预训练 DFT 数据的出现,MACE 和 NequIP 从专用 MLIPs 扩展到预训练 MLIPs,得到预训练的检查点模型文件。预训练模型需要处理更多种元素种类的数据,不同元素通过原子类型的嵌入表示传入模型,而无需改变 MACE 和 NequIP 主干网络的结构。

1)基于 SO(3) 等变的预训练模型。MACE 和 NequIP 将几何信息以严格的群表示形式嵌入模型,是等变预训练 MLIPs 的典型技术路线之一。MACE 基于 MPtrj 弛豫轨迹数据集训练得到 MACE-MP-0 检查点。MACE 基于 OMat24 预训练数据集得到 MACE-MPA-0 检查点。MACE-MP-0 和 MACE-MPA-0 可泛化到无机晶体体系[17]。MACE 基于 SPICE 数据集训练得到的 MACE-OFF 检查点,可用于分子数据集的预测[74]。NequIP 基于 MPtrj 预训练数据集得到 Nequip-MP-L 检查点[75]。Koker 等人[76] 借助 JAX 框架进行 NequIP 的训练,同时替换优化器为 Muon,得到 Nequix MP 检查点。为了支持更大规模预训练数据集(例如 OMat24)的训练,Tan 等人[75] 借助 PyTorch 2.0 compiler 实现 NequIP 更高效的训练和推理,得到 Nequip-OAM-XL 和 Nequip-OAM-L 检查点。Park 等人[77] 在 NequIP 上通过并行优化,支持大规模分子动力学,提出可扩展等变神经网络。SevenNet-l3i5 为 SevenNet 在 MPtrj 数据集上训练得到的检查点。在 SevenNet 的基础上,Lee 等人[78] 提出 FlashTP 算子,优化 NequIP 中的张量积操作,在不损失精度的情况下将张量积运算速度提升数倍,在 OMat24 数据集上得到 EquFlash 检查点。

NequIP、MACE 等基于 SO(3) 卷积的 MLIPs 通过高阶张量耦合获得了优异精度,然而随着阶数 L 的增加计算量和显存占用变得不可忽视。其瓶颈主要来自于消息传递里计算基于球谐/不可约表示(irreps)的张量积操作,由于全张量积的计算复杂度为 $O(L^6)$,使得模型通常只能用到很低的阶数 L(如 L=2 或 L=3),其中 L 为最高角动量阶数。为了提升等变模型的效率,Passaro 和 Zitnick[79] 提出将球谐卷积从 SO(3) 群降阶为在 SO(2) 上的卷积,在满足数学等价的同时可降低计算成本。

2)基于 SO(2) 等变的预训练模型。Passaro 和 Zitnick 通过将节点表示的“主轴”与边方向对齐,把 SO(3) 卷积数学上等价为 SO(2) 卷积,把复杂度从 $O(L^6)$ 降到 $O(L^3)$,从而使用更高的角动量阶数在计算上可行[72]。直观上的解释为:球谐函数 $Y_l^m(\theta,\phi)$ 当主轴固定时,相当于 θ 角度固定,3D 旋转里只剩下绕该边旋转的自由度,因此消息传递只需对角度 ϕ 的旋转保持等变即可。SO(3) 卷积降为 SO(2) 卷积后,避免了全张量积操作,进而减少了高阶 irreps 的计算与内存压力。

eSCN 在消息传递层中采用 SO(2) 卷积,在每条边 $s \to t$ 的消息传递中,将节点表示的主轴旋转对齐到边的方向,从而使球谐函数在边的方向上变得稀疏,避免显式计算 CG 系数。eSCN 在 OC20/OC22 大规模数据集上展示了对方向敏感任务(如力的预测、弛豫)的优势[79]。eSEN 在边的卷积上复用 eSCN 的 SO(2) 卷积操作,并在每个消息传递层的最后,加入 envelope 函数,对距离相关的边消息进行平滑化处理,以改善能量面导数的连续性,同时保证数值稳定性[80]。UMA 为了实现跨领域任务的拟合,将 eSEN 的 SO(2) 卷积内部多处线性算子替换为 Mixture of Linear Experts (MoLE)[16]。

eSEN 基于 MPtrj 数据集训练得到 eSEN-30M-MP 检查点,基于 OMat24 数据集训练得到 eSEN-30M-OAM。UMA 在 5 个大规模 DFT 数据集上(材料数据集 OMat24、分子数据集 OMol25、催化数据集 OC20++、分子晶体数据集 OMC25、MOFs 数据集 ODAC25,合计约 4.59 亿训练样本)混合训练得到 3 个不同参数规模的模型检查点文件:UMA-sm、UMA-md、UMA-lg。

3)弱旋转等变的预训练模型。目前存在一种非严格等变的模型,即不在模型中硬编码等变性,而是通过在可学习的函数空间中施加软约束(类似正则化)的方式让原子受力具有近似的旋转等变性。此种通过约束保证等变性的也可被认为是无对称性约束的模型。

ORB v3[81] 中引入 equigrad 用于量化旋转不变性的程度，并可作为训练时的正则化方法。ORB v3 在输入处插入一个恒等旋转矩阵 $\boldsymbol{R}$，然后计算预测能量 E 关于该旋转矩阵 $\boldsymbol{R}$ 的梯度。ORB v3 把 $\boldsymbol{R}$ 写成反对称零矩阵的指数映射形式：$\boldsymbol{R}=\mathrm{e}^{(\boldsymbol{G}-\boldsymbol{G}^{\mathrm{T}})}$，再计算预测能量 E 对该零矩阵 $\boldsymbol{G}$ 的梯度，如式(13)所示：

$$\Delta_{\mathrm{rot}}=\frac{\partial E(\boldsymbol{r}^{\mathrm{T}}\boldsymbol{R},\boldsymbol{h}\boldsymbol{R})}{\partial \boldsymbol{G}}\Big|_{\boldsymbol{G}=\boldsymbol{0}}, \tag{13}$$

其中 $\boldsymbol{r}$ 是原子坐标，$\boldsymbol{h}$ 是晶胞矩阵。若模型严格旋转不变，则 $\Delta_{\mathrm{rot}}=\boldsymbol{0}$。ORB v3 在训练时对 $\|\Delta_{\mathrm{rot}}\|$ 做 L2 正则，即可在不改网络为显式等变结构的前提下，直接惩罚“能量随全局旋转变化”的趋势，从而诱导模型学习旋转不变性/等变性。

2.3 无对称性约束的 MLIPs

无约束的 MLIP 包含不对模型进行对称性的约束、不对模型进行保守性约束等。无对称性约束的模型是指不对模型的中间表示施加明确的对称性约束，仅通过数据或额外损失项来自行学习旋转对称性质[12]。无保守性约束指 MLIP 模型的原子受力计算不严格来自于能量对原子位置的负梯度。无约束的模型往往在某些指标上表现突出(如直接力预测可提高结构优化精度)，但可能缺乏严格的物理一致性。因此对于需要长时间动力学稳定性的任务，无约束的 MLIPs 可能导致能量不守恒或体系动力学不稳定[26]。无约束的 MLIPs 适合用于静态结构进行一次性推理及微小的对称性误差不敏感的任务。

2.3.1 无对称性约束的专用 MLIPs 及有代表性模型介绍

2.1 和 2.2 节中介绍的不变性和等变性都是对称性要求下的约束，不变性和等变性约束的加入相当于在函数空间中加入 inductive bias。不变性和等变性 MLIPs 在模型训练过程中可能限制函数空间。无约束模型不要求中间特征必须按群表示严格变换，允许模型在训练时有更大的搜索空间[82]。

一种无对称性约束的实现方式是数据增强。数据增强是计算机视觉中常用的提升模型泛化能力的方法[83]。ForceNet 是较早运用数据增强的无约束专用 MLIP。ForceNet 用旋转数据增强鼓励力的预测满足物理约束，并预测不稳定性指标(旋转后再旋转回，与不经旋转的力的预测对比，越小越接近协变)。ForceNet 表示在数据集较多的场景下，结合旋转增强，能显著降低预测不稳定性指标。在 OC20 的 S2F (atomic force prediction)评测中，ForceNet 用远少于 DimeNet++的训练成本达到相近甚至更优的原子受力预测结果[84]。

另一个有代表性的数据增强的无对称性约束模型为 FAENet。FAENet 利用主成分分析方法在每个原子构型上构建若干个局部输入点，并对局部输入点进行规范化投影，再对模型输出施加相应逆变换，从而在不限制主干网络的情况下经验式地实现对称性[82]。FAENet 避免强制用能量梯度求力所带来的 2~4 倍开销。且在 OC20 数据集上表现出精度与速度的折中，在 IS2RE 的预测上推理吞吐相对 DimeNet++提升约 21 倍、相对 ComENet 提升约 6 倍，并取得更低误差。

2.3.2 无保守性约束的 MLIPs 及有代表性模型介绍

无保守性约束的 MLIPs 包括直接力预测的 MLIPs，当通过输出头直接预测力，不再被强制为能量的负梯度，属于无保守性约束的模型，如等变的 Equiformer V2[85]。目前公开提供非保守检查点的模型还包括弱等变的 Orb-v3、等变的 eSEN 等。例如，Orb-v3 同时发布了直接力与保守力版本；eSEN 也针对 OMol25、OC25 等数据集提供了直接力检查点。此外，Open Catalyst 模型库中的不变的 GemNet-dT、GemNet-OC、SpinConv 和等变的 SCN、eSCN 等模型也公开了直接预测原子力的 S2EF 检查点。这些模型虽然可能在架构上满足几何不变性/等变性，但其预测力不被强制为预测能量的负梯度，因此属于无保守性约束的 MLIP。

由于无保守性约束的 MLIP 模型通过同时设置能量输出头和力输出头，即分别拟合能量与力标签，因而不严格满足能量-力一致性。直接力预测避免了通过能量反向传播计算力的过程，可降低训练和推理开销，并允许模型针对力预测目标进行独立优化，因此适合原子受力预测、静态结构弛豫和高通量结构筛选等任务。对于长时间分子动力学和要求严格能量守恒的任务，则应优先采用保守模型，或对直接力模型的能量漂移和动力学稳定性进行专门验证。

2.4 MLIPs 分类及总结

根据对称性保持的方式 MLIPs 可分为不变的、等变的和无约束的 MLIPs 三大类。图 2 最上面部分列举了一些有代表性的不变 MLIPs 模型，基于手工描述符的 HDNNP 是 MLIPs 的鼻祖[1]。此后 MLIPs 多为端到端嵌入的方式。端到端嵌入的 MLIPs 又可分为基于全连接神经网络和基于消息传递的 MLIPs，代表模型为 DP、SchNet。这些模型属于不变的模型，它们利用原子间距离、角度等变量保持平移与旋转不变性，不变的模型具有较高的计算效率[12]。随着

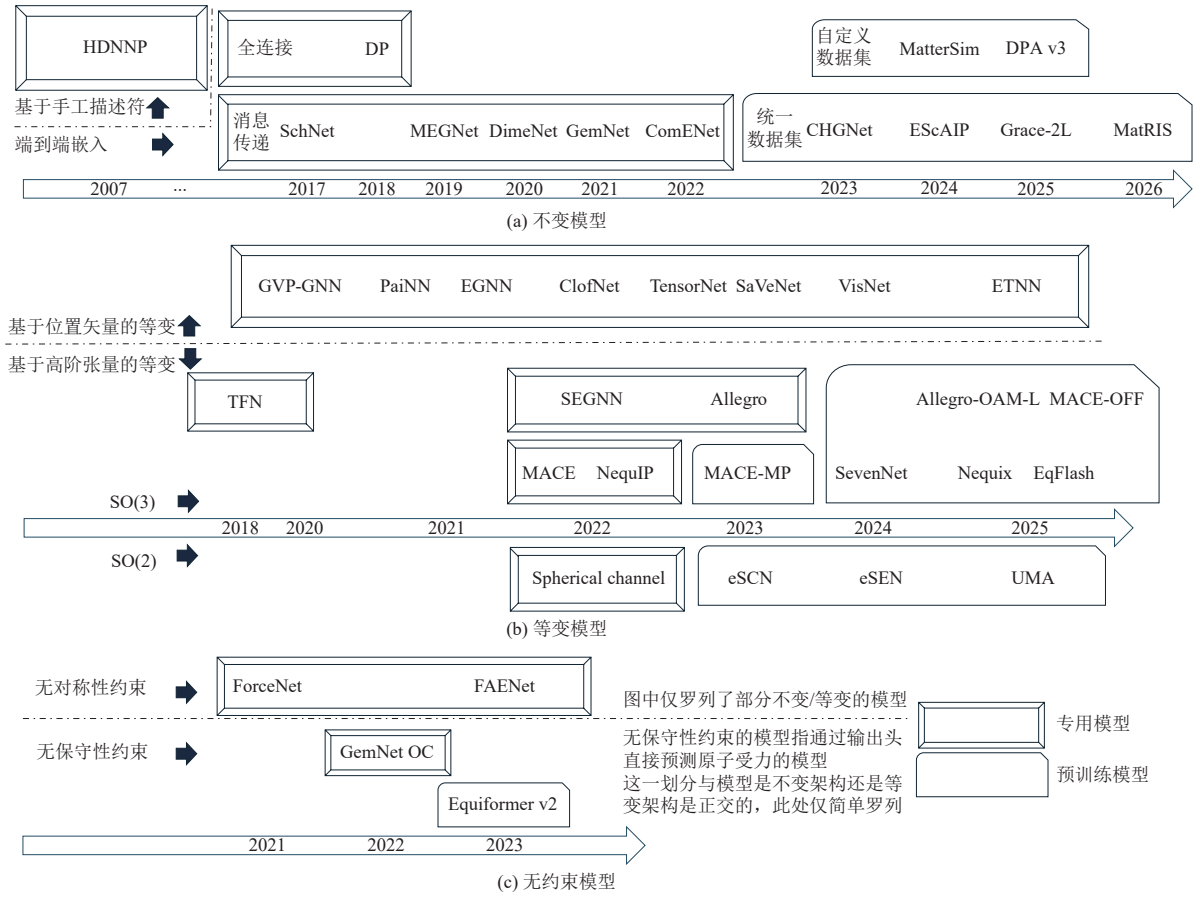


Fig. 2 Representative MLIP models

图 2 有代表性的 MLIPs 模型

DFT 数据集的积累，预训练模型开始不断被提出。预训练模型根据采用的数据集，可以分为自定义数据集的预训练模型和统一数据集的预训练模型。其中自定义数据集的预训练模型包括 DPA2、MatterSim。DPA2 为多任务预训练模型，它们选择了不同泛函的数据一起训练，MatterSim 生成了大量非平衡态构型作为训练数据集。MPtrj 的开源允许大量的模型基于 MPtrj 进行训练。统一数据集的预训练模型包括 CHGNet、MatRIS、Grace 等。不变的模型由于缺乏高阶张量信息，部分模型通过引入更高阶的多体相互作用特征（如高阶线图）以提升模型的表达能力。

等变 MLIPs（图 2 中间部分）通过显式引入矢量或张量表征来保持物理对称性，在预测精度和数据效率上通常优于不变模型[12]。该类模型分为 2 条技术路径：一是以 EGNN、PaiNN 为代表的基于位置矢量的低阶等变模型；二是基于高阶张量的模型，包括采用 SO(3) 群表示的 NequIP、MACE 以及为了降低计算开销而降阶至 SO(2) 卷积的 eSCN、UMA 等。等变的 MLIP 方法严格遵循物理对称性，通过在模型中引入不可约表示和张量运算来保证对称约束，从而在预测精度上往往优于不变 MLIPs 方法[19-42,44-58]，但当角动量阶数进一步增加时，其计算开销和内存开销成为瓶颈[79]。

无对称性约束 MLIPs（图 2 最下面部分）不施加硬性架构约束[12]，主要通过数据增强（如 ForceNet）来诱导模型学习对称性。无保守性约束的 MLIPs 分别拟合能量和力。这类模型结构极其灵活且推理速度快，但由于缺乏严格的物理一致性保障，在长时动力学模拟中可能存在能量守恒等稳定性隐患[26]。

上述分类主要从对称性约束和几何表征方式概括 MLIPs 的模型架构差异，其特点归类在表 1 中。不同类型 MLIP 模型的比较归纳在表 2 中。

目前，MLIPs 的适用范围已从无机材料和小分子体系进一步推广到蛋白质等生物大分子体系。在生物分子场景中，MLIP 成为生物大分子体系模拟中的一个环节。例如 AI2BMD 展示了“几何深度学习势函数+蛋白片段化+混合溶剂处理”的系统化路线，其

Table 1 Classification and Characteristics of MLIP Models

表 1 MLIPs 模型的分类及特点

不变专用 MLIPs[年份]	特点	不变预训练 MLIPs[年份]	特点
HDNNP[2007]	基于手工构造的原子环境描述符和局域能量分解，是早期高维神经网络势的代表。	CHGNet[2023]	基于 Materials Project 轨迹数据预训练，引入电荷/磁矩相关信息，适合材料结构弛豫和能量、力、应力预测。
DeePMD[2018]	采用端到端嵌入和局域环境映射，兼顾 DFT 精度与大规模分子动力学效率。	MatterSim[2024]	以 M3GNet 为 backbone 的预训练模型，在多元素、多温度和压力范围的大规模 DFT 数据上预训练。
SchNet[2017]	通过连续滤波卷积和标量消息传递处理原子间相互作用，保持旋转、平移和置换不变性。	DPA v3[2025]	Deep Potential 系列的大原子模型，基于高阶线图 LiGS 与数据集/任务编码机制实现多任务、多数据源训练，提升跨体系泛化能力。
MEGNet[2019]	将原子、键和全局态统一建模，可同时表达局域结构和材料整体状态信息。	Grace-2L[2025]	面向通用材料模拟的预训练图网络势，侧重较大化学空间上的迁移预测。
DimeNet[2020]	引入方向消息传递，通过距离和键角增强几何表达能力。	MatRIS[2026]	面向材料体系的预训练 MLIP，强调统一数据集上的结构弛豫和多任务泛化。
GemNet[2021]	在距离、角度和二面角等几何量上构造更丰富的消息传递，增强多体相互作用表达。		
ComENet[2022]	利用较完整且高效的几何信息构造不变特征，在表达能力和计算效率之间折中。		
等变专用 MLIPs[年份]	**特点**	**等变预训练 MLIPs[年份]**	**特点**
TFN[2018]	早期张量场网络，基于球谐函数和不可约表示实现 SE(3) 等变。	Spherical Channel Network[2022]	基于球面通道表示的等变模型，常用于催化和表面相关数据集中的预训练势建模。
GVP-GNN[2020]	同时处理标量和向量特征，使几何向量随旋转协同变换，适合三维结构建模。	MACE-MP[2023]	基于 MACE 架构在材料数据上预训练，利用 ACE 高阶多体表达提升材料体系迁移能力。
PaiNN[2021]	通过标量-向量交互实现 E(3) 等变消息传递，提高力和方向敏感性质预测能力。	eSCN[2023]	Spherical Channel Network 的高效版本，利用 SO(2)/球面通道思想降低等变计算开销。
EGNN[2021]	直接更新原子坐标并保持 E(n) 等变性，结构相对简洁、计算开销较低。	SevenNet[2024]	可扩展的等变图神经网络势，面向材料体系预训练和高效分子动力学应用。
ClofNet[2022]	借助局域坐标框架编码三维几何信息，在保持等变性的同时增强方向表达。	eSEN[2025]	面向大规模原子体系的高效等变模型，强调预训练后的跨体系泛化和推理效率。
SEGNN[2022]	采用可转向表示进行 E(3) 等变图消息传递，可系统处理高阶几何特征。	Allegro-OAM-L [2025]	基于 Allegro 的大规模预训练势，强调局域等变多体相互作用和大体系可扩展性。
Allegro[2023]	局域等变多体模型，避免全局消息传递，具有较好的大规模并行和 MD 扩展能力。	MACE-OFF[2025]	面向离平衡构型和分子体系的 MACE 预训练力场，适合有机/分子模拟场景。
MACE[2022]	将原子簇展开与等变消息传递结合，可表达高阶多体相互作用。	Nequix[2025]	基于 NequIP/等变图网络思想的预训练势，强调等变结构带来的数据效率和力预测精度。
NequIP[2022]	典型 E(3) 等变神经网络势，数据效率高，在小数据高精度力场训练中表现突出。	EqFlash[2025]	面向快速推理的等变预训练模型，侧重精度、速度和大规模模拟之间的平衡。
TensorNet[2023]	使用张量特征表达原子局域环境，通过张量运算保持旋转等变性。	UMA[2025]	通用原子模型，依托大规模、多来源数据预训练，面向广泛材料和分子任务。
SaVeNet[2023]	结合标量和向量信息进行消息传递，在表达方向相互作用的同时控制模型复杂度。	ORB v3[2025]	弱旋转等变的预训练原子模型，更多依赖数据规模、训练目标和约束条件获得泛化能力，适合高通量预测与结构优化任务。
VisNet[2024]	通过几何向量和标量特征交互提升三维原子体系的能量与力预测。		
ETNN[2025]	使用欧氏张量特征描述局域环境，强化高阶几何相互作用表达。		
无对称性约束 MLIPs[年份]	**特点**	**无保守性约束 MLIPs[年份]**	**特点**
ForceNet[2021]	直接学习原子受力，可借助数据增强学习几何对称性；推理灵活，但严格能量守恒不一定由结构保证。	GemNet-OC[2022]	面向 OC20 大规模催化数据训练的方向消息传递模型
FAENet[2023]	通过 frame averaging 等策略引入几何一致性，在不完全依赖硬编码等变结构的情况下提升稳定性。	EquiformerV2[2024]	提供直接力预测检查点。同时属于等变模型，为改进的等变 Transformer 架构，采用高阶表示并用 eSCN/SO(2) 卷积提升效率。

注：表 1 仅列举部分具有代表性的模型，用于说明相关技术路线的发展脉络，并非对该领域所有模型的完备总结。

势函数基于 ViSNet。由于生物大分子体系的模拟需考虑构象空间覆盖、长程相互作用、溶剂环境等系统因素，因此其采用 MLIP 进行模拟时需先片段化。其中 AI2BMD 利用蛋白质由有限种氨基酸组成的结构先验，将任意蛋白质拆解为 21 类蛋白单元，以此降低 DFT 数据生成与模型训练难度。后续再进行长程相互作用补偿和溶剂耦合等处理[86]。

3 MLIPs 模型的训练优化

3.1 优化器算法

训练 MLIPs 时最常用的方法为梯度下降法，形式上可表示为式(14)，其中权重的更新沿着负梯度方向以一定的步长η移动，典型的代表是 SGD、

Table 2 Comparison of Different Types of MLIPs

表 2 不同类型的 MLIPs 比较

类型	对称性处理方式	代表模型	主要优势	主要局限
不变 MLIPs	通过距离、角度、二面角、标量消息或手工描述符保持平移、旋转和置换不变性	HDNNP、DP、SchNet、DimeNet、GemNet、ComENet、CHGNet、MatRIS 等	结构相对简单，数值稳定，训练和推理效率较高	标量化表示可能削弱方向信息和高阶多体相互作用表达
等变 MLIPs	显式维护矢量/张量特征，使特征随旋转按群表示协同变换	PaiNN、EGNN、NequIP、MACE、Allegro、eSCN、UMA 等	物理归纳偏置明确，数据效率高，力和方向敏感性质预测能力强	张量积、球谐函数和不可约表示带来较高计算和显存开销，模型实现和部署较为复杂
无约束 MLIPs	不在网络结构中严格强制对称性，主要依赖数据增强	ForceNet、FAENet 等	架构灵活，推理速度快	严格能量守恒和旋转/平移一致性不一定由 MLIP 网络结构保证，长时间 MD 稳定性和外推可靠性需额外评估
	直接力预测	Equiformer V2、GemNet-OC 等		

Adam 优化器[87-88]。在保证收敛精度的前提下增大训练的批大小(batchsize)，可以提高模型的训练效率。然而专用 MLIPs 中如 DP，当采用 Adam 优化器增加 batchsize 时，往往伴随着几乎同比例迭代次数的增加[89]。

$$w_{t+1}=w_t-\eta\nabla L(f(x_i;w_t),y_i)\ 。\tag{14}$$

以扩展卡尔曼滤波(extended Kalman filter, EKF)理论为基础得到的优化器算法相比一阶优化方法具有更快的收敛速度[90]。EKF 于 1989 年就已用于神经网络权重更新[91]，并在基于手工描述符的专用 MLIPs 模型训练中得到应用[92]，也称为全局的 EKF。随着 MLIPs 模型参数量的增加，全局 EKF 在更新权重时，计算和内存开销成为瓶颈。为此，Hu 等人[23]通过将大参数量的层拆分、邻近的小参数量的层合并，对全局的参数矩阵进行解耦合，得到对角块形式的误差协方差矩阵，将全局 EKF 计算复杂度($O(N^2)$)降低至 $O(N\cdot N_b)$(此处 N 为神经网络的权重数，N_b 为重组后的块大小)。基于 EKF 的优化器在训练专用 MLIPs 时，如 DP、二体三体余弦基函数的势能模型(2B3B-Cosine)、二体三体高斯基函数模型(2B3B-Gaussian)、多项张量势(MTP)以及谱邻域分析势(SNAP)时，取得了 8~10 倍的训练加速效果[93]。

3.2 算子优化

除了算法层面的改进，高性能计算在提高 MLIPs 训练的计算效率上发挥着重要作用。MLIPs 模型中均包含邻居列表构建流程，高效实现邻居列表计算可降低训练的成本。部分 MLIPs 采用无框架的实现方法，牺牲深度学习框架带来的编程上的灵活性，以实现更高的计算效率。然而随着 MLIPs 在模型结构上的多样性、复杂性，借助深度学习框架可为用户提供更友好的编程体验。同时为了兼顾计算效率，人们主要关注计算访存上存在瓶颈的算子，通过计算图重构、算子融合的方式来提高训练效率并减少内存占用。相较于人为的检测计算图中存在的冗余操作并消除和手工编写算子，自动编译工具可实现类似功能并达到加速效果，同时还降低了代码维护和编写难度，提高对不同硬件架构的可移植性。除此之外，一些专用算子库可以进一步提升 MLIPs 的训练效率。下面分别介绍邻居列表构建/无框架的 MLIPs 实现、计算图重构/算子融合、自动编译工具、专用算子库。

1)邻居列表构建/无框架的 MLIPs 实现。邻居列表构建是 MLIPs 训练中最常规的操作之一。常用的邻居搜索算法包括 Verlet 表[94]、链表(cell-linked list)以及线性缩放方法等。得益于 CUDA 更直接的底层内存控制、更强的编译器优化，且无需 Python 调度以此减小解释器开销，DeePMD-kit 和 GPUMD 等高性能 MLIPs 在邻居列表构建中使用了 CUDA 进行实现[47]。GPUMD 采用无框架的训练方式，用 CUDA 在 GPU 上实现端到端训练。GPUMD 采用以原子为粒度的 CUDA 线程映射策略，在对应 CUDA kernel 中将一个线程映射到一个中心原子上。其中一个 kernel 用于处理该原子的局域环境相关计算。在该 kernel 中，与中心原子 i 对应的线程将遍历 i 的邻居原子，根据 NEP 中的解析函数构建局域环境描述符，并在同一 kernel 内完成前向神经网络计算，得到该原子的 E_i 及其对描述符分量的导数。径向和角向描述符相关的力及 virial 计算被进一步拆分到后续 CUDA kernels 中完成。GPUMD 中以原子为中心的计算组织方式减少了中间数据搬移和线程间通信需求，有利于提高 GPU 上 NEP 模型的计算效率[43]。无框架的实现可为 MLIPs 的计算过程做极限融合与内存优化，是一种高效的实现方式。然而一定程度上牺牲了 MLIPs 结构设计的灵活性。

2)计算图重构/算子融合。依赖 PyTorch 等深度学习框架进行 MLIPs 训练时，会启动大量碎片化的小算子，导致不可忽视的算子启动开销和不必要的内存 I/O 开销。计算图重构与算子融合技术能够消

除深度学习计算图中冗余的张量操作，通过降低算子启动频率与显存访存开销，有效解决计算效率低下的问题。FastDimeNet++通过手工重构计算图并融合算子以提升算力利用率。由于原始的 DimeNet++模型在原子-边-角度三元组等不同层级间存在大量冗余的张量扩展操作，导致触发了超一万个的小 kernel，GPU 利用率不足 5%。FastDimeNet++引入了一系列训练优化的融合策略，包括共享输入的水平融合、共享输出的水平融合以及批量矩阵乘法等 3 种线性层融合方案，将多维度的线性变换融合为大型高并行度的算子[95]。Hu 等人[89]发现 DP 模型中对称性算子的实现会启动大量算子，用 CUDA 实现了描述符计算算子的前向计算、一阶导数计算和二阶导数的计算，通过融合中间张量以减少 CPU 与 GPU 之间的数据搬运。

对于预训练 MLIPs 中的 CHGNet，Zhou 等人[96]发现 CHGNet 的训练效率受限于计算图的串行依赖关系与大量细粒度 kernel。FastCHGNet 消除 CHGNet 中原子特征、键特征的显式依赖关系，实现原子、键、角度特征的并发更新。同时 FastCHGNet 把逐样本的串行图处理方式改为按照 batch 批量处理，提高 GPU 的利用率。Zhou 等人[96]重写了 Group-Softmax、SiLU、LayerNorm 等算子以加速预训练模型 MatRIS 的训练过程。

3）自动编译工具。即时（just in time，JIT）编译技术作为运行时常用的优化手段，可将高层的计算图在执行时转换为低层优化代码。在 MLIP 训练中，JIT 常用于优化小 kernel 的调用，实现对重复张量算子的自动融合。FastDimeNet++在实现中大量采用 PyTorch 的 JIT 编译，结果表明在不影响能量和力计算物理一致性的前提下，结合 JIT 可使训练速度提升 2~4 倍[95]。Tan 等人[75]在 NequIP 模型上用 torch.compile 接口将连续的逐元素运算、高阶张量积和维度置换等合并为单一 GPU kernel，大幅降低了算子在 CPU 端的启动延迟和内存 I/O 开销。实验表明该方法在保证数值精度和能量-力守恒一致性的情况下，torch.compile 相比于 TorchScript 的训练加速了 2.4~5.0 倍。

4）专用算子库。针对含不可约表示的等变 MLIPs（如预训练模型 MACE），其中包含昂贵的张量积操作。e3nn 是等变网络的 Python 标准实现库，包括 irreps 表达、球谐函数表示、CG 系数耦合等操作[97]。cuEquivariance 把 e3nn 常用等变算子在 GPU 上进行高效的实现，主要包括张量积（Tensor-Product）、对称规约（contraction）等算子，是面向 $O(3)$/SO(3) 等变表示（irreps）运算的常用加速库。FlashTP 进行了进一步的优化，把张量积层内部的多步算子尽可能合并为少数算子，并利用 CG 系数矩阵的稀疏性，用 COO 格式存储 CG 系数矩阵里非零元的索引，以降低存储开销同时减少无效的计算[78]。

3.3 负载均衡

传统 data loader 往往采用固定大小的 batch，即每个 batch 包含固定数量的样本。然而预训练 MLIPs 在预训练过程中用到的预训练数据集，如 OMat24、MPtrj、sAlex 等，往往数据量大，每个样本的图规模小，一般为几十到几百原子/节点。Zhou 等人[96]发现以 MPtrj 为例的预训练数据集中每个样本对应的原子数、边数、角度数呈现明显的长尾分布。平均原子/边/角数量分别为 28.5、1 717.4、9 091.6，方差达到 6.82×10^2、2.96×10^6、1.14×10^8，意味着少量大图会拉长一次迭代的计算时间，造成计算快的小图需等待大图的计算结果，进而降低多 GPU 训练效率。因此，面向预训练 MLIPs 的训练过程，须在 mini-batch 的构造阶段引入负载均衡策略，以保证更均衡的资源使用，最小化等待时间。由于预训练数据集中包含由原子数、边数、角度数组成的多种计算负载，下面分别介绍由基于不同物理量进行负载均衡的策略。

1）基于原子的负载均衡策略。预训练 MLIPs 在训练时进行负载均衡的核心思想是使每个计算设备承担相近的工作量，避免因数据样本大小不均导致的闲置和同步等待。其中 Firoz 等人将 mini-batch 的构造抽象为一个多目标装箱问题，以原子数量代表图规模/样本大小，并设定每个箱（batch）的容量上限（即每个 batch 最大允许的总原子数）。整体思路为按图规模降序排序，将最大图优先分配，然后采用循环轮转的方式将剩余图逐一尝试放入当前 batch 列表中。如果某个 batch 还能容纳当前图（剩余容量足够），则将其加入该 batch，并减少相应剩余容量。若无法容纳，则暂时将该 batch 标记为“已满”。采用该负载均衡策略，Firoz 等人[98]表明在包含 2.6 百万分子、原子数跨度从 1 到 768 原子的混合数据集上，使用 740 块 GPU 训练时，每个 epoch 时间从 12 min 缩短到 2 min。

2）基于原子-边的负载均衡策略。Koker 等人[76]发现节点数只能粗略表征模型计算量；在 NequIP 模型中，边数决定了消息计算的次数，因此更直接影响实际 FLOPs 与显存占用。因此在 Nequix 中，同时对总节点数和总边数设上限，动态把样本装进一个 batch，直到接近这 2 个的上限为止。Nequix 兼顾节

点数和边数，能避免出现节点不多但边数极多的图使迭代过程中负载不均衡[76]。

3）基于原子-边-角度的负载均衡策略。FastCHGNet将训练负载定义为原子数、键数与角数的总和，先按该负载对样本排序，再采用轮转方式将样本依次分配到各 GPU，从而让每张 GPU 在一个迭代步内的mini-batch 整体上更接近，缓解 batch 内负载不均问题。预训练模型 MatRIS 由于建模了原子图和键图，模型训练过程中也包含原子、边和角度信息[96]。MatRIS在训练时先对全数据集做全局的 shuffle，再切分为多个 chunk，并将各 chunk 依次送入“排序+装箱”的队列里进行负载均衡的处理。通过全局 shuffle 和分 chunk保留训练样本随机性，在 chunk 内再做负载均衡。

3.4 两阶段训练

预训练 MLIPs 用于 MD、声子谱等下游任务时，能量守恒是必须满足的基本物理约束。力的计算应当来自势能对原子坐标的梯度，以此保证轨迹长期积分的稳定性与物性计算可信度。然而在训练过程中，保守力、高精度的计算过于昂贵。两阶段训练是指在第 1 阶段，先用相对便宜的配置进行训练，在第 2 阶段再用物理正确但更贵的配置进行训练。

1）直接力与保守力的两阶段训练。eSEN 在一阶段训练中在骨干网络上添加一个直接力的预测头，以更低成本进行高效表征学习。随后丢弃直接力的预测头，在同一骨干上切换为保守力的计算。实验进行以下两阶段配置：先进行 60 个 epoch 的直接力训练，再进行 40 个 epoch 的保守微调。两阶段训练方案的损失比从头训练 100 个 epoch 的保守模型更低，两阶段训练方案下得到的最终的预训练 MLIPs将总的训练时间降低约 40% [80]。采用直接力的第1 阶段训练可视为“表示学习的热身/预热”；采用保守力的第 2 阶段训练可视为“物理约束下的参数微调”。

2）BF16 与 FP32 的两阶段训练。UMA 在采用上述的直接力与保守力的两阶段训练外，进一步将数值精度引入两阶段训练过程中。第 1 阶段以直接力方式进行训练时，采用低精度 BF16 进行训练。第 2 阶段以保守力方式进行训练时，采用 FP32 进行训练。在预训练 UMA 时，尽管第 1 阶段采用 BF16 相对FP32 会有一定精度损失，但第 1 阶段预训练的过程中具有更为稳定的收敛性。在第 2 阶段，将数值精度切换到 FP32 后，由 1 阶段训练 BF16 带来的精度退化问题可恢复，且仅需少于 1% 的迭代步数即可恢复精度[16]。

3.5 模型轻量化

MLIP 目前采用的模型轻量化策略主要包括模型剪枝、模型蒸馏和模型量化。模型剪枝从结构层面删除冗余层以直接减少模型参数量和 FLOPs。模型蒸馏通过高精度教师模型指导小型学生模型学习，使小型学生模型在较低推理成本下逼近教师模型。模型量化则通过降低权重、激活或中间张量的数值位宽来减少显存占用并提升低精度硬件上的吞吐率。

1）模型剪枝。在 MLIP 中，通常采用结构化剪枝方法，例如移除模型中的部分交互层或消息传递层来减少模型参数规模。Ghunaim 等人[99]针对分子性质预测的模型 JMP 分析不同消息传递层对下游任务的贡献，发现靠后的消息传递层的精度收益逐渐减小，因此在微调前直接从预训练 JMP-L 中删除靠后的部分消息传递层。该方法通过缩短消息传递深度降低计算量，不改变单层内部权重结构，属于较粗粒度的层剪枝。Kong 等人[100]在删除低贡献的消息传递层后，进一步结合图划分和多 GPU 分布式推理，使 MLIP 模型能够扩展到百万原子级模拟。这类剪方法容易在常规 GPU 上获得实际加速，但也可能因消息传递深度降低而削弱长程信息传播和复杂局域环境表达能力。

2）模型蒸馏。模型蒸馏通常以参数量大的高精度教师模型指导参数量小的学生模型进行训练，使学生模型在较低推理成本下逼近教师模型的预测分布[101]。MLIPs 在蒸馏时，最直接的损失函数的设计是考虑能量误差和力的误差。已有工作尝试通过对齐教师模型与学生模型的节点特征、边特征得到高精度的学生模型，以加速分子图神经网络推理[102]。由于 MD 轨迹的稳定性受到势能面梯度及其局部曲率的影响，学生模型若仅拟合静态能量和力，可能在长时间积分中出现能量漂移、构象分布偏差或势垒预测错误。因此，Amin 等人[103]进一步提出面向MLIP 的 Hessian-based 蒸馏策略，通过教师模型的能量的 Hessian 矩阵指导学生模型学习更准确的势能面，从而构建参数量更小的学生模型。

3）模型量化。模型量化通过降低权重、激活或中间张量的数值精度来减少显存占用并提高低精度硬件上的吞吐率[104-106]。量化通常不改变模型结构和计算的浮点操作次数（FLOPs），而将神经网络中的高精度数值表示替换为低比特宽度的数据表示。一方面能够显著减少模型存储开销和显存占用，同时降低乘加运算的计算成本。MLIPs 的量化需要格外关注数值稳定性，因为原子受力往往需要通过能量对

原子坐标求梯度得到，低精度数值表示可能在梯度计算、归一化、长时间分子动力学模拟等过程中被放大。现有 MLIP 训练已经尝试使用混合精度思想。例如前文提到的 UMA 采用 BF16 与 FP32 的两阶段策略，先利用 BF16 提升直接力训练效率，再在保守力阶段切换至 FP32 恢复数值精度。

4 MLIPs 模型的评估

由于现阶段的 MLIPs 种类多，仅根据能量/力的精度来评价一个 MLIP 是不够的[26]。如何综合全面地评价一个 MLIP 是一个有挑战性的工作。为实现公开、公平的对比评测，研究社区构建了多类公开数据集与评价指标体系。下文按任务类型列出常用评估基准。

本文列举现有的 3 类常用评估基准：第 1 类是能量和力的基础预测任务（4.1 节），主要使用 MD17/rMD17、MD22、QM9、ANI-1x、SPICE、TorsionNet-500 等数据集，考察模型对势能面及其梯度的基本拟合能力。第 2 类是无机材料评估基准（4.2 节），有代表性的为 Matbench Discovery，此外还包括 MatCalc-Bench、MDR phonon、LAMBench 等测评任务，评估模型在无机晶体结构弛豫、形成能、热力学稳定性、弹性性质和声子谱等任务中的可靠性。第 3 类是有机分子的评估基准（4.3 节），有代表性的为 FAIR Chemistry Leaderboard，此外还包括 OMC 等评测任务，可系统评估 MLIP 在有机分子、生物分子、电解液、金属配合物和中性有机物等化学域上的表现。

4.1 分子能量/力基础任务评估

常用的分子数据集包括 MD17 系列（小分子的分子动力学轨迹数据）、MD22（包含更大体系如三肽、脂类等）以及 QM9（约 13 万小分子单点能量和性质）等。此外，TorsionNet-500（500 个分子构象的扭转势能曲线）用于评估模型预测势垒高度的准确性[107]。分子数据集的评估侧重能量和原子力的准确度，常采用平均绝对值误差（MAE）和均方根误差（RMSE）衡量能量（单位一般为 meV 或 kcal/mol）和力（meV/Å 或 kcal/mol·Å）的偏差。TorsionNet-500 不仅比较扭转路径上的能量误差，还评估势垒高度误差并统计势垒误差大于 1 kcal/mol 的分子数。

专用 MLIP 模型如不变的 SchNet 在 MD17 数据集上原子受力的 MAE 为 50~100 meV/Å，DimeNet++ 在 QM9 上大幅提高能量预测精度（例如基态能量 U0 的 MAE 仅约 6.3 meV）。NequIP 通过引入 E(3) 等变，在 MD17 数据集上仅采用 1 000 个训练样本时，aspirin 分子的力的 MAE 仅约 20 meV/Å量级，比 sGDML 等模型展现出更低的误差。MACE 在 rMD17 数据集上达到更低的误差（绝大多数小分子能量 MAE 小于 10 meV，力 MAE 小于 40 meV/Å）。对于扭转势预测，24 层的 DPA3 在 TorsionNet-500 上各项误差优于 MACE、AIMNet2，展现了强大的泛化能力。目前来看，诸如 NequIP、MACE 等模型在 MD 和 QM 系列的中小分子数据集的预测上已可接近 DFT 精度。

4.2 无机晶体性质预测基准

针对固体材料，关注晶体结构的能量与性质预测，包括平衡态的形成能以及非平衡态的力学和热力学属性等[108]。常用数据集有 Materials Project 晶体数据（MPtrj，包含 158 万个结构帧）和 MatPES-PBE 数据集。测评排行榜包括 Matbench Discovery[19]、MatCalc-Bench[108]、LAMBench[109] 等。

Matbench Discovery 是当前较为广泛使用的模型测评排行榜之一，是一个主要面向高通量发现新稳定无机晶体的评测框架和在线榜单，用来比较各种 MLIPs 在材料发现任务上的表现。Matbench Discovery 采用 MPtrj 轨迹数据作为训练集之一，每个轨迹数据包含能量、力、应力、磁矩。另外一个规模更大的训练数据集为 OMat24+MPtrj+sAlex。该榜单引入分布显著偏移的 WBM（Wang-Botti-Marques，约 25.7 万个候选无机晶体）数据集作为测试集[110]，用于考察模型在外推材料空间中的稳定性与鲁棒性。Matbench Discovery 设计了热力学稳定性预测、结构弛豫质量与晶格热导率预测等多个任务，并通过 $F1$ 分数、发现加速因子（discovery acceleration factor，DAF）等指标刻画模型的材料发现能力。Matbench Discovery 目前包含 40 余个预训练 MLIP 检查点文件。由于 Matbench Discovery 设置了众多的指标，为了方便对模型进行排序，Matbench Discovery 选取了一个综合指标 CPS（combined performance score），并用该综合指标 CPS 衡量模型在真实材料发现流程中的整体效果。其计算基于 3 项互补指标：稳定性指标 $F1$（衡量在有限 DFT 预算下识别稳定晶体的能力）、弛豫结构的 RMSD（衡量预测基态几何的准确性）以及热导率误差 K_{SRME}（衡量势能面曲率/声子相关性质的可靠性）。

总体来说，Matbench Discovery 要求模型对 WBM 候选无机晶体进行结构优化并预测形成能，主要考察结构弛豫质量和稳定材料筛选能力。MatCalc-

Bench 设置了多项评价指标，能够从结构弛豫精度、形成能预测、弹性模量估计以及势能面局部曲率刻画等角度，系统评估模型对材料结构与性质关系的建模能力。MDR Phonon 基准以最大声子频率、振动熵、Helmholtz 自由能和定容热容等指标为评价对象，系统考察模型对晶格振动、声子热力学性质及势能面二阶曲率的刻画能力。

4.3 分子性质预测基准

FAIR Chemistry Leaderboard/OMol25 leaderboard 是一个有代表性的分子体系评估基准[111]，其基础任务是 S2EF（structure to energy and forces），即给定原子结构预测总能量和原子力。此外，该 leaderboard 还设置了多个面向化学应用的下游评估任务。1）Ligand pocket 评估蛋白-配体相互作用。2）Ligand strain 评估配体应变能和构象 RMSD。3）Conformers 关注最低能构象识别和构象集合 RMSD。4）Protonation 评估不同质子化态之间的能量差。5）Distance scaling 同时考察短程和长程分子间相互作用的能量/力差。6）IE/EA 评估电离能和电子亲和能相关能量/力差。7）Spin gap 关注不同自旋态之间的能量差。这些任务覆盖构象、非共价相互作用、质子化、电荷转移和自旋态等传统能量/力测试集难以充分反映的化学问题。OMol25 leaderboard 能够检验模型在有机分子、生物分子、电解液等不同化学领域问题上的迁移能力。

中国科学技术大学[112]提出一个面向有机分子晶体（organic molecular crystals, OMC）的 MLIPs 测评榜单（简称 OMC-Bench）。该榜单不仅比较能量/力的预测误差，同时强调 OMC 的关键挑战不仅在于局域相互作用的拟合精度，还包含大体系泛化、共晶/多晶型结构稳定性、声子与温度相关自由能等下游任务的一致性与可迁移性。

从任务设置看，榜单分为 5 个测评任务，5 个任务覆盖 OMC 建模中最常用的工作流链条。1）以大原子数的 OMC 体系（原子数 $N \geqslant 300$）为对象，核心指标为力与应力的 MAE。2）共晶结构匹配，主要关注弛豫后结构与参考结构的一致性。3）多晶型能量排序，聚焦多晶型（polymorph）的相对能量排序是否正确，采用 KPE 衡量排序误差，并用 FRA（全排序正确家族占比）评价在多家族上的整体成功率，把能量回归误差转化为对材料筛选更直接的排序任务。4）与晶格动力学与热力学相关的声子计算，评估最高声子频率的 MAE，并用归一化声子态密度的 KL 散度衡量整体谱形差异，在此基础上，进一步比较由声子导出的熵等指标的 MAE。5）升华焓的预测，直接对接实验可观测量，强调模型在固相晶体与气相分子间的热力学一致性。目前榜单上已涵盖数十个预训练 MLIPs，如 AtomBit、CHGNet、SevenNet-13i5、MACE-MPA-0、ORB v3、Matris-OAM 等。

5 总结与展望

机器学习原子间势（MLIPs）通过学习高维势能面（PES），在保持接近量子力学精度的同时，将分子动力学模拟推进到更大时空尺度，成为从头算 MD 与经验力场之间兼顾精度与效率的关键折中方案。从对称性归纳偏置出发，MLIPs 可系统划分为不变、等变与无约束 3 类。与此同时，MLIPs 的范式正在由专用 MLIPs 走向预训练 MLIPs。目前模型与训练技巧处于协同演进的状态。一方面，高效训练技巧决定了模型迭代速度：训练加速能够缩短模型迭代的周期。另一方面，模型的设计同样决定训练技巧的研究重点。MLIPs 模型已经处于快速扩展、多路线并行发展的阶段，仅凭单一数据集或单一误差指标已难以综合评估 MLIPs 模型的能力。因此，构建公平、客观、可复现的测评榜单至关重要。一个合格的榜单应基于统一的评测协议，覆盖多任务、多分布（ID/OOD）的测评，以此体现出模型的可靠性、可迁移性。

从 MLIPs 的发展趋势看，MLIPs 正在由面向特定元素、特定相或特定化学体系的专用模型，逐渐走向跨元素、跨结构类型和跨任务的预训练模型。然而，当前真正具备跨领域、多任务能力的 MLIP 模型仍然较少，UMA、MatRIS-MoE[113]等工作初步展示了混合专家模型在多数据源、多体系任务中的潜力。未来，随着高精度训练数据集的进一步扩充，有望设计更强大的多任务 MLIPs 模型，使得其能学习到更本征的原子嵌入表示。例如即便一些高能态数据集没有在训练数据集中出现，也能实现较为准确的预测。

从未来对 DFT 数据要求的角度来看，随着算力资源变得更加丰富，设计能力更强的 MLIPs 模型变得可行。然而模型参数量的变大对高质量 DFT 数据量的需求将急剧增加。因此，未来需要建设统一标准的更大规模 DFT 数据体系，包括清晰的数据生成协议、覆盖平衡态与非平衡态结构的采样策略、可追溯的计算参数记录。只有在数据标准化和数据质量控制进一步完善的基础上，MLIPs 才能向更通用的原子模型发展。

从测评基准的角度来看，评测体系也需要从分散的单项 benchmark 走向更统一、更全面的评价基准。未来更理想的评测体系应同时覆盖精度、效率和泛化性。目前的榜单更多地报告精度和泛化性，然而在实际下游任务上，效率问题极为关键。榜单也可纳入推理吞吐、显存占用、大原子尺度系统模拟下的可扩展性等效率指标。效率指标的测评可以方便用户在不同任务不同精度需求下选用最适合的 MLIP 模型。

作者贡献声明：胡思宇负责方法整体设计、调研及论文撰写；谭光明和贾伟乐提出指导和修改意见。其中谭光明（tgm@ict.ac.cn）和贾伟乐（jiaweile@ict.ac.cn）为共同通信作者。

参 考 文 献

[1] Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces[J]. Physical Review Letters, 2007, 98(14): 146401

[2] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry[C]//Proc of the 34th Int Conf on Machine Learning. Sydney: PMLR, 2017, 70: 1263−1272

[3] Jia Weile, Wang Han, Chen Mohan, et al. Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning[C]//Proc of Int Conf for High Performance Computing, Networking, Storage and Analysis. Piscataway, NJ: IEEE, 2020: 1−14

[4] Guo Zhuoqiang, Lu Denghui, Yan Yujin, et al. Extending the limit of molecular dynamics with ab initio accuracy to 10 billion atoms[C]//Proc of the 27th ACM SIGPLAN Annual Symp on Principles and Practice of Parallel Programming. New York: ACM, 2022: 205−218

[5] Parr R G, Yang Weitao. Density-Functional Theory of Atoms and Molecules[M]. Oxford: Oxford University Press, 1995

[6] MacKerell A D. Empirical force fields for biological macromolecules: Overview and issues[J]. Journal of Computational Chemistry, 2004, 25(13): 1584−1604

[7] Zhang Linfeng, Wang Han, Car R, et al. Phase diagram of a deep potential water model[J]. Physical Review Letters, 2021, 126(23): 236001

[8] Bonati L, Parrinello M. Silicon liquid structure and crystal nucleation from ab initio deep metadynamics[J]. Physical Review Letters, 2018, 121(26): 265701

[9] Marcolongo A, Binninger T, Zipoli F, et al. Simulating diffusion properties of solid-state electrolytes via a neural network potential: Performance and training scheme[J]. ChemSystemsChem, 2020, 2(3): e1900031

[10] Merchant A, Batzner S, Schoenholz S S, et al. Scaling deep learning for materials discovery[J]. Nature, 2023, 624(7990): 80−85

[11] Yang Han, Hu Chenxi, Zhou Yichi, et al. MatterSim: A deep learning atomistic model across elements, temperatures and pressures[J]. arXiv preprint, arXiv: 2405.04967, 2024

[12] Duval A, Mathis S V, Joshi C K, et al. A hitchhiker's guide to geometric GNNs for 3D atomic systems[J]. arXiv preprint, arXiv: 2312.07511, 2023

[13] Yan Keqiang, Bohde M, Kryvenko A, et al. A materials foundation model via hybrid invariant-equivariant architectures[J]. arXiv preprint, arXiv: 2503.05771, 2025

[14] Devereux C, Smith J S, Huddleston K K, et al. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens[J]. Journal of Chemical Theory and Computation, 2020, 16(7): 4192−4202

[15] Deng Bowen, Zhong Peichen, Jun K, et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling[J]. Nature Machine Intelligence, 2023, 5(9): 1031−1041

[16] Wood B M, Dzamba M, Fu Xiang, et al. UMA: A family of universal models for atoms[J]. arXiv preprint, arXiv: 2506.23971, 2025

[17] Batatia I, Benner P, Chiang Y, et al. A foundation model for atomistic materials chemistry[J]. The Journal of Chemical Physics, 2025, 163(18): 184110

[18] Zhang Duo, Liu Xinzijian, Zhang Xiangyu, et al. DPA-2: A large atomic model as a multi-task learner[J]. npj Computational Materials, 2024, 10(1): 293

[19] Riebesell J, Goodall R E A, Benner P, et al. A framework to evaluate machine learning crystal stability predictions[J]. Nature Machine Intelligence, 2025, 7: 836−847

[20] Barroso-Luque L, Shuaibi M, Fu Xiang, et al. Open materials 2024 (OMat24) inorganic materials dataset and models[J]. arXiv preprint, arXiv: 2410.12771, 2024

[21] Levine D S, Shuaibi M, Spotte-Smith E W C, et al. The open molecules 2025 (OMol25) dataset, evaluations, and models[J]. arXiv preprint, arXiv: 2505.08762, 2025

[22] Smith J S, Isayev O, Roitberg A E. ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost[J]. Chemical Science, 2017, 8(4): 3192−3203

[23] Hu Siyu, Zhang Wentao, Sha Qiuchen, et al. RLEKF: An optimizer for deep potential with ab initio accuracy[C]//Proc of the AAAI Conf on Artificial Intelligence. Washington, DC: AAAI, 2023, 37(7): 7910−7918

[24] Zhang Yuzhi, Wang Haidi, Chen Weijie, et al. DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models[J]. Computer Physics Communications, 2020, 253: 107206

[25] Morrow J D, Gardner J L A, Deringer V L. How to validate machine-learned interatomic potentials[J]. The Journal of Chemical Physics, 2023, 158(12): 121501

[26] Fu Xiang, Wu Zhenghao, Wang Wujie, et al. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations[J]. Transactions on Machine Learning

Research, 2023: 1−31

[27] Bihani V, Mannan S, Pratiush U, et al. EGraFFBench: Evaluation of equivariant graph neural network force fields for atomistic simulations[J]. Digital Discovery, 2024, 3(4): 759−768

[28] Zuo Yunxing, Chen Chi, Li Xiangguo, et al. Performance and cost assessment of machine learning interatomic potentials[J]. The Journal of Physical Chemistry A, 2020, 124(4): 731−745

[29] Behler J. Constructing high-dimensional neural network potentials: A tutorial review[J]. International Journal of Quantum Chemistry, 2015, 115(16): 1032−1050

[30] Behler J. Four generations of high-dimensional neural network potentials[J]. Chemical Reviews, 2021, 121(16): 10037−10072

[31] Behler J. Perspective: Machine learning potentials for atomistic simulations[J]. The Journal of Chemical Physics, 2016, 145(17): 170901

[32] Unke O T, Chmiela S, Sauceda H E, et al. Machine learning force fields[J]. Chemical Reviews, 2021, 121(16): 10142−10186

[33] Friederich P, Hase F, Proppe J, et al. Machine-learned potentials for next-generation matter simulations[J]. Nature Materials, 2021, 20: 750−761

[34] Ko T W, Ong S P. Recent advances and outstanding challenges for machine learning interatomic potentials[J]. Nature Computational Science, 2023, 3: 998−1000

[35] Anstine D M, Isayev O. Machine learning interatomic potentials and long-range physics[J]. The Journal of Physical Chemistry A, 2023, 127(11): 2417−2431

[36] Freitas R, Cao Yifan. Machine-learning potentials for crystal defects[J]. MRS Communications, 2022, 12: 510−520

[37] Noordhoek K, Bartel C J. Accelerating the prediction of inorganic surfaces with machine learning interatomic potentials[J]. Nanoscale, 2024, 16: 6365−6382

[38] Wu Zonghan, Pan Shirui, Chen Fengwen, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4−24

[39] Bronstein M M, Bruna J, Cohen T, et al. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges[J]. arXiv preprint, arXiv: 2104.13478, 2021

[40] Esteves C. Theoretical aspects of group equivariant neural networks[J]. arXiv preprint, arXiv: 2004.05154, 2020

[41] Nguyen Q T, Schatzki L, Braccia P, et al. Theory for equivariant quantum neural networks[J]. PRX Quantum, 2024, 5(2): 020328

[42] Han Jiaqi, Cen Jiacheng, Wu Liming, et al. A survey of geometric graph neural networks: Data structures, models and applications[J]. Frontiers of Computer Science, 2025, 19(11): 1911375

[43] Fan Zheyong, Wang Yanzhou, Ying Penghua, et al. GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations[J]. The Journal of Chemical Physics, 2022, 157(11): 114801

[44] Huang Yufeng, Kang Jun, Goddard W A, et al. Density functional theory based neural network force fields from energy decompositions[J]. Physical Review B, 2019, 99(6): 064103

[45] Novikov I S, Gubaev K, Podryabinkin E V, et al. The MLIP package: Moment tensor potentials with MPI and active learning[J]. Machine Learning: Science and Technology, 2020, 2(2): 025002

[46] Thompson A P, Swiler L P, Trott C R, et al. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials[J]. Journal of Computational Physics, 2015, 285: 316−330

[47] Wang Han, Zhang Linfeng, Han Jie. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics[J]. Computer Physics Communications, 2018, 228: 178−184

[48] Schutt K, Kindermans P J, Sauceda Felix H E, et al. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions[J]. Advances in Neural Information Processing Systems, 2017, 30: 991−1001

[49] Gasteiger J, Gross J, Gunnemann S. Directional message passing for molecular graphs[C]// Proc of the 8th Int Conf on Learning Representations. Addis Ababa: OpenReview.net, 2020: 1−13

[50] Gasteiger J, Becker F, Gunnemann S. GemNet: Universal directional graph neural networks for molecules[J]. Advances in Neural Information Processing Systems, 2021, 34: 6790−6802

[51] Gasteiger J, Giri S, Margraf J T, et al. Fast and uncertainty-aware directional message passing for non-equilibrium molecules[J]. arXiv preprint, arXiv: 2011.14115, 2020

[52] Liu Yi, Wang Limei, Liu Meng, et al. Spherical message passing for 3D molecular graph [C]// Proc of the 10th Int Conf on Learning Representations. Virtual: OpenReview.net, 2022:1−20

[53] Ji Shuiwang, Lin Yuchao, Liu Haoran, et al. ComENet: Towards complete and efficient message passing for 3D molecular graphs[J]. Advances in Neural Information Processing Systems, 2022, 35: 650−664

[54] Zhou Yuanchang, Hu Siyu, Zhang Xiangyu, et al. MatRIS: Toward reliable and efficient pretrained machine learning interaction potentials[J]. arXiv preprint, arXiv: 2603.02002, 2026

[55] Zhang Duo, Peng Anyang, Cai Chun, et al. A graph neural network for the era of large atomistic models[J]. arXiv preprint, arXiv: 2506.01686, 2025

[56] Zeng Jinzhe, Zhang Duo, Peng Anyang, et al. DeePMD-kit v3: A multiple-backend framework for machine learning potentials[J]. Journal of Chemical Theory and Computation, 2025, 21(9): 4375−4385

[57] Lysogorskiy Y, Bochkarev A, Drautz R. Graph atomic cluster expansion for foundational machine learning interatomic potentials[J]. arXiv preprint, arXiv: 2508.17936, 2025

[58] Tran R, Lan J, Shuaibi M, et al. The Open Catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts[J]. ACS Catalysis, 2023, 13(5): 3066−3084

[59] Satorras V G, Hoogeboom E, Welling M. E(n) equivariant graph neural networks[C]//Proc of the 38th Int Conf on Machine Learning. Virtual: PMLR, 2021, 139: 9323−9332

[60] Han Jiaqi, Huang Wenbing, Rong Yu, et al. Equivariant graph hierarchy-based neural networks[J]. Advances in Neural Information Processing Systems, 2022, 35: 9176−9187

[61] Huang Wenbing, Han Jiaqi, Rong Yu, et al. Equivariant graph mechanics networks with constraints[J]. arXiv preprint, arXiv: 2203.06442, 2022

[62] Schutt K, Unke O, Gastegger M. Equivariant message passing for the prediction of tensorial properties and molecular spectra[C]//Proc of the 38th Int Conf on Machine Learning. Virtual: PMLR, 2021, 139: 9377−9388

[63] Du Wenxuan, Zhang He, Du Yuanqi, et al. SE(3) equivariant graph neural networks with complete local frames[C]//Proc of the 39th Int Conf on Machine Learning. Baltimore: PMLR, 2022, 162: 5583−5608

[64] Zhang Yan, Cen Jiacheng, Han Jiaqi, et al. Improving equivariant graph neural networks on large geometric graphs via virtual nodes learning[C]//Proc of the 41st Int Conf on Machine Learning. Vienna: PMLR, 2024, 235: 58662−58679

[65] Battiloro C, Karaismailoglu E, Tec M, et al. E(n) equivariant topological neural networks[C]//Proc of the 13th Int Conf on Learning Representations. Singapore: OpenReview.net, 2025: 1−43

[66] Wang Yusong, Wang Tong, Li Shaoning, et al. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing[J]. Nature Communications, 2024, 15(1): 313

[67] Phillips A C. Introduction to Quantum Mechanics[M]. Hoboken: John Wiley & Sons, 2013

[68] Batzner S, Musaelian A, Sun L, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials[J]. Nature Communications, 2022, 13(1): 2453

[69] Thomas N, Smidt T, Kearnes S, et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds[J]. arXiv preprint, arXiv: 1802.08219, 2018

[70] Fuchs F, Worrall D, Fischer V, et al. SE(3)-transformers: 3D roto-translation equivariant attention networks[J]. Advances in Neural Information Processing Systems, 2020, 33: 1970−1981

[71] Tholke P, De Fabritiis G. TorchMD-Net: Equivariant transformers for neural network based molecular potentials[C]//Proc of the 10th Int Conf on Learning Representations. Virtual: OpenReview.net, 2022:1−20

[72] Batatia I, Csanyi G, Kovacs D P, et al. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields[J]. Advances in Neural Information Processing Systems, 2022, 35: 11423−11436

[73] Xu Z, Yu H, Bohde M, et al. Equivariant graph network approximations of high-degree polynomials for force field prediction[J]. Transactions on Machine Learning Research, 2024: 1−28

[74] Kovacs D P, Moore J H, Browning N J, et al. MACE-OFF: Short-range transferable machine learning force fields for organic molecules[J]. Journal of the American Chemical Society, 2025, 147(21): 17598−17611

[75] Tan Chuin Wei, Descoteaux M L, Kotak M, et al. High-performance training and inference for deep equivariant interatomic potentials[J]. arXiv preprint, arXiv: 2504.16068, 2025

[76] Koker T, Kotak M, Smidt T. Training a foundation model for materials on a budget[C]//Proc of the NeurIPS 2025 Workshop on AI for Accelerated Materials Discovery. San Diego: OpenReview.net, 2025: 1−8

[77] Park Y, Kim J, Hwang S, et al. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations[J]. Journal of Chemical Theory and Computation, 2024, 20(11): 4857−4868

[78] Lee S Y, Kim H, Park Y, et al. FlashTP: Fused, sparsity-aware tensor product for machine learning interatomic potentials[C]//Proc of the 42nd Int Conf on Machine Learning. Vancouver: PMLR, 2025, 267: 33143−33156

[79] Passaro S, Zitnick C L. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs[C]//Proc of the 40th Int Conf on Machine Learning. Honolulu: PMLR, 2023, 202: 27420−27438

[80] Fu Xiang, Wood B M, Barroso-Luque L, et al. Learning smooth and expressive interatomic potentials for physical property prediction[J]. arXiv preprint, arXiv: 2502.12147, 2025

[81] Rhodes B, Vandenhaute S, Simkus V, et al. Orb-v3: Atomistic simulation at scale[J]. arXiv preprint, arXiv: 2504.06231, 2025

[82] Duval A A, Schmidt V, Hernandez-Garcia A, et al. FAENet: Frame averaging equivariant GNN for materials modeling[C]//Proc of the 40th Int Conf on Machine Learning. Honolulu: PMLR, 2023, 202: 9013−9033

[83] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [C] //Proc of the 9th Int Conf on Learning Representations. Virtual: OpenReview.net, 2021: 1−21

[84] Hu Weihua, Shuaibi M, Das A, et al. ForceNet: A graph neural network for large-scale quantum calculations[J]. arXiv preprint, arXiv: 2103.01436, 2021

[85] Liao Yi-Lun, Wood B, Das A, et al. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations[C]//Proc of the 12th Int Conf on Learning Representations. Vienna: OpenReview.net, 2024

[86] Wang Tong, He Xinheng, Li Mingyu, et al. Ab initio characterization of protein molecular dynamics with AI2BMD[J]. Nature, 2024, 635: 1019−1027

[87] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533−536

[88] Kingma D P, Ba J. Adam: A method for stochastic optimization[C]//Proc of the 3rd Int Conf on Learning Representations. San Diego: OpenReview.net, 2015: 1−15

[89] Hu Siyu, Zhao Tong, Sha Qiuchen, et al. Training one DeePMD model in minutes: A step towards online learning[C]//Proc of the 29th ACM SIGPLAN Annual Symp on Principles and Practice of Parallel Programming. New York: ACM, 2024: 257−269

[90] Heimes F. Extended Kalman filter neural network training: Experimental results and algorithm improvements[C]//Proc of 1998 IEEE Int Conf on Systems, Man, and Cybernetics. Piscataway, NJ:

IEEE, 1998, 2: 1639−1644

[91] Singhal S, Wu L. Training multilayer perceptrons with the extended Kalman algorithm[C]//Advances in Neural Information Processing Systems 1. San Mateo: Morgan Kaufmann, 1989: 133−140

[92] Witkoskie J B, Doren D J. Neural network models of potential energy surfaces: Prototypical examples[J]. Journal of Chemical Theory and Computation, 2005, 1(1): 14−23

[93] Hu Siyu, Zhou Yuanchang, Zhao Tong, et al. Neural network force field training based on reorganized layer-wised extended Kalman filtering[J]. Journal of Software, 2025, 36(9): 4093−4109 (in Chinese)
(胡思宇, 周远昌, 赵曈, 等. 基于层重组扩展卡尔曼滤波的神经网络力场训练[J]. 软件学报, 2025, 36(9): 4093-4109)

[94] Chialvo A A, Debenedetti P G. On the use of the Verlet neighbor list in molecular dynamics[J]. Computer Physics Communications, 1990, 60(2): 215−224

[95] Zhu Feiwen, Futrega M, Bao Han, et al. FastDimeNet++: Training DimeNet++ in 22 minutes[C]//Proc of the 52nd Int Conf on Parallel Processing. New York: ACM, 2023: 274−284

[96] Zhou Yuanchang, Hu Siyu, Wang Chen, et al. FastCHGNet: Training one universal interatomic potential to 1.5 hours with 32 GPUs[C]//Proc of 2025 IEEE Int Parallel and Distributed Processing Symp. Piscataway, NJ: IEEE, 2025: 236−246

[97] Geiger M, Smidt T. e3nn: Euclidean neural networks[J]. arXiv preprint, arXiv: 2207.09453, 2022

[98] Firoz J S, Pellegrini F, Geiger M, et al. Optimizing data distribution and kernel performance for efficient training of chemistry foundation models: A case study with MACE[C]//Proc of the 34th Int Symp on High-Performance Parallel and Distributed Computing. New York: ACM, 2025: 1−13

[99] Ghunaim Y, Villa A, Ignacz G, et al. Towards faster and more compact foundation models for molecular property prediction[C]// Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2025: 48−57

[100] Kong Liang, Shim J, Hu Guoliang, et al. Scalable foundation interatomic potentials via message-passing pruning and graph partitioning[J]. npj Computational Materials, 2026, 12: 180

[101] Gou Jianping, Yu Baosheng, Maybank S J, et al. Knowledge distillation: A survey[J]. International Journal of Computer Vision, 2021, 129(6): 1789−1819

[102] Kelvinius F E, Georgiev D, Toshev A, et al. Accelerating molecular graph neural networks via knowledge distillation[J]. Advances in Neural Information Processing Systems, 2023, 36: 25761−25792

[103] Amin I, Raja S, Krishnapriyan A. Towards fast, specialized machine learning force fields: Distilling foundation models via energy Hessians[J]. arXiv preprint, arXiv: 2501.09009, 2025

[104] Deng Lei, Li Guoqi, Han Song, et al. Model compression and hardware acceleration for neural networks: A comprehensive survey[J]. Proceedings of the IEEE, 2020, 108(4): 485−532

[105] Gholami A, Kim S, Dong Z, et al. A survey of quantization methods for efficient neural network inference[M]//Low-Power Computer Vision. Boca Raton, FL: CRC, 2022: 291−326

[106] Jacob B, Kligys S, Chen B, et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 2704−2713

[107] Rai B K, Sresht V, Yang Q, et al. TorsionNet: A deep neural network to rapidly predict small-molecule torsional energy profiles with the accuracy of quantum mechanics[J]. Journal of Chemical Information and Modeling, 2022, 62(4): 785−800

[108] Kaplan A D, Liu Runze, Qi Ji, et al. A foundational potential energy surface dataset for materials[J]. arXiv preprint, arXiv:2503.04070, 2025

[109] Peng Anyang, Cai Chun, Guo Mingyu, et al. LAMBench: A benchmark for large atomic models[J]. arXiv preprint, arXiv: 2504.19578, 2025

[110] Wang H C, Botti S, Marques M A L. Predicting stable crystalline compounds using chemical similarity[J]. npj Computational Materials, 2021, 7(1): 12

[111] FAIR Chemistry. FAIR Chemistry Leaderboard[EB/OL]. [2026-05-25]. https://huggingface.co/spaces/facebook/fairchem_leaderboard

[112] AI for Science Institute, University of Science and Technology of China. OMC-Bench[EB/OL]. [2026-05-25]. https://aisci.ustc.edu.cn/mlip-omcs/#/

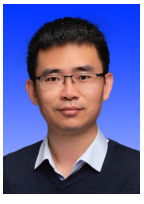
[113] Zhou Yuanchang, Wang Hongyu, Du Yiming, et al. Breaking the training barrier of billion-parameter universal machine learning interatomic potentials[J]. arXiv preprint, arXiv: 2604.15821, 2026



Hu Siyu, born in 1996. PhD, assistant professor. Her main research interests include AI for Science, high-performance computing, and machine learning interatomic potentials.
胡思宇, 1996 年生。博士, 助理研究员。主要研究方向为 AI for Science、高性能计算、机器学习原子间势。



Tan Guangming, born in 1980. PhD, professor, PhD supervisor. His main research interests include parallel algorithm design and analysis, parallel programming design and optimization, computer architecture, and high-performance computing.
谭光明, 1980 年生。博士, 研究员, 博士生导师。主要研究方向为并行算法设计与分析、并行程序设计与优化、计算机体系结构、高性能计算。



Jia Weile, born in 1985. PhD, professor, PhD supervisor. His main research interests include high-performance computing, AI for Science, artificial intelligence, and massively parallel computing.
贾伟乐, 1985 年生。博士, 研究员, 博士生导师。主要研究方向为高性能计算、AI for Science、人工智能、大规模并行计算。