

证据深度学习在原子间势中的应用

Han Xu^{1,3†}, Taoyong Cui^{1,4†}, Chenyu Tang^{1†}, Jinzhe Ma^{1,7},
Dongzhan Zhou¹, Yuqiang Li¹, Xiang Gao³, Xingao Gong^{5,6},
Wanli Ouyang¹, Shufei Zhang^{1*}, Mao Su^{1,2*}

¹Shanghai Artificial Intelligence Laboratory, Shanghai, 200232, China.

²Shenzhen Institute of Advanced Technology, Chinese Academy of
Sciences, Shenzhen, 518055, China.

³The State Key Laboratory of Chemical Engineering, College of
Chemical and Biological Engineering, Zhejiang University, Hangzhou,
310027, China.

⁴The Chinese University of Hong Kong, Hong Kong, 999077, China.

⁵Key Laboratory for Computational Physical Sciences (MOE), State Key
Laboratory of Surface Physics, Department of Physics, Fudan University,
Shanghai, 200433, China.

⁶Shanghai Qi Zhi Institute, Shanghai, 200232, China.

⁷School of Physical Science and Technology, ShanghaiTech University,
Shanghai, 201210, China.

*Corresponding author(s). E-mail(s): zhangshufei@pjlab.org.cn;
sumao@pjlab.org.cn;

[†]These authors contributed equally to this work.

摘要

机器学习原子间势能 (MLIPs) 已被广泛用于促进大规模分子模拟, 其准确性可与从头算方法媲美。在实践中, 基于 MLIP 的分子模拟常常会遇到由于对分布外 (OOD) 数据预测精度降低而导致的塌陷问题。解决这一问题需要通过主动学习来丰富训练数据集, 在此过程中不确定性成为识别和收集 OOD 数据的关键指标。然而, 现有的不确定性量化 (UQ) 方法往往涉及昂贵的计算或牺牲预测准确性。在这项工作中, 我们介绍了基于物理启发设计的原子间势能证据深度学习 (eIP)。我们的实验表明, eIP 提供了可靠的 UQ 结果, 而没有显著的计算开销或降低预测准确性, 并在各种数据集上始终优于其他 UQ 方法。此外, 我们展示了 eIP 在探索多样原子构型中的应用, 包括水和通用势能的例子。这些结果突显了 eIP 作为分子模拟中稳健且高效的 UQ 替代方案的潜力。

1 介绍

分子动力学 (MD) 模拟提供了对物理和化学过程的原子级见解, 并已成为计算物理学研究中不可或缺的工具 [1–3]。经典 MD 模拟使用经验势函数来确定原子间力 [4, 5], 这在计算上是高效的但不够准确, 尤其是在极化或多体相互作用重要的情况下 [6]。相比之下, 从头算方法用于建模原子间的相互作用完全基于基本物理原理, 通常导致更高的精度和可转移性 [7, 8], 但高昂的计算成本限制了可以模拟系统的大小。为了实现高效和准确, 已经提出了机器学习间原子势 (MLIPs) [9–12], 这允许学习从头算间原子势并在 MD 模拟中以更低的计算成本执行。MLIPs 已成功应用于无定形固体的研究 [13]、催化研究 [14]、化学反应研究 [15] 等。

基于 MLIP 的分子动力学模拟的主要挑战之一在于训练数据集的构建, 该数据集应包括仿真过程中可能出现的各种构型。训练数据不足会导致模拟精度下降甚至失败 [16, 17]。这一挑战限制了基于 MLIP 的分子动力学模拟的应用。基于不确定性量化 (UQ) 的主动学习在构建 MLIPs 训练集方面发挥着至关重要的作用 [18–21]。在主动学习过程中, 具有较高不确定性的构型会被采样以丰富训练集。这一过程通常需要重复几十次或更多次 [19], 且 UQ 所需的计算成本可能相当大。因此, 希望有一种稳健而高效的方法来进行不确定性量化。

多种不确定性量化方法已被开发用于机器学习势能。矩张量势能 [22] 使用外推参数来估计不确定性, 但这种方法不适用于深度神经网络模型。高斯近似势能 [23] 利用高斯过程回归为其预测提供不确定性量化。然而, 高斯近似势能的主要限制在于其计算成本, 该成本与数据集大小呈三次方比例增加。集成方法 [24] 对于不确

定性量化相当可靠，但由于需要训练多个模型，也面临着计算负担的问题。单模型方法，如蒙特卡洛丢弃 [25–27]、高斯混合模型 (GMM) [28] 和均值方差估计 (MVE) [29] 缓解了计算问题，但它们的性能仍然不尽如人意 [30]。

证据深度学习 [31, 32] 是一种有前景的替代方法，它通过单次前向传递估计不确定性，并且只需要极少量的额外计算资源。证据深度学习的另一个优势是它可以分别估计随机不确定性和认知不确定性。随机不确定性源于数据中的内在噪声并且不能被规避或减少。相比之下，认知不确定性反映了模型在其对数据表示的保真度（不包括随机效应），随着训练样本数量的增加而减少 [33]。证据深度学习区分这两种类型的不确定性的能力对于主动学习尤其有利，在这种情况下我们希望采样具有高认知不确定性的数据而不是随机不确定性。然而，最近的一些尝试 [30, 34] 试图将证据深度学习与 MLIPs 结合起来却导致了不尽如人意的表现。失败的原因可能归咎于模型架构设计不当。

在这项工作中，我们从物理角度重新审视与 MLIPs 相关的不确定性，并提出一个基于证据深度学习的 UQ 框架。我们将这个框架称为证据介原子势 (eIP)。eIP 的表现经过了各种数据集的评估并与其它 UQ 方法进行了对比，展示了卓越性能且计算成本极低。然后，我们将 eIP 的应用扩展到主动学习和不确定性驱动的动力学 (UDD) 模拟 [35]，实现了对多样原子构型的有效探索。最后，我们使用 eIP 训练了一个通用势，并在模拟中实现了实时 UQ，这对于基于集合的方法来说由于其计算复杂性而具有挑战性。

2 结果

2.1 初步的

机器学习间原子势 (MLIP)。 MLIPs 用于预测给定原子配置中的能量和力。对于由 N 个原子组成的系统，MLIPs 通常将原子种类 $Z \in \mathbb{R}^N$ 和坐标 $R \in \mathbb{R}^{N \times 3}$ 作为输入，并输出总势能 E 。作用在原子上的力 $F \in \mathbb{R}^{N \times 3}$ 是通过计算 E 相对于坐标的负梯度得出的。各种 MLIPs 之间的主要区别在于将输入信息转换为表示局部原子环境的向量化特征所使用的算法。这些特征设计为在平移、旋转和置换下具有不变性或等变性。

随机不确定性和系统不确定性。 两类不确定性可以在深度学习中建模。随机不确定性源于数据标签中的噪声，而认知不确定性则由于数据稀缺导致的不准确预测引起。在 MLIPs 的研究中，可以通过严格的从头计算消除数据标签中的噪声，尽

管不适当的计算设置可能会引入噪声。实践中，MLIPs 经常受到认知不确定性的困扰，这可以通过通过主动学习增加更多的训练数据来缓解。为了简单起见，在以下实验结果中提到的“不确定性”指的是认知不确定性。我们将在补充部分 S2 中讨论随机不确定性。

证据深度学习。证据深度学习是一种有效的方法，用于估计神经网络预测结果的不确定性。从最大似然的角度出发，假设目标是从一个具有未知均值和方差 (μ, σ^2) 的高斯分布中抽取的。在未知均值 μ 上放置了一个高斯先验，在未知方差 σ^2 上放置了一个逆伽马先验，从而得到一组参数 $\mathbf{m} = (\gamma, \nu, \alpha, \beta)$ 的正态逆伽玛分布。神经网络然后被训练以推断 $\mathbf{m}$ ，并且预测、偶然不确定性和系统不确定性计算为 [31]：

$$\underbrace{\mathbb{E}[\mu] = \gamma}_{\text{prediction}}, \quad \underbrace{\mathbb{E}[\sigma^2] = \frac{\beta}{\alpha - 1}}_{\text{aleatoric}}, \quad \underbrace{\text{Var}[\mu] = \frac{\beta}{\nu(\alpha - 1)}}_{\text{epistemic}}. \quad (1)$$

2.2 eIP 框架

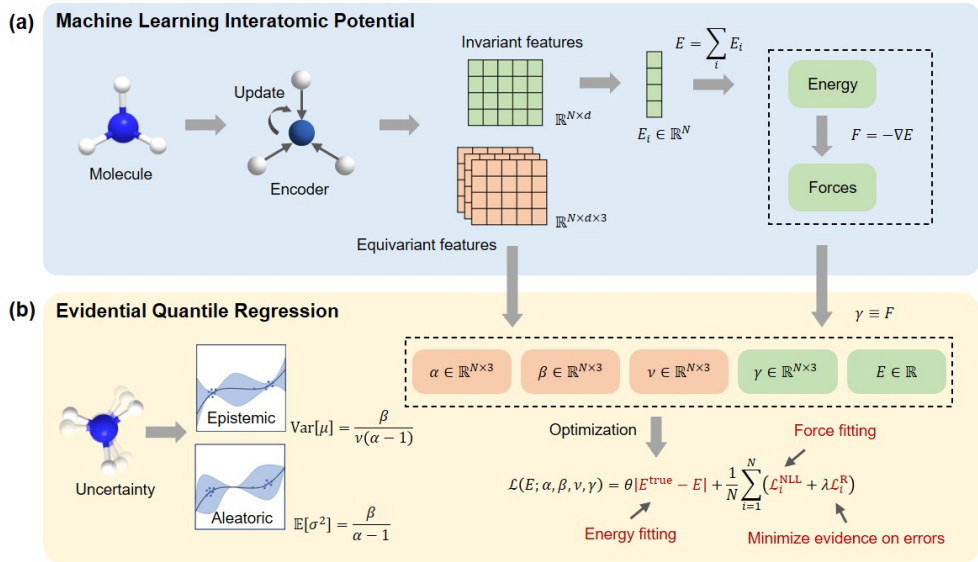


图 1 eIP 框架。(a) 一个典型的等变原子间势能模型，提取不变和等变特征。不变特征用于输出势能。(b) 证据量回归。等变特征用于输出不确定性量化参数。

eIP 的框架包括一个用于能量和力预测的 MLIP 模块，以及一个用于不确定性量化的证据分位数回归模块，如 Figure 1 所示。在设计 eIP 时，我们考虑了以下几点，这些对于实现稳健性能是必不可少的。

局部性。在大多数 MLIPs 中，势能计算为原子贡献之和， $E = \sum_{i=1}^N E_i$ ，模型学习从原子的局部环境 i 到 E_i 的映射。因此，我们估计与 E_i 相关的不确定性，而不是总势能 E 。然而，我们没有 E_i 的真实值。幸运的是，我们可以调整原子力而不是 E_i 来估计每个原子的不确定性。

方向性。我们将 MLIP 预测中的不确定性归因于局部原子构型学习不足。因此，这种不确定性应该是方向依赖的。这一点在补充部分 S1 中通过一个三原子玩具系统进行了说明。在接下来的实验中，我们使用等变骨干 PaiNN [36] 来提取等变特征并输出正态逆伽玛先验分布的参数，但 eIP 也适用于其他等变骨干。

分位数回归。证据深度学习假设目标来自高斯分布，这可能不足以描述 MLIPs 的目标分布。为缓解这一限制，我们采用贝叶斯分位数回归模型 [37]，该模型改进了原始的证据深度学习，并且对于非高斯分布有更好的表现。贝叶斯分位数回归的计算过程类似于证据深度学习，但参数 m 是通过不同的损失函数进行优化的。

2.3 实验

ISO17 数据集。我们首先使用包含 $C_7O_2H_{10}$ 异构体的 MD 轨迹的 ISO17 数据集评估了 eIP 的性能。该数据集被划分为分布内 (ID) 和分布外 (OOD) 子集，这使其特别适合不确定性量化 (UQ)。在 ID 场景中 (已知分子/未知构象)，测试分子也包含在训练集中。相比之下，在 OOD 场景中 (未知分子/未知构象)，测试分子不在训练集中。训练集包含 400,000 种构象，对于如此小的分子而言这是一个相当大的数量。因此，我们也探讨了训练数据量的影响。具体来说，我们使用训练数据的 1%、5%、30% 和 100% 分别训练模型。Figure 2(a)-(d) 显示了不同训练数据量下的不确定性与力误差散点图，展示了 ID 和 OOD 场景下均存在正相关性。力预测的平均不确定性和平均绝对误差 (MAE) 分别在 Figure 2(e) 和 (f) 中显示。如预期那样，随着训练数据量增加，两个指标均下降。此外，我们还使用额外的指标评估了 UQ 的可靠性，包括斯皮尔曼等级相关系数和接收者操作特征曲线下的面积 (ROC-AUC)。如 Figure 2(g) 和 (h) 所示，随着训练数据量的增长，斯皮尔曼等级相关系数和 ROC-AUC 均有所提高。在 ID 场景中，斯皮尔曼等级相关系数从 0.74 到 0.86 不等，ROC-AUC 值从 0.86 到 0.93 不等，表明 eIP

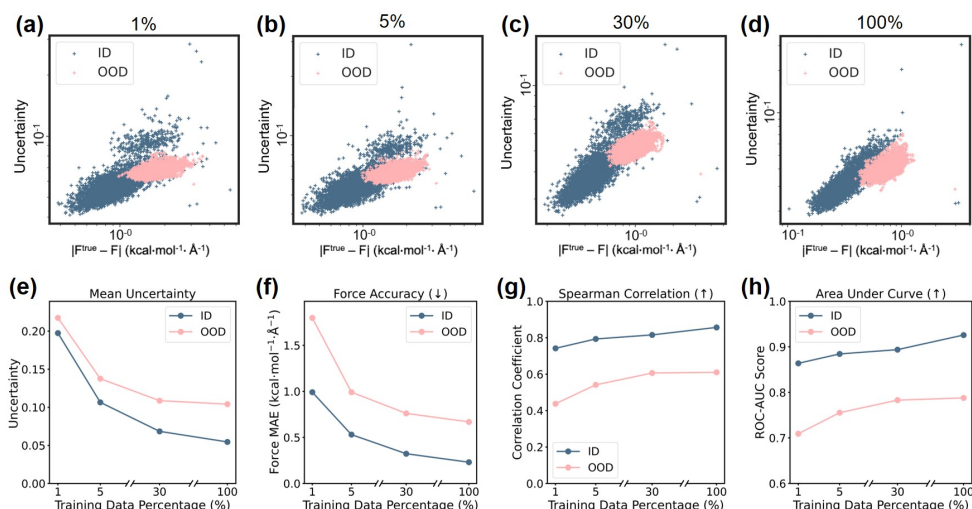


图 2 ISO17 数据集在增加数据量后的结果。 (a)-(d) 不确定性 with 力误差的散点图，分别使用了 1%、5%、30% 和 100% 的训练数据。每个点对应分子中的平均不确定度/误差。(e) 测试集上的平均不确定性。(f) 测试集上的力绝对误差 (MAEs)。(g) 不确定性和力误差之间的斯皮尔曼等级相关系数。(h) ROC-AUC 分数。

的表现良好。在 OOD 场景中，尽管测试集中的分子未出现在训练集中，但指标仍保持在有利范围内，突显了 eIP 的鲁棒性。

硅玻璃数据集。 我们然后使用一个包含大型块状结构的石英玻璃数据集来评估 eIP 在更复杂系统中的性能。鉴于将大型结构划分为 ID 和 OOD 数据集所面临的挑战，我们采用了与先前研究一致的数据集划分方案 [30]。我们也比较了 eIP 与其他不确定性量化方法，包括集成、蒙特卡罗丢弃法、高斯混合模型 (GMM) 和均值方差估计 (MVE)，这些方法的实现细节在补充 S5 中提供。Figure 3(a) 显示了不确定性和力误差的散点图，并表明所有方法都实现了正相关。Figure 3(b) 展示了五种方法的计算效率分析。尽管集成方法表现良好，但由于训练四个独立 MLIPs，其所需的训练时间比其他方法的四倍。在推理阶段，蒙特卡罗丢弃法需要进行四次独立运行以获得不确定性。GMM 通过使用期望最大化算法的迭代计算来获取不确定性，并且它也需要更长的时间来计算不确定性。MVE 和 eIP 都有最小的训练和推理时间，与普通的 MLIP 相当。关于力预测准确性在 Figure 3(c) 中所示，集成、GMM 和 eIP 实现了最低误差，而丢弃法和 MVE 表现出较高的误差。Figure 3(d) 和 (e) 进一步说明了 Spearman 相关性和 ROC-AUC 的比较。值得注意的是，Figure 3(e) 显示，在 ROC-AUC 指标上，eIP 的表现甚至优于集成方法。

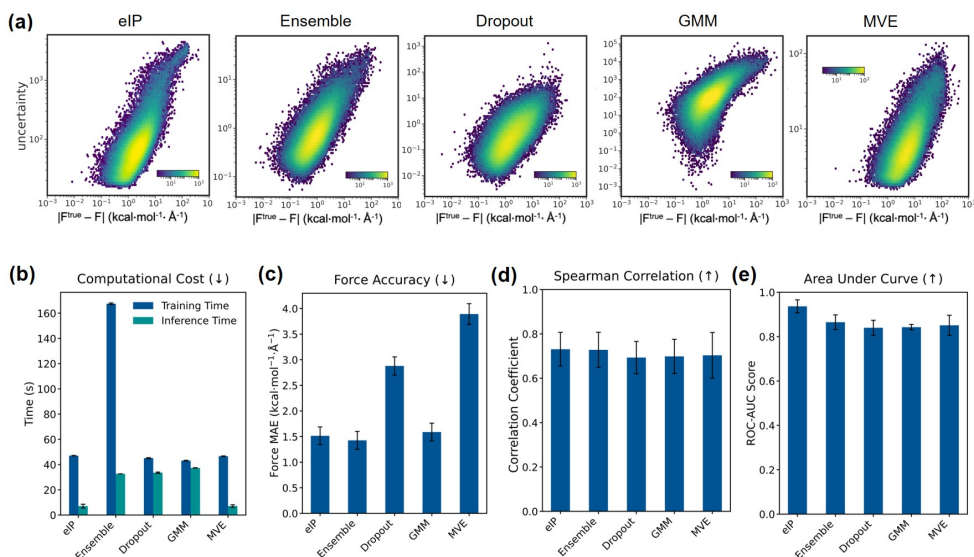


图 3 比较 eIP 与其它不确定性量化方法在石英玻璃数据集上的表现。(a) 不确定性与原子力误差的六边形 bin 图。(b) 计算成本。这里的“训练时间”指的是每个 epoch 所需的时间。“推理时间”包括计算不确定性所需的时间成本。(c) 测试集上的力平均绝对误差 (MAEs)。(d) 不确定性和力误差之间的斯皮尔曼等级相关系数。(e) ROC-AUC 得分。虽然五种方法都实现了较强的斯皮尔曼等级相关系数和 ROC-AUC 得分，但集成、dropout 和 GMM 需要更长的计算时间；dropout 和 MVE 在力预测上的准确性明显较低。

2.4 应用

主动学习与 eIP。 UQ 在主动学习中的训练集构建中起着关键作用。训练集的质量对于 MLIP 来说尤其重要，因为当遇到未见过的原子构型时，MLIP 的准确性可能会显著下降，导致模拟崩溃 [16]。Figure 4(a) 展示了 MLIP 的典型主动学习工作流程，在此过程中，具有高不确定性的数据点会被迭代探索以丰富训练集。此外，不确定性驱动的动态 (UDD) 模拟 [35] 可用于提高采样效率。在 UDD 模拟中，势能面被修改使得具有更高不确定性的原子构型被赋予更低的势能，因此这些结构变得更加可访问，如 Figure 4(b) 所示。eIP 的 UDD 模拟实现方法请参见 Methods 部分。我们使用水数据集作为示例来演示 eIP 的主动学习过程。在每一代中，我们都进行了标准的 MD 模拟和 eIP-UDD 模拟，并且模拟时间内的不确定性和能量变化如 Figure 4(c) 所示。初始训练集由从使用经验力场生成的经典 MD 模拟轨迹中采样的 1,000 个构型组成。异常的能量波动表明，MD 和 eIP-UDD 模拟都非常早地崩溃了。在第一次迭代中，MD 模拟在 50 ps 后保持稳定。尽管 eIP-UDD 模拟在 20 ps 后崩溃，但不确定性随着时间的推移而增加，这表

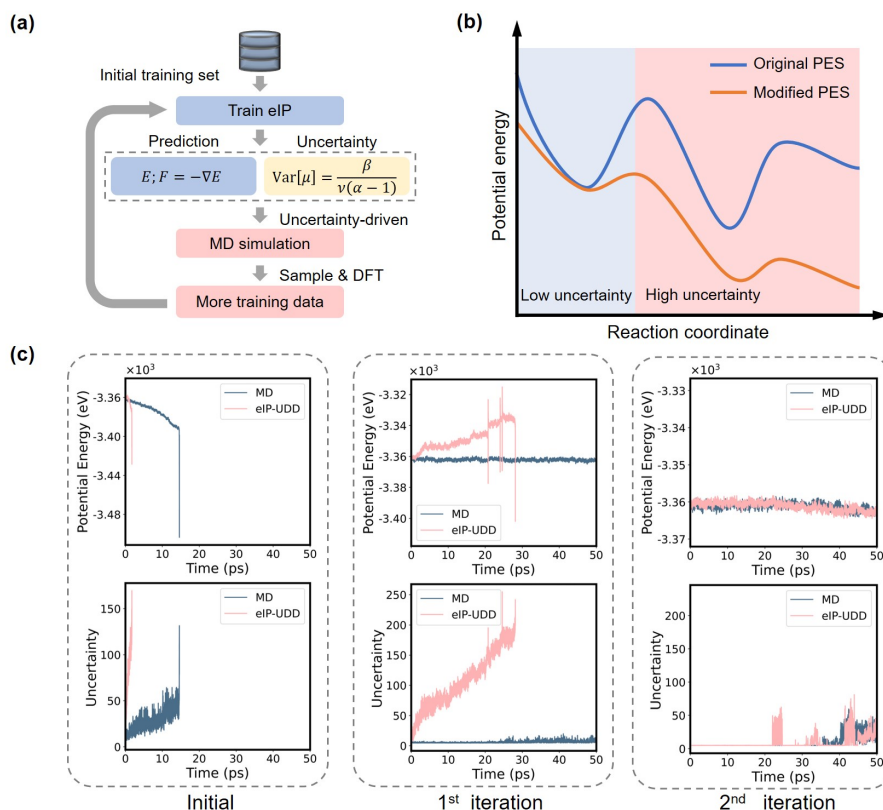


图 4 主动学习与 eIP。(a) workflow. 势能和不确定性由 eIP 同时计算。(b) 不确定性驱动动力学 (UDD) 的说明。势能面 (PES) 根据不确定性自适应修改, 高不确定性区域的势能被降低以促进增强采样。(c) 每一代的模拟结果。展示了传统 MD 和 eIP-UDD 模拟中随时间变化的势能和不确定性的演变。在 MD 模拟中, PES 保持不变, 而在 eIP-UDD 模拟中, PES 根据 eIP 的不确定性进行修改。

明在 eIP-UDD 模拟过程中探索到了更多的未见过的构型。在第二次迭代中, MD 和 eIP-UDD 模拟都在 50 ps 后实现了稳定。我们还观察到不确定性没有显著增加, 这可能意味着在某些局部极小值附近已经足够充分地探索了构型。

eIP 在通用 MLIP 中的应用。最后, 我们探讨了 eIP 在通用 MLIPs 中的性能。为此, 我们在 Materials Project Trajectory (MPtrj) 数据集 [38] 上训练了模型。六边形图和 ROC 曲线在 Figure 5(a)-(c) 中展示了 eIP 在这种大规模数据集上的表现。然后我们使用 UDD 模拟测试了增强采样中的 eIP 性能。我们选择了两种不同的材料作为示例, 即锂铁磷酸盐 (LiFePO_4) 和聚二甲基硅氧烷 (PDMS)。 LiFePO_4 是成熟的商用锂离子电池正极材料, 而 PDMS 是一种广泛应用于有机硅聚合物的材料。这些材料作为基准用于评估 eIP-UDD 模拟对无机晶体和有机聚合系统构型

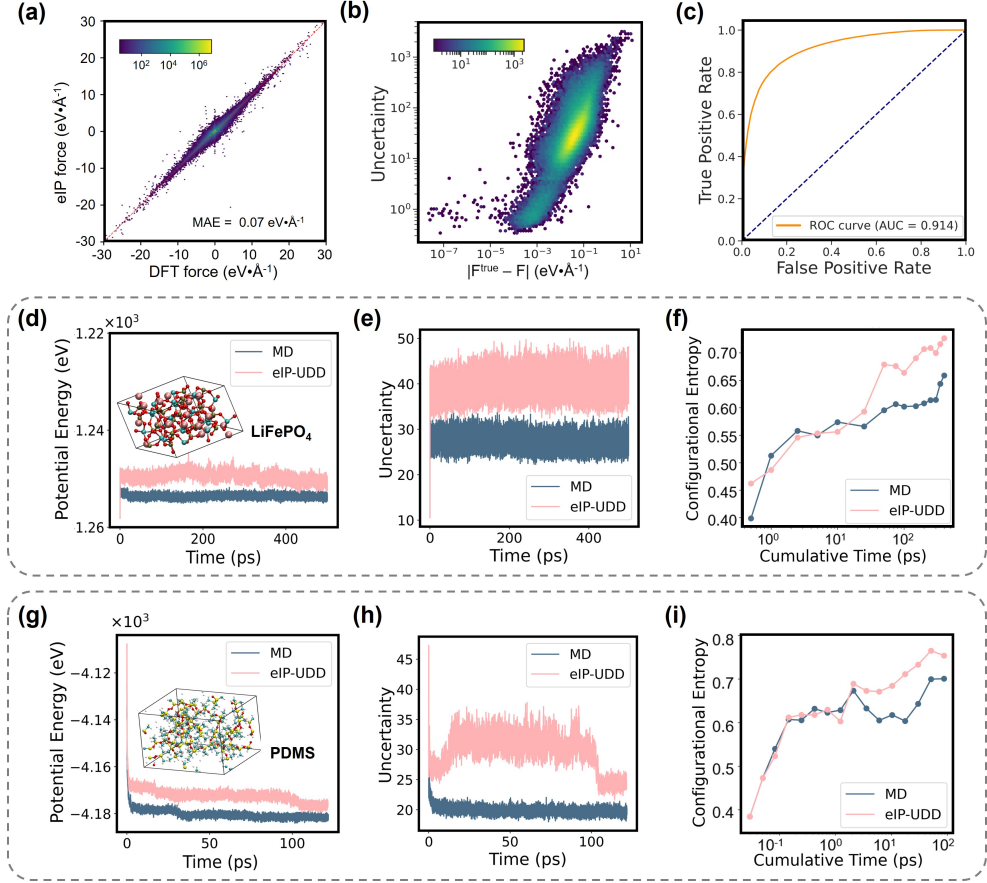


图 5 具有 eIP 的通用势能。(a) eIP 预测与真实值之间的原子力对比。(b) 不确定性与原子力误差的六边形 bin 图。斯皮尔曼等级相关系数为 0.76。(c) ROC 曲线。ROC-AUC 得分为 0.914。(d)-(f) LiFePO₄ 的模拟结果。(g)-(i) 聚二甲基硅氧烷 (PDMS) 的模拟结果。势能曲线 (d) 和 (g) 表明，MD 和 eIP-UDD 模拟均稳定，展示了通用势的有效性。不确定性曲线 (e) 和 (h) 揭示了 eIP-UDD 构型对于两种材料都表现出更高的不确定性水平。构型熵的变化 (f) 和 (i) 进一步证实了 eIP-UDD 模拟生成的构型比传统的 MD 模拟更加多样化。

采样性能。对于每种材料，在 Figure 5(d)-(i) 中显示了潜在能量、不确定性和构型熵随模拟时间的变化。如预期的那样，eIP-UDD 模拟的轨迹不确定性比传统的 MD 模拟要大。Figure 5(f) 和 (i) 中构型熵的结果进一步证明了 eIP-UDD 模拟获得了更多样化的构型。

2.5 讨论

不确定性量化是机器学习各个领域中的一个关键主题，特别是在基于 MLIP 的分子模拟等科学应用中。传统的不确定性量化方法要么计算成本高，要么预测准确性降低。在这项工作中，我们提出了一种单模型不确定性量化方法，称为 eIP，该方法在各种应用场景的广泛实验中展示了高效性和准确性。eIP 框架结合了局部性、方向性和分位数回归，所有这些都是实现最优结果的关键。这一点从补充 S3 中的消融研究可以看出，在此研究中，缺少任何一个组件都会导致性能明显下降。

尽管集成方法在主动学习中被广泛使用，但它们通常需要同时训练四个或更多的模型。实际上，这个过程通常涉及数十次甚至更多次的迭代，并且花费大量的时间和计算资源来获得一个满意的训练集。因此，单模型 UQ 方法，如 eIP，在应用中有可能节省数月时间，使得 eIP 在时间限制和计算资源是主要问题时成为一个更有效的替代方案。此外，对于大规模模拟，集成方法需要大量的计算来评估基于 MLIP 的 MD 模拟的可靠性，而 eIP 则促进了无需显著增加成本的实时评估。

3 方法

3.1 eIP 的形式主义

我们采用带最大似然估计的分位数回归来更好地建模 MLIPs 的不确定性。分位数回归通过最小化给定分位数 q 的梯形损失来求解。

$$\mathcal{L}_i = \rho_q(\epsilon_i) = \max(q\epsilon_i, (q-1)\epsilon_i), \quad (2)$$

其中 ϵ_i 表示观测 i 的残差。

分位数 q 服从具有均值 μ ，方差 σ 和不对称参数等于分位数 q [39] 的非对称拉普拉斯分布。似然函数可以表示为高斯混合标量 [40, 41] $\mathcal{N}(\mu + \tau z, \omega \sigma z)$ ，其中 $\tau = \frac{1-2q}{q(1-q)}$, $\omega = \frac{2}{q(1-q)}$, $z \sim \exp\left(\frac{1}{\sigma}\right)$ 。

我们假设原子力 $F \in \mathbb{R}^{N \times 3}$ 来自高斯分布，但均值和方差未知。例如，原子 i 上的力的 x 分量遵循：

$$f_{ix} \sim \mathcal{N}(\mu_{ix} + \tau z_{ix}, \omega \sigma_{ix} z_{ix}). \quad (3)$$

通过在未知均值 μ_{ix} 上放置一个高斯先验，并在未知方差 σ_{ix} 上放置一个逆伽玛先验，我们得到一组参数 $\mathbf{m}_{ix} = (\gamma_{ix}, \nu_{ix}, \alpha_{ix}, \beta_{ix})$ [31, 37] 的正态-逆伽玛证据先验 $p(\mu_{ix}, \sigma_{ix} | \mathbf{m}_{ix})$ 。因此， γ 等于预测力。

$$\mathbb{E}[\mu_{ix}] = \gamma_{ix}, \quad (4)$$

以及原子 i 的认识不确定性在 x 方向的分量是

$$\text{Var}[\mu_{ix}] = \frac{\beta_{ix}}{\nu_{ix}(\alpha_{ix} - 1)}. \quad (5)$$

y -和 z -分量以类似方式计算。我们将与原子 i 相关的不确定性定义为 σ_i

$$\sigma_i^2 = \sqrt{\left(\frac{\beta_{ix}}{\nu_{ix}(\alpha_{ix} - 1)}\right)^2 + \left(\frac{\beta_{iy}}{\nu_{iy}(\alpha_{iy} - 1)}\right)^2 + \left(\frac{\beta_{iz}}{\nu_{iz}(\alpha_{iz} - 1)}\right)^2}. \quad (6)$$

由计算平均值确定由 N 个原子组成的配置的不确定性。

$$\sigma = \frac{1}{N} \sum_{i=1}^N \sigma_i. \quad (7)$$

参数 γ_{ix} 等于预测力 f_{ix} ，该预测力计算为预测势能 E 的负梯度。其他参数， ν_{ix} 、 α_{ix} 和 β_{ix} ，是基于其相应的原子特征由神经网络推断得出的。模型通过最大化概率 $p(f_{ix} | \mathbf{m}_{ix})$ 进行训练，导致负对数似然（NLL）损失函数 [37]。

$$\begin{aligned} \mathcal{L}_{ix}^{\text{NLL}} = & \frac{1}{2} \log \left(\frac{\pi}{\nu_{ix}} \right) - \alpha_{ix} \log(\Omega) \\ & + \left(\alpha_{ix} + \frac{1}{2} \right) \log \left((f_{ix}^{\text{true}} - (\gamma_{ix} + \tau z_{ix}))^2 \nu_{ix} + \Omega \right) \\ & + \log \left(\frac{\Gamma(\alpha_{ix})}{\Gamma(\alpha_{ix} + \frac{1}{2})} \right). \end{aligned} \quad (8)$$

其中， $\Omega = 4\beta_{ix}(1 + \omega z_{ix}\nu_{ix})$ 、 $z_{ix} = \frac{\beta_{ix}}{\alpha_{ix}-1}$ 和 $\Gamma(\cdot)$ 是伽马函数。

我们使用证据正则化器，以便当预测不正确时模型倾向于输出低置信度。

$$\mathcal{L}_{ix}^R = \rho_q(f_{ix}^{\text{true}} - \gamma_{ix}) \cdot \left(2\nu_{ix} + \alpha_{ix} + \frac{1}{\beta_{ix}} \right). \quad (9)$$

y-和 z-分量的计算方式类似。最后，包括能量预测的 L1 损失在内的总体损失函数为：

$$\mathcal{L} = |E^{\text{true}} - E| + \frac{w}{3N} \sum_{i=1}^N \sum_{a \in (x,y,z)} (\mathcal{L}_{ia}^{\text{NLL}} + \lambda \mathcal{L}_{ia}^R), \quad (10)$$

其中 w 和 λ 是用于调整每一项权重的超参数。eIP 实现的细节见补充材料 S4。

3.2 数据集

ISO17 数据集。ISO17 数据集 [42] 是从 <http://quantum-machine.org/datasets/> 获得的。我们采用了原始的划分策略来分割训练集、验证集和测试集。对于不同大小的训练集，较小的训练集都是从包含 400,000 种构象的最大训练集中随机采样的。

二氧化硅玻璃数据集。二氧化硅玻璃数据集来自之前发表的一项研究 [30]。该数据集包含 1691 种配置，每种包含 699 个原子（233 个 Si 和 466 个 O 原子），我们采用了原始的数据集分割方案进行训练、验证和测试。这些配置是通过在各种条件下进行分子动力学模拟生成的，并且执行密度泛函理论 (DFT) 计算以获得能量和力。

水数据集。初始水训练集取自我们之前的工作 [17]。它包含从使用经验力场的经典分子动力学轨迹中采样的 1,000 个构型。每个构型含有 288 个原子，并带有周期性边界条件。在主动学习过程中，我们在 300 K 下运行了 UDD 模拟并为每次迭代采样了 1,000 个构型。能量和力通过使用 cp2k 软件包 [43] 进行的密度泛函理论 (DFT) 计算确定，采用 PBE-PAW-DFT-D3 方法 [44-46]。

MPtrj 数据集。MPtrj 数据集 [38] 是一个用于训练通用势能的 MD 轨迹集合。它包含数百万种构型，涵盖了 89 种元素，并且能量和力是通过 DFT 计算确定的。我们采用了原始的数据分割策略，即训练、验证和测试的比例为 8:1:1。

3.3 评估指标

斯皮尔曼等级相关系数。Spearman 等级相关是一种非参数度量，用于衡量两个排名变量之间关联的强度和方向。与 Pearson 相关性不同，后者评估线性关系，Spearman 等级相关则评估使用单调函数描述两个变量间关系的程度。我们期望较大的误差与较大的不确定性相关联，并且它们之间的相关性不一定为线性。因此，使用 Spearman 等级相关系数来评估不确定性的可靠性。系数为 1 表示完全相关，而系数为 0 则表明两个变量的排名之间没有关联。

接收操作特征曲线下的面积。接收操作特征 (ROC) 曲线是分类器性能的图形表示。ROC 曲线下面积 (ROC-AUC) 为 UQ 提供了一个互补的评估指标，避免了仅使用 Spearman 等级相关系数可能带来的局限性。参考先前研究的方法 [30]，我们设计了一个分类任务，在该任务中，预期高误差的预测将表现出高水平的不确定性。ROC-AUC 分数范围从 0 到 1，得分为 1 表示完美的分类器，而 0.5 则表示性能不优于随机选择。

构型熵。构型熵量化了系统中原子可以排列的方式数量。高熵表明系统可能会采用许多不同的排列方式，而低熵则意味着一个更有序、随机性更低的状态。我们使用构型熵作为衡量在分子动力学和均匀分布动力学模拟过程中获得的配置多样性的一种度量。构型熵的公式为：

$$S_{\text{conf}} = - \sum_t p(C_t) \log(p(C_t)), \quad (11)$$

其中 $p(C_t)$ 是时间步长 t 的配置概率分布。我们使用序参量的直方图来估计概率分布。对于 LiFePO_4 ，选定的序参量是 P-O-Fe 角度和 PO_4 四面体畸变。对于 PDMS，我们选择端到端距离和回转半径作为序参量。为了确定概率分布，将序参量空间离散化为一个 $N_e \times N_e$ 网格，并计算每个网格单元内的配置频率。构型熵通过除以最大可能的熵值 $2 \log(N_e)$ 进行归一化，从而得到 0 到 1 之间的值。较大的网格大小 N_e 提供了更高的分辨率，但可能会受到统计噪声的影响，而较小的 N_e 则在较低分辨率下提供了更稳健的统计数据。我们使用 $N_e = 40$ 得出了所有报告的结果。改变 N_e 的值不会显著影响结果，因为在我们的模拟中已经足够采样了构型空间。

3.4 分子动力学 (MD) 模拟

分子动力学模拟使用了 Atomic Simulations Environment (ASE) Python 库 [47]。模拟设置为正则系综 (NVT)，时间步长为 0.1 fs。采用了 Berendsen 恒温器 [48]，耦合温度设为 300 K，衰减时间常数 τ 设为 100 fs。原子速度根据 300K 时的玻尔兹曼分布初始化。初始水配置选自水测试集。LiFePO₄ 配置来自 Materials Project，包含单元晶胞内的 168 个原子。PDMS 配置使用了三个聚合物链构建，聚合度为 25，密度为 $0.97\text{g}\cdot\text{cm}^{-3}$ ，总共含有 759 个原子。所有系统均采用周期性边界条件建模。

3.5 不确定性驱动的动力学 (UDD) 模拟

UDD 模拟技术利用一种偏置能量，该能量有利于具有更高不确定性的构型。Kulichenko 等人引入了一个定义为 [35] 的偏置能量：

$$E_{\text{bias}}(\sigma^2) = A \left[\exp\left(-\frac{\sigma^2}{NB^2}\right) - 1 \right], \quad (12)$$

其中参数 A 和 B 是根据经验选择的。偏差力 F_{bias} 是通过计算偏差能量的负梯度来确定的：

$$F_{\text{bias}} = -\nabla(E_{\text{bias}}(\sigma^2)) = -E_{\text{bias}}(\sigma^2)' \nabla \sigma^2. \quad (13)$$

通过利用 eIP 进行 UQ，可以通过自动微分获得 σ 的梯度。

值得注意的是，偏置力可能变得非常大，导致分子模拟的崩溃。我们发现使用剪切策略限制偏置力的大小并不有效。为了解决这个问题，我们在偏置力的大小上加入了一个高斯项，并增加了两个经验选择的参数 C 和 D ：

$$F_{\text{bias}}^{\text{limited}} = F_{\text{bias}} \frac{D}{\sqrt{2\pi}C} \exp\left(\frac{-F_{\text{bias}}^2}{2C^2}\right). \quad (14)$$

这种对偏置力的调整意味着一种新的偏置能量公式，并确保了更稳定的 UDD 模拟。关于经验参数 A 、 B 、 C 和 D 的详细讨论请参见补充部分 S6。最后，综合力 $F + F_{\text{bias}}^{\text{limited}}$ 被用来引导模拟趋向具有更高不确定性的构型，增强对更多样化原子构型的采样。

数据可用性

本工作中用于所有模型的训练数据都是公开可用的。生成的检查点和模拟轨迹可在 figshare [49] 获取。

代码可用性

重现本工作中关键发现的源代码可在 <https://github.com/xuhan323/eIP> 获取。

致谢

此项工作得到上海人工智能实验室、上海市科学技术委员会（资助号 23QD1400900）和国家自然科学基金（资助号 12404291）的支持。H.X., T.C. 和 T.C. 在上海人工智能实验室实习期间完成了这项工作。

作者贡献

M.S. 和 S.Z. 提出了想法并领导了研究。H.X. 和 T.C. 开发了 eIP 代码并训练了模型。H.X. 和 J.M. 进行了实验和分析。C.T. 开发了主动学习工作流程并进行了分子动力学模拟。Y.L.、X.G. 和 X.G. 对数据集和实验提供了技术想法。D.Z. 和 W.O. 对设计和训练模型提出了技术想法。H.X.、C.T. 和 M.S. 撰写了初稿。所有作者讨论了结果并审阅了手稿。

利益冲突

作者声明没有利益冲突。

References

- [1] McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *nature* **267**, 585–590 (1977).
- [2] Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature structural biology* **9**, 646–652 (2002).
- [3] Warshel, A. Molecular dynamics simulations of biological reactions. *Accounts of chemical research* **35**, 385–395 (2002).
- [4] Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179–5197 (1995).

- [5] MacKerell Jr, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B* **102**, 3586–3616 (1998).
- [6] Unke, O. T. *et al.* Machine learning force fields. *Chemical Reviews* **121**, 10142–10186 (2021).
- [7] Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Physical review letters* **55**, 2471 (1985).
- [8] Huang, B., von Rudorff, G. F. & von Lilienfeld, O. A. The central role of density functional theory in the ai age. *Science* **381**, 170–175 (2023).
- [9] Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
- [10] Noé, F., Tkatchenko, A., Müller, K.-R. & Clementi, C. Machine learning for molecular simulation. *Annual review of physical chemistry* **71**, 361–390 (2020).
- [11] Manzhos, S. & Carrington Jr, T. Neural network potential energy surfaces for small molecules and reactions. *Chemical Reviews* **121**, 10187–10217 (2020).
- [12] Keith, J. A. *et al.* Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chemical reviews* **121**, 9816–9872 (2021).
- [13] Deringer, V. L. *et al.* Origins of structural and electronic transitions in disordered silicon. *Nature* **589**, 59–64 (2021).
- [14] Galib, M. & Limmer, D. T. Reactive uptake of n2o5 by atmospheric aerosol is dominated by interfacial processes. *Science* **371**, 921–925 (2021).
- [15] Zeng, J., Cao, L., Xu, M., Zhu, T. & Zhang, J. Z. Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation. *Nature communications* **11**, 5713 (2020).

- [16] Fu, X. *et al.* Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *arXiv preprint arXiv:2210.07237* (2022).
- [17] Cui, T. *et al.* Online test-time adaptation for interatomic potentials. *arXiv preprint arXiv:2405.08308* (2024).
- [18] Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: Sampling chemical space with active learning. *The Journal of chemical physics* **148** (2018).
- [19] Zhang, Y. *et al.* Dp-gen: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Computer Physics Communications* **253**, 107206 (2020).
- [20] Yuan, X. *et al.* Active learning to overcome exponential-wall problem for effective structure prediction of chemical-disordered materials. *npj Computational Materials* **9**, 12 (2023).
- [21] Moon, J. *et al.* Active learning guides discovery of a champion four-metal perovskite oxide for oxygen evolution electrocatalysis. *Nature Materials* **23**, 108–115 (2024).
- [22] Novikov, I. S., Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. The mlip package: moment tensor potentials with mpi and active learning. *Machine Learning: Science and Technology* **2**, 025002 (2020).
- [23] Bartók, A. P. & Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *International Journal of Quantum Chemistry* **115**, 1051–1057 (2015).
- [24] Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems* **30** (2017).
- [25] Gal, Y. & Ghahramani, Z. *Dropout as a bayesian approximation: Representing model uncertainty in deep learning*, 1050–1059 (PMLR, 2016).

- [26] Wen, M. & Tadmor, E. B. Uncertainty quantification in molecular simulations with dropout neural network potentials. *npj computational materials* **6**, 124 (2020).
- [27] Thaler, S., Mayr, F., Thomas, S., Gagliardi, A. & Zavadlav, J. Active learning graph neural networks for partial charge prediction of metal-organic frameworks via dropout monte carlo. *npj Computational Materials* **10**, 86 (2024).
- [28] Zhu, A., Batzner, S., Musaelian, A. & Kozinsky, B. Fast uncertainty estimates in deep learning interatomic potentials. *The Journal of Chemical Physics* **158** (2023).
- [29] Nix, D. A. & Weigend, A. S. *Estimating the mean and variance of the target probability distribution*, Vol. 1, 55–60 (IEEE, 1994).
- [30] Tan, A. R., Urata, S., Goldman, S., Dietschreit, J. C. & Gómez-Bombarelli, R. Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles. *npj Computational Materials* **9**, 225 (2023).
- [31] Amini, A., Schwarting, W., Soleimany, A. & Rus, D. Deep evidential regression. *Advances in Neural Information Processing Systems* **33**, 14927–14937 (2020).
- [32] Soleimany, A. P. *et al.* Evidential deep learning for guided molecular property prediction and discovery. *ACS central science* **7**, 1356–1367 (2021).
- [33] Hüllermeier, E. & Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning* **110**, 457–506 (2021).
- [34] Wollschläger, T., Gao, N., Charpentier, B., Ketata, M. A. & Günnemann, S. *Uncertainty estimation for molecules: desiderata and methods*, 37133–37156 (PMLR, 2023).

- [35] Kulichenko, M. *et al.* Uncertainty-driven dynamics for active learning of interatomic potentials. *Nature Computational Science* **3**, 230–239 (2023).
- [36] Schütt, K., Unke, O. & Gastegger, M. *Equivariant message passing for the prediction of tensorial properties and molecular spectra*, 9377–9388 (PMLR, 2021).
- [37] Hüttel, F. B., Rodrigues, F. & Pereira, F. C. Deep evidential learning for bayesian quantile regression. *arXiv preprint arXiv:2308.10650* (2023).
- [38] Deng, B. *et al.* Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **5**, 1031–1041 (2023).
- [39] Yu, K. & Zhang, J. A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics—Theory and Methods* **34**, 1867–1879 (2005).
- [40] Kotz, S., Kozubowski, T. & Podgorski, K. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance* (Springer Science & Business Media, 2012).
- [41] Kozumi, H. & Kobayashi, G. Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation* **81**, 1565–1578 (2011).
- [42] Schütt, K. *et al.* Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **30** (2017).
- [43] Kühne, T. D. *et al.* Cp2k: An electronic structure and molecular dynamics software package-quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics* **152** (2020).
- [44] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **77**, 3865 (1996).

- [45] Blöchl, P. E. Projector augmented-wave method. *Physical review B* **50**, 17953 (1994).
- [46] Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of chemical physics* **132** (2010).
- [47] Larsen, A. H. *et al.* The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- [48] Berendsen, H. J., Postma, J. v., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81**, 3684–3690 (1984).
- [49] han xu. Evidential Deep Learning for Interatomic Potential (2025). URL https://figshare.com/articles/dataset/Evidential_Deep_Learning_for_Interatomic_Potential/28805819.