针对动力学可观测量优化力场的路径概率梯度

Zhongyao Wang[1,2,3,*], Lun Song[1,4,*], Yakun Zeng[5,*],Xingao Gong[6],Rui Shi[5,†],MaoSu[1,7,†]

[1]上海人工智能实验室，中国上海 *200232*[2]复旦大学智能机器人与先进制造学院，中国上海 *200433*[3]上海创新研究院，中国上海 *200231*[4]华中科技大学，中国武汉 *430074*[5]浙江大学物理学院，中国杭州 *310058*[6]计算物理科学重点实验室（教育部），计算物理科学研究所，表面物理国家重点实验室及复旦大学物理系，中国上海 *200433*[7]中国科学院深圳先进技术研究院，中国深圳 *518055*

arXiv:2611.32851v1 [physics.chem-ph] 24 Nov 2026

*这些作者贡献相同。 †通信作者：ruishi@zju.edu.cn, sumao@pjlab.org.cn。**机器学习力场继承了用于生成其训练数据的电子结构方法所带来的误差，而通过实验测量对其进行校准，为突破这一局限提供了直接途径。然而，由于对混沌分子轨迹求导会产生不稳定的梯度，针对动力学可观测量进行这类校准仍然很困难。在此，我们提出 PathGrad，通过对朗之万路径概率求导，而非直接对轨迹求导。由此得到的梯度估计量根据热噪声和力导数在线累积，其方差在轨迹长度增加时仍保持稳定。我们通过有限差分计算验证了该梯度估计量，并利用它针对实验扩散系数及结构约束，对熔融铜和水的势能函数进行微调。对于水，同时优化扩散和氧—氧径向分布函数，也能改善剪切黏度和转动弛豫的预测，而这两者均未纳入优化目标。在单一温度下进行校准，也能改善多个动力学可观测量的表观活化能，使其温度依赖性更接近实验结果。此外，与独立的更高层次量子化学参考相比，优化后的势能函数表现出更小的力误差，这表明与实验结果校准能够提高其保真度**

势能面中超出拟合目标的部分。因此，PathGrad 建立了一个通用

框架，可将实验可观测量直接纳入力场优化。

1 引言

分子动力学（MD）模拟的预测准确性取决于力场的质量。经典经验力场通常基于具有物理依据的函数形式构建，并通过拟合量子力学计算与实验可观测量的组合进行参数化。机器学习力场大幅提升了势能面的灵活性，能够高精度地再现第一性原理能量和力[1]。然而，仅以自下而上的方式训练的机器学习力场仍受其参考电子结构方法的局限。近期的实验基准凸显了这一差异，表明即使势能能够准确拟合量子力学力，也不一定能再现实验可观测量[2]。因此，一个日益重要的研究方向是：直接根据实验测量改进力场[3]。

近期研究已证明，利用与时间无关的可观测量进行实验对齐是可行的。可微分轨迹重加权（DiffTRe）能够针对静态可观测量优化机器学习力场，而无需对分子轨迹求导[4]。后续研究结合量子力学数据与实验数据来改进机器学习力场[5],，而通过实验密度对齐，液体电解质性质的预测得到了改善[6]。然而，尽管在静态实验性质对齐方面取得了这些进展，将扩散系数和黏度等动力学可观测量直接纳入力场优化，仍然要困难得多。

动力学可观测量能够提供有关分子运动和弛豫过程的信息，而仅凭静态性质无法完整刻画这些过程。不同于可由在平衡状态下采样的瞬时构型评估的静态结构与热力学量，动力学可观测量取决于分子轨迹沿时间的关联。一个自然的方法是对这些轨迹关于力场参数求导。然而，分子动力学本质上具有混沌性，微小的参数扰动可能导致轨迹敏感性呈指数增长，使长时间梯度估计在数值上变得不稳定[7]。近期，伴随微分[8] 和可逆分子模拟[9]方法的发展，使得能够在力场优化中使用动力学可观测量。尽管如此，这些方法仍保留了单条轨迹对参数扰动的敏感性，通常需要截断梯度传播或关联函数，以控制数值不稳定性。这类近似可能损害梯度的准确性，尤其是对于受长时间动力学支配的可观测量而言。

一种概念上不同的途径，是对路径的概率而非路径本身求导。这类得分函数（似然比）估计器[10,11] 不会继承混沌动力学的敏感性。对于朗之万动力学，路径概率可由Girsanov/Onsager–Machlup作用量得到，并已用于路径重加权[12,13] 和参数敏感性分析[14,15]。Bolhuis等人[16] 利用路径重加权将速率常数施加到力场参数上，但仅适用于参数不超过两个的低维模型的过阻尼动力学。路径得分能否驱动针对输运系数的多参数力场优化，至今仍是一个悬而未决的问题。

本文提出 PathGrad，这是一种路径概率梯度框架，用于针对实验动力学可观测量优化力场。PathGrad 建立在似然比敏感性分析的基础上，根据离散化朗之万轨迹的概率测度评估参数梯度，并使用仿真过程中累积的热噪声和力导数。由于静态可观测量也通过与构型得分相同的协方差形式纳入，因此可以同时优化动力学和结构目标。我们首先在二维玩具模型上将该估计器与有限差分进行比较，以验证其有效性；随后将 PathGrad 应用于熔融铜的嵌入原子模型（EAM）[17] 和水的神经进化势（NEP）[18] 。对于熔融铜，PathGrad 微调一个 12参数嵌入原子势，使其符合实验扩散系数，同时保持晶格常数和体积模量。对于水，则针对实验扩散系数和 O–O 径向分布函数（RDF）优化 NEP 模型。尽管 O–H 和 H–H 径向分布函数、剪切黏度以及旋转弛豫时间均未纳入损失函数，但这些指标也有所改善。值得注意的是，在单一温度下进行对齐，也能改善这些动力学可观测量的表观活化能，使其温度依赖性与实验结果显著更加一致。此外，相对于 CCSD(T)级参考数据的力误差也有所降低，这表明这种实验对齐提升了拟合目标之外的势能面保真度。这些结果表明，PathGrad 是一种可直接根据实验可观测量优化力场的通用框架。

2 结果

路径概率梯度（PathGrad）。我们考虑一个力场$\mathbf{F}_\theta(\mathbf{r}) = -\nabla_{\mathbf{r}} U_\theta(\mathbf{r})$，其参数为 θ，其中$\mathbf{r}$ 表示系统中所有粒子的坐标。压力和径向分布函数等静态性质仅取决于$\mathbf{r}$ 和 θ，而扩散系数等动力学性质则取决于完整轨迹 $\Gamma = \{\mathbf{r}_t\}$。可观测量 O 对参数的梯度

可通过以下方式计算：

$$\nabla_\theta \langle O \rangle_\theta = \langle \nabla_\theta O \rangle_\theta + \langle (O - \langle O \rangle) \nabla_\theta \log p_\theta \rangle_\theta \tag{1}$$

其中 p_θ 为概率密度函数。对于静态可观测量，这可化简为热力学涨落公式，该公式是 DiffTRe[4]等系综重加权方案的基础。对于动力学可观测量，最直观的方法是将动力学可观测量视为初始位置和力场参数的函数。在这种情况下，我们需要使用可微仿真[19,20]等方法直接对轨迹求导。然而，在分子动力学轨迹中，该梯度的方差会随轨迹长度呈指数增长，从而妨碍其实际应用。为克服这一限制，我们提出将整个轨迹而非初始位置视为随机变量。在这种情况下，等式（1）右侧的第一项消失，现在的目标是推导轨迹的概率密度。我们考虑离散化的朗之万动力学

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n \Delta t \tag{2}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \left[\frac{\mathbf{F}_\theta(\mathbf{r}_n)}{M} - \gamma \mathbf{v}_n \right] \Delta t + \sqrt{\frac{2\gamma k_B T \Delta t}{M}}\, \xi_n \tag{3}$$

其中 $\xi_n \sim \mathcal{N}(\mathbf{0},\ \mathbf{I})$ 是相互独立的标准高斯向量。为便于说明，这里使用欧拉–丸山格式；数值实现则采用补充材料中[21,22] 详细介绍的 A-BO 朗之万积分器。对于轨迹 $\Gamma = \{\mathbf{r}_n,\ \mathbf{v}_n\}$，可通过对等式（3）求逆来推断每一步的噪声：

$$\xi_n(\Gamma, \theta) = \sqrt{\frac{M \Delta t}{2\gamma k_B T}} \left[\frac{\mathbf{v}_{n+1} - \mathbf{v}_n}{\Delta t} + \gamma \mathbf{v}_n - \frac{\mathbf{F}_\theta(\mathbf{r}_n)}{M} \right] \tag{4}$$

由于每个 ξ_n 都是标准高斯，在给定初始条件 $\mathbf{r}_0$ 和 $\mathbf{v}_0$ 的情况下，路径概率为

$$p_\theta(\Gamma |\, \mathbf{r}_0, \mathbf{v}_0) = \prod_n \exp\left[-\tfrac{1}{2} \xi_n(\Gamma, \theta)^2 \right], \tag{5}$$

注意，归一化常数可以忽略，因为式 (1) 不要求概率密度归一化。当 $\Delta t \to 0$ 趋于极限时，该概率密度便成为昂萨格–马赫卢普作用量。总概率为 $p_\theta(\Gamma) = p_\theta(\Gamma|\mathbf{r}_0,\mathbf{v}_0)\, p_\theta(\mathbf{r}_0,\mathbf{v}_0)$，其中 $p_\theta(\mathbf{r}_0,\mathbf{v}_0)$ 是

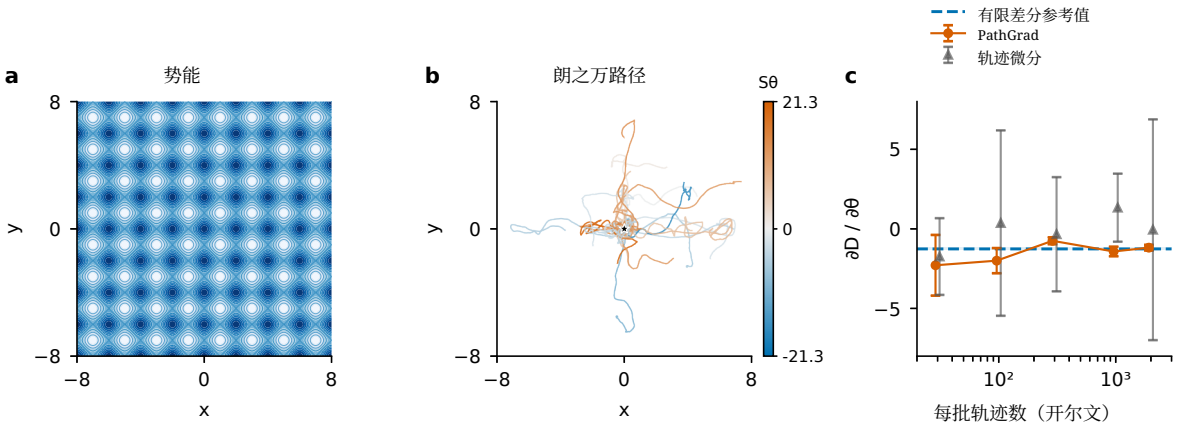


图1：二维玩具模型上的路径得分微分。（**a**）起伏势能$U(x,\ y;\ \theta)\ =\ -\theta[\cos(\pi x)\ +\cos(\pi y)]$，其中 $\theta\ =\ 0.5$。颜色越深表示势能越低；相邻极小值之间的势垒为 2θ。（**b**）24条从原点静止状态出发的示例性朗之万轨迹，按其路径得分着色 $S_\theta[\Gamma_i]$。梯度由扩散估计值与路径得分之间的协方差确定。（**c**）PathGrad 和直接轨迹微分得到的 $\partial D/\partial\theta$ 估计值随每批轨迹数的变化。点和误差棒表示五个独立批次的均值和样本标准差。蓝色虚线及阴影带表示有限差分参考值 -1.261 ± 0.093，该值由五个批次（每批 $K\ =\ 2000$）得出。

玻尔兹曼分布。对数概率的梯度为

$$\nabla_\theta \log p_\theta(\Gamma) = S_\theta[\Gamma] - \beta\nabla_\theta U(\mathbf{r}_0) \tag{6}$$

其中，$S_\theta[\Gamma] = \nabla_\theta \log\ p_\theta(\Gamma|\mathbf{r}_0,\mathbf{v}_0)$ 是路径得分函数。由式（4）和式（5），可得

$$S_\theta[\Gamma] = \sqrt{\frac{\Delta t}{2\gamma M k_B T}} \sum_n \xi_n \cdot \nabla_\theta \mathbf{F}_\theta(\mathbf{r}_n) \tag{7}$$

该路径得分 $S_\theta[\Gamma]$ 在标准朗之万模拟过程中通过逐步累积 $\boldsymbol{\xi}_n$ 和 $\nabla_\theta\mathbf{F}_\theta(\mathbf{r}_n)$ 进行在线计算。它不需要对轨迹求导，从而避免了混沌动力学中的李雅普诺夫不稳定性。微调过程如下：在每次迭代中，我们并行运行多个独立模拟。在模拟过程中，我们不仅演化位置和速度，还根据公式（7）累积路径得分函数。然后，我们计算每个模拟的动力学可观测量，并使用公式（1）计算梯度。

玩具模型。 我们首先在一个可以通过有限差分独立估计梯度的系统上验证路径得分估计器。单个粒子在二维波纹势 $U(x,\ y;\ \theta) = -\theta[\cos(\pi x) + \cos(\pi y)]$, 下进行朗之万动力学，其中 θ 设定了相邻极小值之间势垒的高度（图 1a）。增大 θ 会抑制热激活跃迁并降低扩散系数 D，因此 $\partial D/\partial\theta < 0$。我们从 $\mathbf{r}_0 = (0,\ 0)$ 出发传播 K 条独立轨迹，并根据它们的均方位移（MSD）估计 D。

图1b展示了不同的轨迹。每条轨迹 Γ_i 都按其路径得分$S_\theta[\Gamma_i] = \partial_\theta \log p_\theta[\Gamma_i]$着色。得分为正意味着参数 θ 略微增大时，这条路径的相对似然会增加；得分为负则意味着其似然会降低。扩散得更远的轨迹往往得分为负，而移动性较弱的轨迹往往得分为正。因此，它们的负协方差 $\langle\ (D_i - \langle D\rangle)\ \partial_\theta \log\ p_\theta[\Gamma_i]\rangle$ 给出 $\partial D/\partial\theta < 0$。协方差量化了参数 θ 略微变化时，高移动性路径与低移动性路径之间的概率如何重新分配。

图1c比较了随着批量大小 K 增加，路径得分估计器与直接轨迹微分的表现。路径得分估计值收敛至有限差分参考值，其统计不确定性约按 $1/\sqrt{K}$下降。相比之下，即使在 $K = 2000$时，直接轨迹微分的变化仍然很大，得到 -0.06 ± 6.94，其离散范围远大于梯度值本身。在势垒顶端附近，动力学包含一个不稳定方向，因此微小的参数扰动可能会被大幅放大，并显著改变穿越事件[7,23]发生的时机或是否发生。路径得分微分根据采样路径统计权重的变化估计梯度，因此无需传播这种轨迹敏感性。

微调熔融铜的嵌入原子势。Mishin等人[17] 的嵌入原子模型（EAM）根据晶格结构、弹性性质和从头算数据拟合而成，并能准确描述铜中的缺陷。在这里，我们使用一个拟合于 Mishin 势能的 12 参数解析表示（方法），它能够再现晶体晶格性质，却无法再现实验扩散系数。我们将其作为首个原子级演示，通过微调势能，在保持结构保真度的同时恢复目标扩散系数。

具体而言，我们同时针对三个目标微调势能：扩散系数 D、晶格常数 a_0和体积模量 B。结构约束通过一个 $4 \times 4 \times 4$ 面心立方超胞的总能量 $U(L)$相对于其边长 L 的导数施加。在平衡超胞边长 $L_0 = 4a_0$ 处，压力为零，从而得到体积模量

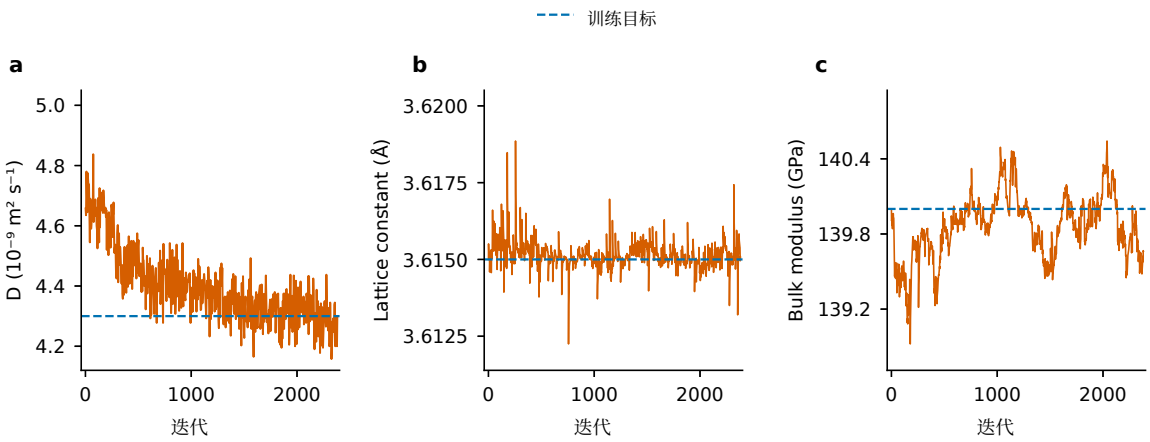


图 2：微调熔融铜的 EAM势。(a) 朗之万扩散估计值 D 随优化迭代的变化，并标出实验目标 $D_{\mathrm{target}} = 4.3 \times 10^{-9}\,\mathrm{m^2\,s^{-1}}$ （虚线）。**(b)** 平衡晶格常数 a_0 和**(c)** 体积模量 B 随优化迭代的变化，目标值分别为 $a_0 = 3.615$Å 和$B = 140$GPa（虚线）。

为 $B = (9L)^{-1}\partial^2 U/\partial L^2|_{L=L_0}$。组合损失为

$$\mathcal{L} = \lambda_1\left(D - D_{\mathrm{target}}\right)^2 + \left[\lambda_2\left(\frac{\partial U}{\partial L}\right)^2 + \lambda_3\left(\frac{\partial^2 U}{\partial L^2} - 9BL\right)^2\right]_{L=L_0}, \tag{8}$$

wh其中 D_{target} 是实验目标[24–27]，而 λ_1， λ_2， λ_3 是可调权重。

优化轨迹如图 2 所示。扩散系数从较大的值开始，并在最初几百次迭代内收敛至目标值。在校正 D 的过程中，结构约束会暂时受到扰动，随后在接下来的迭代中逐渐恢复。

水的机器学习力场微调。 接下来，我们将 PathGrad 应用于机器学习力场，此时参数空间更大，势能也更加灵活。神经进化势（NEP）[18] 在保持对参考量子化学数据的高保真度的同时，具有很高的计算效率。当使用接近 CCSD(T) 精度的 MB-pol 参考数据[28,29] 进行训练时，CCSD(T) 是分子量子化学的基准“黄金标准”[30]，NEP模型能够准确再现水的实验径向分布函数、密度和扩散系数[31]。然而，当使用采用 SCAN 元广义梯度近似泛函的密度泛函理论（DFT）数据[32,33], 进行训练时，扩散系数显著低于实验目标（补充表 S2）。我们从一个在 SCAN数据集[35], 上训练的公开可用的NEP模型[34]出发，对其进行微调，以恢复实验扩散系数，同时

拟合 O–O径向分布函数（径向分布函数），以防止结构退化，所使用的组合损失为

$$\mathcal{L} = (D - D_{\mathrm{target}})^2 + \lambda_S \sum_i \big[g(r_i) - g_0(r_i)\big]^2, \tag{9}$$

其中 $g(r)$按帧计算为高斯展宽的径向分布函数（宽度 $\varepsilon = 0.1$Å），$g_0(r)$为以相同宽度平滑的中子散射参考[36]。高斯展宽提供平滑、可微的目标，并抑制传统直方图的高方差分箱噪声。径向分布函数梯度由式(1)得出，使用构型得分 $-\beta\nabla_\theta U_\theta$(r)以及$\beta = 1/(k_{\mathrm{B}}T)$。

优化结果如图3所示。扩散系数从 $0.61 \times 10^{-9}\ \mathrm{m^2\,s^{-1}}$ 上升，并在约450次迭代后达到 $2.3 \times 10^{-9}\mathrm{m^2\,s^{-1}}$ 的训练目标（图3a）。径向分布函数的均方根误差也从0.230降至0.055（图3b）。相应地，最终的氧-氧径向分布函数明显趋近于X射线散射参考[37]（图4a），而未在损失函数中受到约束的O–H和H–H径向分布函数也趋近于其实验参考值（图4b、c）。我们还每10次迭代计算一次相对于CCSD(T)级MB-pol参考数据集[31]的力均方根误差（图3c）。尽管损失函数中未包含CCSD(T)力，力均方根误差在前450次迭代中仍下降了7.5%。这表明对动力学的校正并非以牺牲力场保真度为代价；相反，微调后的模型在力空间中也更接近高水平参考。

随后，我们将剪切黏度 η 和分子内 H–H 矢量的二阶旋转弛豫时间 τ_2 与实验测量进行比较，以确定微调是否改善了对损失函数中未包含的可观测量的描述。如图 4d–f 所示，初始模型的扩散和重新取向较慢，且黏度过高；微调后，所有性质均有所改善。

由于模型是使用单一温度 300 开尔文下的轨迹进行微调的，我们还评估了它在其他温度下的表现。在不重新训练的情况下，我们评估了 D, η 和 τ_2 在 288.15、298.15 和 308.15 K 下的表现，各温度均采用该温度下的实验密度[38]（图 5）。微调前，模型与实验结果的差异会随着降温而增大；微调后，模型在整个 20 K 温度区间内都与测量值接近，尽管这些温度均未纳入训练。值得注意的是，微调还使这些动力学可观测量的表观活化能显著接近实验值（补充图S3和表 S4）。因此，单一温度下的实验对齐不仅改善了

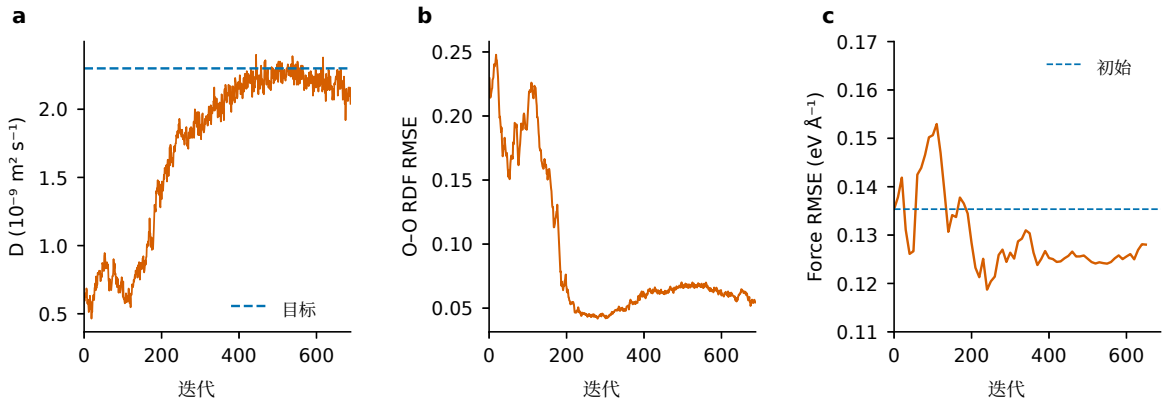


图 3：水势微调。(a) 优化过程中的扩散及其目标值（虚线）。**(b)** 优化过程中的氧-氧径向分布函数均方根误差。**(c)** 相对于 CCSD(T) 级别参考值的力均方根误差[31]，每 10 次迭代评估一次。

分子输运和弛豫的数值，也改善了它们的温度依赖性，这表明对底层势能面的修正具有明确的物理意义。

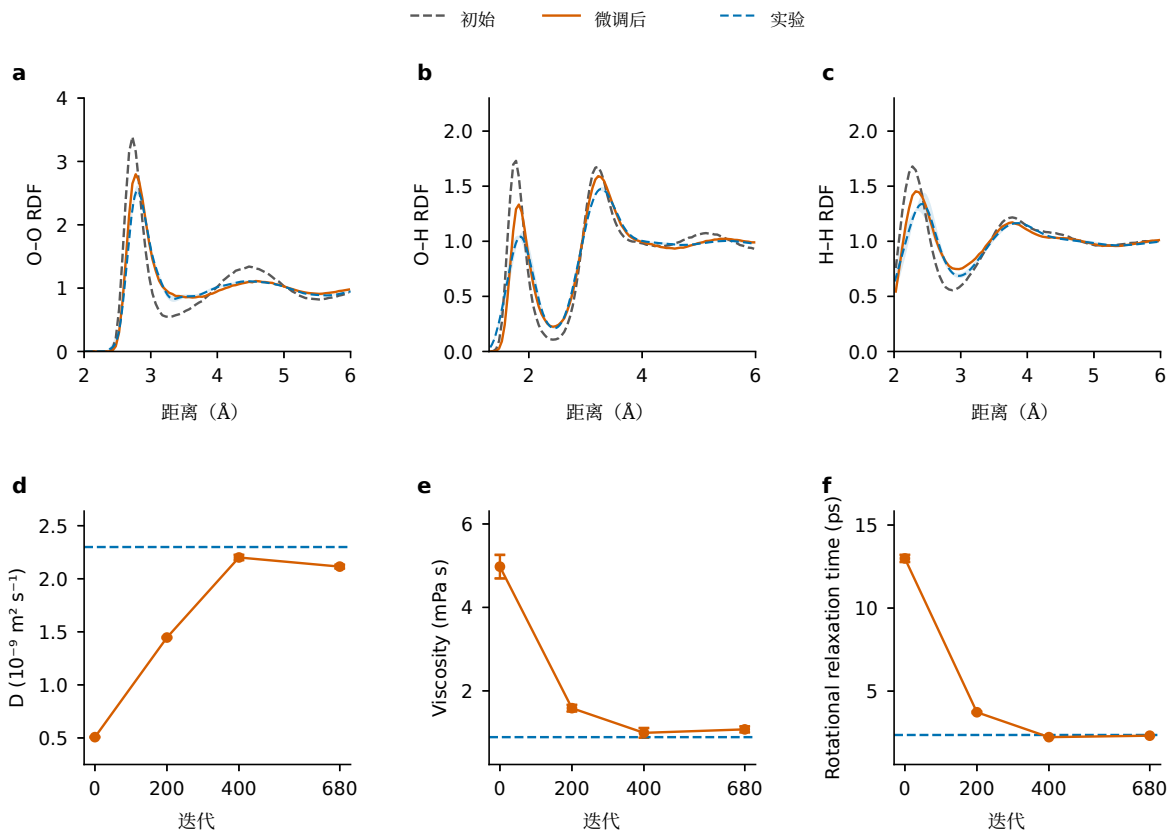


图4：微调前后的水结构与动力学。所有结果均在$T = 298.15$开尔文和 $\rho = 0.99704\,\mathrm{g\ cm^{-3}}$条件下评估。**(a–c)** 初始模型和微调后模型（第680次迭代时）的氧-氧、氧-氢和氢-氢径向分布函数。蓝色虚线为实验参考值：氧-氧径向分布函数来自299.5开尔文下的X射线散射[37]；氧-氢和氢-氢径向分布函数来自298开尔文下对散射数据进行的经验势结构精修（EPSR）[39]。阴影表示相应来源报告的不确定性或变化范围。氧-氢（$r \geq 1.3$Å）和氢-氢（$r \geq 2.0$Å）仅显示分子间区域。**(d–f)** 第0、200、400和680次迭代时的扩散系数 D、剪切黏度η 和氢-氢转动弛豫时间 τ_2 。符号表示五条独立的5纳秒轨迹的均值，误差棒为样本标准差。部分误差棒过小，在图中无法看见。虚线表示实验值[40–43]。优化损失函数仅包含 D和氧-氧径向分布函数。估计窗口和对应曲线见补充图S2。

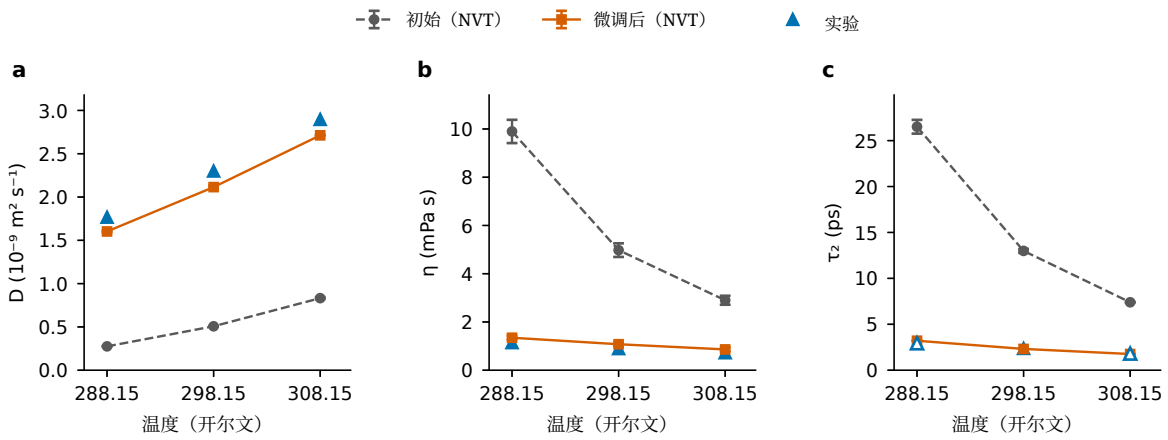


图5：在 300 K 下微调后的温度可迁移性。（**a**）扩散系数 D、（**b**）剪切黏度 η 和（**c**）旋转弛豫时间 τ_2 在 288.15、298.15 和 308.15 K 下的数值，分别为在 300 K 优化前（初始，灰色）和优化后（微调后，橙色）（方法）。各温度下的模拟均在相应实验密度[38]下采用 NVT 系综进行。符号和误差棒分别表示五条独立轨迹的平均值和样本标准差。部分误差棒过小，在图中无法辨认。实验参考值[40–43] 以三角形表示。空心三角形表示 288.15 和 308.15 K 下的 τ_2 参考值，这些数值根据拟合的实验曲线数字化提取[44] （见补充表S3）。

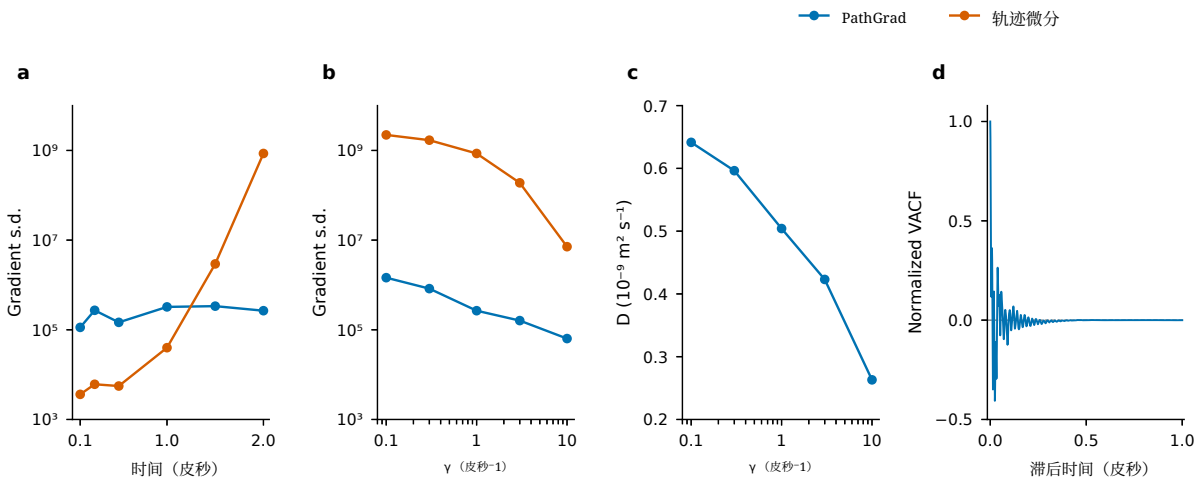


图6：初始 NEP 水模型的梯度标准差与动力学。（**a**）梯度标准差随轨迹长度的变化，$\gamma = 1$皮秒$^{-1}$。更长轨迹的结果见补充图 S1。（**b**）梯度标准差随 γ 的变化，轨迹长度为 2 皮秒。两种方法均用于估算扩散观测量相对于势参数的梯度。（**c**）扩散系数 D随 γ的变化。每个数据点均由一条 5 纳秒轨迹通过氧原子MSD拟合得到。（**d**）在 $\gamma = 1$皮秒$^{-1}$时的全原子速度自相关函数（VACF），以零时滞处的系综平均值归一化。

3 讨论

对路径概率而非轨迹本身求微分，可以避免混沌分子动力学中随轨迹长度呈指数增长的梯度方差。对于 NEP 水模型，在 2皮秒轨迹上进行直接轨迹微分时，其方差已经比 PathGrad 方差大几个数量级（图 6a），并且梯度从第一次优化步骤开始就发生数值溢出。相比之下，路径得分的方差仅呈线性增长（见补充材料），这与似然比估计量[10,45]的预期一致。同时，以时间起点为平均基准的扩散估计值的统计误差会随 t减小。这两种趋势在很大程度上相互补偿，因此梯度估计量的方差仅较弱地依赖于轨迹长度，如补充图 S1所示。

一个重要的实际考量是朗之万摩擦系数 γ的选择，它决定了梯度噪声与动力学保真度之间的权衡。在轨迹长度固定时，路径得分的方差按 $1/\gamma$ 缩放（图 6b），因此较小的摩擦需要更多独立轨迹用于训练。较大的摩擦会产生噪声更小的梯度，但也会更强地扰动输运过程（图 6c）。较小的 γ 值对内禀动力学的干扰较小，所得结果也更接近微正则（NVE）系综中的结果。本文采用的数值 $\gamma = 1$ 皮秒$^{-1}$是朗之万恒温器的常用选择，也用于可逆模拟[9]。

每条独立轨迹的长度是另一个关键参数。由于 PathGrad 方差几乎不随轨迹长度变化（图 6a 和补充图 S1），因此轨迹长度可以根据物理因素而非数值因素来选择：轨迹应足够长，使目标可观测量进入相关的长时间动力学状态。根据速度自相关函数（VACF）的衰减情况（图 6d），我们为铜选择了 1 皮秒，为水选择了 2 皮秒，这对于扩散系数而言已经足够。对于弛豫较慢的可观测量，例如过冷液体中的剪切黏度或重新取向，则需要更长的时间窗口；对于这些目标，PathGrad 方差对长度的弱依赖性尤其具有优势。当有额外的计算资源可用时，我们增加的是独立轨迹的数量，而不是其长度，因为梯度方差会随着 $1/K$ 减小。

机器学习力场采用规模不断扩大的电子结构数据集进行训练，但其准确度仍受参考方法的限制，参考方法通常为密度泛函理论。利用实验数据进行学习，是突破这一上限最直接的途径。分子动力学模拟预测的实验可观测量大致分为结构量和动力学量。构象重加权[4] 已能为结构可观测量提供高效且通用的梯度，

而对动力学可观测量求导仍然是一项挑战[8,9,16]。PathGrad 在标准朗之万模拟过程中累积一个未经截断的梯度估计量，从而通过单一方法对结构和动力学可观测量求导。原则上，它适用于任何能够表示为有限长度朗之万路径平均值的可观测量。水体系的结果展示了这种方法的潜力。具体而言，尽管损失函数中没有使用量子化学数据，相对于 CCSD(T)级参考数据的力误差仍降低了约 7.5%。因此，PathGrad 提供了一条统一的基于梯度的路径，可将结构和动力学实验数据都作为机器学习力场的直接学习目标，同时消除限制动力学训练目标时间窗口的梯度不稳定性。

方法

仿真与微调方案 所有仿真均采用朗之万 A–BO 离散化，系统专属参数汇总于表 S5。

玩具模型。 在无量纲单位下，我们设定 $M = T = 1$、 $\gamma = 0.5$、 $\Delta t = 0.005$ 和 $\theta = 0.5$。每条轨迹均从势能最低点 $\mathbf{r}_0 = (0,0)$ 处静止开始，并积分 2000 步。中心有限差分采用扰动 $\theta \pm 0.02$，并使用共同随机数。路径分数协方差在每个批次内分别对 D_k 和 $S_\theta[\Gamma_k]$ 去中心化。误差棒表示五个独立批次估计值的样本标准差。

铜。 我们采用一个包含 12 个参数的解析表示，该表示拟合了 Mishin等人[17],的 EAM势，其中包括嵌入函数 $F(\bar{\rho}) = F_C[\exp(-F_k\bar{\rho}) - 1]$（两个参数）、对势 $\psi(r) = \sum_{i=1}^{4} C_i \exp(-k_i r)$（八个参数），以及宿主电子密度函数$\rho(r) = C_\rho \exp(-k_\rho r)$（两个参数），其中 $\bar{\rho}$ 为相邻原子电子密度之和。我们在 $T = 1500$ K 和质量密度 $\rho = 7.94$ g cm^{-3} 条件下，使用 $\gamma = 1$ ps^{-1} 和 $t = 1$ 模拟 256 个铜原子，并采用式（8）中的组合损失。结构约束根据一个 $4 \times 4 \times 4$ 面心立方超胞（256 个原子）的总能量 $U(L)$ 计算，该超胞边长为 L，在 $L_0 = 4a_0 = 14.46$ Å 处评估；这对应于实验晶格常数 $a_0 = 3.615$ Å 和体积模量 $B = 140$ GPa[46,47]。每个优化步骤中，在更新参数之前，先将动力学梯度投影到与结构梯度正交的方向上。

水。 从一个公开可用的NEP模型[34] 出发，该模型使用基于SCAN泛函[35],生成的DFT数据训练。我们模拟了512个水分子（1536个原子），密度为 $\rho = 0.994$g cm^{-3}，模拟时长为$\gamma = 1$ps^{-1}，轨迹长度为 $t = 2$ ps。Langevin恒温器的目标温度设为 $T = 300$ K，初始速度从玻尔兹曼分布中抽取；训练期间记录的动能温度在297至301 K之间（平均为299 K）。结构目标 g_0（r）为中子散射氧-氧径向分布函数[36],，并采用宽度为 $\varepsilon = 0.1$Å的高斯核进行平滑； g（r）则对每一帧使用相同的核进行计算。

输运系数 **分析** 扩散系数由爱因斯坦关系定义：

$$D = \lim_{t\to\infty} \frac{1}{6t} \left\langle |\mathbf{r}(t_0 + t) - \mathbf{r}(t_0)|^2 \right\rangle , \tag{10}$$

其中，尖括号表示对时间起点 t_0 和原子进行平均。实际上，我们通过拟合均方位移（MSD）的线性阶段来提取 D。

水的剪切黏度通过格林–久保关系获得，该关系基于对称无迹应力张量 $\Pi_{\alpha\beta} = (\sigma_{\alpha\beta} + \sigma_{\beta\alpha})/2 - \delta_{\alpha\beta}\,\mathrm{tr}(\sigma)/3$的自相关。由于液态水具有各向同性，可以将 Π 的所有分量结合起来以改善统计结果[48],，从而得到运行积分

$$\eta(t) = \frac{V}{10 k_B T} \int_0^t \sum_{\alpha,\beta} \langle \Pi_{\alpha\beta}(t_0)\Pi_{\alpha\beta}(t_0 + s)\rangle_{t_0}\,\mathrm{d}s,$$

其中 V为体积，求和遍及全部九组笛卡尔分量对（α，β），包括非对角指标的两种排列，因此产生前置因子 $1/10$。实际计算中，我们累积滞后时间上的积分，并将 25–75 ps 范围内的平台均值作为 η。

分子重取向由单位分子内 H–H 向量 $\hat{u}$ 的二阶勒让德时间相关函数表征：

$$\tau_2 = \int_0^\infty C_2^{\mathrm{HH}}(t)\,\mathrm{d}t, \qquad C_2^{\mathrm{HH}}(t) = \big\langle P_2\big(\hat{u}(t_0)\cdot\hat{u}(t_0+t)\big)\big\rangle, \tag{11}$$

其中 $P_2(x) = \frac{1}{2}(3x^2 - 1)$ 为第二勒让德多项式，平均值是对所有水分子和时间起点求得的。在适用于室温附近液态水的极限窄化区，τ_2 是决定分子内 ^{1}H 自旋–晶格弛豫[42,43,49]的旋转相关时间。我们使用梯形法则对积分进行数值计算，积分从 $t = 0$ 开始，延伸至积分不再变化的滞后时间。在 298.15 开尔文时，迭代 0、200、400 和 680 对应的记录窗口分别为 149.5 、 69.5 、 39.5 和 29.5 皮秒；在 288.15 开尔文时，初始和微调后的窗口分别为 349.5 和 49.5 皮秒；在 308.15 开尔文时，则分别为 149.5 和 29.5 皮秒。

参考文献

[1] Unke, O. T.等。 机器学习力场。*Chemical Reviews***121**，10142–10186（2021）。

[2] Mannan, S. 等。 UniFFBench：针对实验测量评估通用机器学习力场。*Nature ComputationalScience* **6**，894–906（2026）。URL https: // doi.org/10.1038/s43588-026-01019-4。

[3] Fröhlking, T.、Bernetti, M.、Calonaci, N. 和 Bussi, G. 朝着与实验可观测量相匹配的经验力场迈进。*The Journal of Chemical Physics* **152** （2020）。

[4] Thaler, S. 和 Zavadlav, J. 通过可微轨迹重加权从实验数据中学习神经网络势。*Nature Communications* **12**，6884（2021）。

[5] Röcken, S. & Zavadlav, J. 通过融合实验与模拟数据构建准确的机器学习力场。*npj Computational Materials***10**, 69 (2024).

[6] Gong, S. 等用于液态电解质开发的预测性机器学习力场框架。自然机器智能**7**, 543–552 (2025).[7] Metz, L., Freeman, C. D., Schoenholz, S. S. & Kachman, T. 梯度并非你所需的一切。*arXiv* 预印本 *arXiv:2111.05803*(2021).[8] Han, B. & Yu, K. 通过可微分子模拟，利用动力学性质改进势能面。*Nature Communications***16**, 816 (2025).[9] Greener, J. G. 用于训练经典力场和机器学习力场的可逆分子模拟。美国国家科学院院刊 **122**, e2426058122 (2025).[10] Glynn, P. W. 随机系统的似然比梯度估计。《*ACM*通讯》 **33**, 75–84 (1990).[11] Mohamed, S., Rosca, M., Figurnov, M. & Mnih, A. 机器学习中的蒙特卡洛梯度估计。《机器学习研究杂志》 **21**, 1–62 (2020).[12] Donati, L., Hartmann, C. & Keller, B. G. 用于路径系综和马尔可夫状态模型的 Girsanov 重加权。*The Journal ofChemical Physics***146**, 244112 (2017).[13] Kieninger, S. & Keller, B. G. 朗之万动力学的路径概率比——精确与近似。*TheJournal of ChemicalPhysics***154** (2021).[14] Warren, P. B. & Allen, R. J. 用 Malliavin 权重采样计算布朗动力学模拟中的灵敏度系数。*Physicalreview letters* **109**, 250601 (2012).[15] Warren, P. B. & Allen, R. J. 通过轨迹重加权进行随机建模中的稳态参数灵敏度分析。*TheJournalof chemicalphysics* **136** (2012).[16] Bolhuis, P. G., Brotzakis, Z. F. & Keller, B. G. 通过路径重加权施加动力学约束，优化分子势模型。*The Journal of Chemical Physics* **159**, 074102 (2023).[17] Mishin, Y., Mehl, M. J., Papaconstantopoulos, D. A., Voter, A. F. & Kress, J. D. 铜的结构稳定性与晶格缺陷：第一性原理、紧束缚及嵌入原子计算。《物理评论 *B*》 **63**, 224106 (2001).[18] Fan, Z. 等神经进化机器学习势：兼顾高精度与低成本的原子模拟及其在热输运中的应用。《物理评论 *B*》 **104**, 104309 (2021).

[19] Schoenholz, S. S. & Cubuk, E. D. JAX MD：可微物理框架。收录于神经信息处理系统进展，第 33 卷，11428–11441 页（Curran Associates, Inc., 2020）。网址：
h
t
t
p
s://
p
r
o
c
e
e
di
ngs.neurips.cc/paper/2020/hash/83d3d4b6c9579515e1679aca8cbc8033-Abstract.html。[20] Wang, X. 等DMFF：用于分子力场开发和分子动力学模拟的开源自动可微平台。*Journal of Chemical Theory andComputation* **19**，5897–5909（2023）。[21] Leimkuhler, B. & Matthews, C. 用于分子采样的随机数值方法的合理构建。《应用数学研究快报》**2013**，34–56（2013）。[22] Leimkuhler, B. & Matthews, C. 通过朗之万动力学实现稳健高效的构型分子采样。*TheJournalof chemicalphysics***138**（2013）。[23] Suh, H. J., Simchowitz, M., Zhang, K. & Tedrake, R. 可微模拟器能否带来更好的策略梯度？收录于第 *39* 届国际机器学习会议论文集，第 162 卷机器学习研究论文集，20668–20696 页（2022）。[24] Meyer, A. 准弹性中子散射揭示的液态铜自扩散。《物理评论 *B*》**81**，012102（2010）。[25] Meyer, A. 利用准弹性中子散射测量液态金属的自扩散系数。收录于*EPJ* 会议网络，第 83 卷，01002（EDP Sciences，2015）。[26] Brillo, J.《多组分液态合金的热物理性质》（De Gruyter Oldenbourg，2016）。[27] Hoyt, J., Sadigh, B., Asta, M. & Foiles, S. 基于原子尺度计算得出的铜–镍体系动力学相场参数。《材料学报》**47**，3181–3187（1999）。网址：https://www.sciencedirect.com/science/article/pii/S1359645499001895。[28] Babin, V., Leforestier, C. & Paesani, F. 开发采用柔性单体的“第一性原理”水势：二聚体势能面、VRT光谱和第二维里系数。*Journal of chemical theory and computation* **9**，5395–5403（2013）。[29] Babin, V., Medders, G. R. & Paesani, F. 开发采用柔性单体的“第一性原理”水势。II：三聚体势能面、第三维里系数和小团簇。*Journal of chemicaltheoryand computation* **10**，1599–1607（2014）。[30] Bartlett, R. J. & Musiał, M. 量子化学中的耦合簇理论。《现代物理学评论》**79**，291–352（2007）。

[31] Xu, K. 等。NEP-MB-pol：用于快速、准确预测水的热力学和输运性质的统一机器学习框架。npj Computational Materials 11，279（2025）。[32] Sun, J., Ruzsinszky, A. & Perdew, J. P. 强约束且规范化恰当的半局域密度泛函。Physical Review Letters 115，036402（2015）。[33] Chen, M. 等。水的从头算理论与建模。Proceedings of the National Academy of Sciences 114，10846–10851（2017）。[34] Fan, Z. 等。qNEP：一种具有动态电荷、用于大规模原子模拟的高效神经进化势。Journal of Chemical Theory and Computation 22，4787–4801（2026）。[35] Xu, K. 等。利用神经进化势准确预测水的热导率。The Journal of Chemical Physics 158，204114（2023）。[36] Soper, A. K. 220 至 673 开尔文、压力最高达 400 MPa 时水和冰的径向分布函数。Chemical Physics 258，121–137（2000）。网址：https://www.sciencedirect.com/science/article/pii/S0301010400001798。[37] Skinner, L. B., Benmore, C. J., Neuefeind, J. C. & Parise, J. B. 压缩性最小值附近水的结构。The Journal of Chemical Physics 141，214507（2014）。[38] Kell, G. S. 0 ° 至 150 ° ℃液态水的密度、热膨胀率和压缩性：大气压和饱和条件下的相关关系与数据表，并按照 1968 年温标进行评述和表述。Journal of Chemical & Engineering Data 20，97–105（1975）。[39] Soper, A. K. 通过辐射总散射实验得出的水的径向分布函数：我们究竟能否确定什么？ISRN Physical Chemistry 2013，279463（2013）。[40] Holz, M., Heil, S. R. & Sacco, A. 用于准确¹H NMR 脉冲场梯度测量校准的水及六种选定分子液体的温度依赖性自扩散系数。Physical Chemistry Chemical Physics 2，4740–4742（2000）。[41] Korson, L., Drost-Hansen, W. & Millero, F. J. 不同温度下水的黏度。The Journal of Physical Chemistry 73，34–39（1969）。[42] Jonas, J., DeFries, T. & Wilbur, D. J. 压缩水中的分子运动。The Journal of Chemical Physics 65，582–588（1976）。

[43] Kumar, R. & Skinner, J. L. 具有显式三分子相互作用的水模拟模型。*The Journal of Physical ChemistryB***112**，8311–8318（2008 年）。

[44] Krynicki, K. 介于 0°开尔文和 100°开尔文之间纯水中的质子自旋-晶格弛豫。《物理学》**32**，167–178（1966 年）。[45] Arampatzis, G.、Katsoulakis, M. A. & Rey-Bellet, L. 复杂随机动力学似然比敏感度指标的高效估计器。*The Journal of Chemical Physics* **144**（2016 年）。[46] Arblaster, J. W. 元素晶体学性质的精选值（ASM International, 2018 年）。[47] Haynes, W. M.（编）《*CRC*化学与物理手册》（CRC出版社，2016 年），第 97 版。[48] González, M. A. & Abascal, J. L. F. 刚性水模型的剪切黏度。*The Journal ofChemical Physics* **132**，096101（2010 年）。[49] Bloembergen, N.、Purcell, E. M. & Pound, R. V. 核磁共振吸收中的弛豫效应。《物理评论》**73**，679–712（1948 年）。

致谢 本工作得到国家自然科学基金委员会（项目编号：12404291）和新一代人工智能国家科技重大项目（2025ZD0121802）的资助。Z.W. 和 L.S. 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 R.S. 提出了这一想法并设计了研究。L.S. 和 Z.W. 发展了路径评分形式体系并实现了代码。Y.Z. 完成了输运系数模拟和分析。X.G.、R.S. 和 M.S. 参与了分析并提供了科学见解。L.S. 撰写了初稿。所有作者均审阅并批准了最终稿件。

利益冲突 作者声明不存在利益冲突。

通信作者材料相关的通信和请求请联系 R.S.（ruishi@zju.edu.cn）或 M.S.（sumao@pjlab.org.cn）。

《利用路径概率梯度针对动力学可观测量优化力场》的补充信息

Zhongyao Wang、Lun Song、Yakun Zeng、Xingao Gong、Rui Shi、Mao S u

目录

S1 理论 **2**

S1.1 系综平均的梯度.2 S1.2 构型得分与结构可观测量.2 S1.3 路径得分与动力学可观测量.3 S1.4 方差.5

S2 水势验证的其他结果。 **7**

图表目录

S1 水梯度的标准差分解.6 S2 水的输运系数提取.9 S3 水动力学的阿伦尼乌斯分析.12

表格目录

S1 水验证协议.8 S2 298.15 开尔文下的输运系数.10 S3 跨温度可迁移性.11 S4 水动力学的表观活化能.11 S5 模拟与优化参数.13

S1 理论

S1.1 系综平均的梯度

我们考虑一个参数为 θ 的力场，并用$\mathbf{r}$表示系统的一个构型，其概率密度为 $p_\theta(\mathbf{r})$。对于依赖于$\mathbf{r}$且可能还依赖于 θ 的可观测量 O，其系综平均为

$$\langle O\rangle_\theta = \frac{\int O(\mathbf{r},\theta)\, p_\theta(\mathbf{r})\,\mathrm{d}\mathbf{r}}{\int p_\theta(\mathbf{r})\,\mathrm{d}\mathbf{r}}. \tag{S1}$$

其参数梯度由商法则得到，

$$\begin{aligned}
\nabla_\theta\langle O\rangle_\theta &= \frac{\int \left(\nabla_\theta O\, p_\theta + O\,\nabla_\theta p_\theta\right)\,\mathrm{d}\mathbf{r}}{\int p_\theta\,\mathrm{d}\mathbf{r}} - \frac{\int \nabla_\theta p_\theta\,\mathrm{d}\mathbf{r}\int O\,p_\theta\,\mathrm{d}\mathbf{r}}{\left(\int p_\theta\,\mathrm{d}\mathbf{r}\right)^2} \\
&= \frac{\int \nabla_\theta O\, p_\theta\,\mathrm{d}\mathbf{r}}{\int p_\theta\,\mathrm{d}\mathbf{r}} + \frac{\int O\,\nabla_\theta \log p_\theta\, p_\theta\,\mathrm{d}\mathbf{r}}{\int p_\theta\,\mathrm{d}\mathbf{r}} - \frac{\int \nabla_\theta \log p_\theta\, p_\theta\,\mathrm{d}\mathbf{r}}{\int p_\theta\,\mathrm{d}\mathbf{r}}\frac{\int O\,p_\theta\,\mathrm{d}\mathbf{r}}{\int p_\theta\,\mathrm{d}\mathbf{r}} \\
&= \langle\nabla_\theta O\rangle_\theta + \langle O\,\nabla_\theta\log p_\theta\rangle_\theta - \langle\nabla_\theta\log p_\theta\rangle_\theta\langle O\rangle_\theta \\
&= \langle\nabla_\theta O\rangle_\theta + \left\langle \left(O-\langle O\rangle_\theta\right)\nabla_\theta\log p_\theta\right\rangle_\theta .
\end{aligned} \tag{S2}$$

这就是随机模拟、运筹学和强化学习 [1–5] 中使用的得分函数或似然比梯度估计量：得分 $\nabla_\theta \log p_\theta$ 通过抽样分布对参数的敏感性，对可观测量涨落进行重新加权。由于式 (S2) 右侧仅涉及 $\nabla_\theta \log p_\theta$，且 $\log p_\theta$ 中的任何加性常数都会在协方差中消去，因此我们可以始终使用未归一化的密度。下文将式 (S2) 应用于两种抽样分布：单个构型的玻尔兹曼分布，由此得到用于径向分布函数等结构可观测量的构型得分（见 S1.2 节）；以及朗之万轨迹上的概率分布，由此得到用于扩散系数等动力学可观测量的路径得分（见 S1.3 节）。

S1.2 构型得分与结构可观测量

对于静态可观测量 $Q(\mathbf{r})$，例如径向分布函数 (RDF)，其采样分布为玻尔兹曼分布 $p_\theta(\mathbf{r}) \propto \exp[-\beta U_\theta(\mathbf{r})]$，其中 $\beta = 1/(k_B T)$，其得分为

$$\nabla_\theta \log p_\theta(\mathbf{r}) = -\beta\,\nabla_\theta U_\theta(\mathbf{r}) + \text{const}, \tag{S3}$$

其中，与构型无关的常数（对数配分函数的梯度）在式 (S2) 的协方差中消去。因此，静态系综平均的梯度可简化为构型形式

$$\nabla_\theta\langle Q\rangle_\theta = \langle\nabla_\theta Q\rangle_\theta - \beta\left\langle\left(Q-\langle Q\rangle_\theta\right)\nabla_\theta U_\theta\right\rangle_\theta, \tag{S4}$$

这就是 DiffTRe 方法 [6] 所依据的估计量。

RDF $g(r)$ 按每个构型计算为高斯平滑直方图：

$$g(r_i) = \frac{1}{4\pi r_i^2 \rho N}\sum_{a=1}^{N}\sum_{b\neq a}\frac{1}{\sqrt{2\pi\varepsilon^2}}\,\exp\!\Big[-\frac{(r_i-|\mathbf{r}_a-\mathbf{r}_b|)^2}{2\varepsilon^2}\Big], \tag{S5}$$

其中 ρ 是数密度，ε 是平滑宽度（我们使用 $\varepsilon = 0.1\,\text{Å}$）。平滑处理可得到平滑的径向分布函数估计，并减少分箱噪声。传统直方图会引入不连续的分箱计数和高方差的散粒噪声。

S1.3 路径得分与动力学可观测量

扩散系数等动力学可观测量取决于整条轨迹，而非单个构型。因此，我们将轨迹本身作为式（S2）中的随机变量。下面先讨论路径概率及其得分的一般结构，包括初始条件的处理（第 S1.3.1 节），然后针对数值实现中使用的 A-BO 兰之万方案，明确计算其得分（第 S1.3.2 节）。

S1.3.1 路径概率、得分函数与初始条件

我们将粒子 i 的质量记为 M_i，并将多粒子朗之万动力学写作

$$\begin{aligned} \mathrm{d}\mathbf{r}_i &= \mathbf{v}_i\,\mathrm{d}t, \\ \mathrm{d}\mathbf{v}_i &= \frac{\mathbf{F}_i(\mathbf{r})}{M_i}\,\mathrm{d}t - \gamma\mathbf{v}_i\,\mathrm{d}t + \sqrt{\frac{2\gamma\,k_BT}{M_i}}\,\mathrm{d}\mathbf{W}_i, \end{aligned} \tag{S6}$$

其中 $\mathbf{F}_i(\mathbf{r}) = -\nabla_{\mathbf{r}_i}U_\theta(\mathbf{r})$，且 $\mathrm{d}\mathbf{W}_i$ 是维纳过程。

数值积分器通过抽取高斯噪声 $\boldsymbol{\xi}_i^{(n)} \sim \mathcal{N}(\mathbf{0},\ \mathbf{I})$，在一个时间步长 Δt 内将相点 $\mathbf{z}^{(n)} = (\mathbf{r}^{(n)}, \mathbf{v}^{(n)})$ 推进至 $\mathbf{z}^{(n+1)}$。对于此处考虑的积分器，可以根据存储的位置和速度事后推断噪声，并且从 $\boldsymbol{\xi}_i^{(n)}$ 到更新后相点的映射具有与参数无关的雅可比行列式，如 S1.3.2 节中明确所示。因此，在给定初始相点 $\mathbf{z}_0 = (\mathbf{r}_0,\mathbf{v}_0)$ 的条件下，轨迹 Γ 的概率除去一个与参数无关的因子后为

$$p_\theta(\Gamma \mid \mathbf{z}_0) \propto \prod_{i,n} \exp\Big(-\frac{|\boldsymbol{\xi}_i^{(n)}|^2}{2}\Big). \tag{S7}$$

由于式 (S2) 仅涉及 $\nabla_\theta \log p_\theta$，与参数无关的雅可比行列式因子和密度的整体归一化因子会相互抵消，因此无需计算。

完整的路径概率可因式分解为

$$p_\theta(\Gamma) = p_\theta(\Gamma \mid \mathbf{z}_0)\; p_\theta(\mathbf{z}_0). \tag{S8}$$

我们定义路径得分

$$S_\theta[\Gamma] \equiv \nabla_\theta \log p_\theta(\Gamma \mid \mathbf{z}_0) = -\sum_{i,n} \boldsymbol{\xi}_i^{(n)} \cdot \nabla_\theta \boldsymbol{\xi}_i^{(n)}, \tag{S9}$$

其显式形式取决于通过推断噪声所采用的积分器，并在第 S1.3.2.[7–9] 节中推导。因此，该得分记录了在原始力场和受扰动的力场下生成观测路径所需噪声之间累积的偏差。

剩下需要说明的是 $p_\theta(\mathbf{z}_0)$，即如何处理初始条件。这里有两种情况。若初始相点固定不变，则 $p_\theta(\mathbf{z}_0)$ 与参数无关，且 $\nabla_\theta \log p_\theta(\Gamma) = S_\theta[\Gamma]$。若初始条件是从该力场的平衡分布中采样得到的，

$$p_\theta(\mathbf{z}_0) \propto \exp\Big[-\beta\Big(U_\theta(\mathbf{r}_0) + \sum_i \tfrac{1}{2}M_i|\mathbf{v}_{i,0}|^2\Big)\Big], \tag{S10}$$

速度服从与参数无关的麦克斯韦–玻尔兹曼分布，而位置则具有第 S1.2 节中的构型得分，得到

$$\nabla_\theta \log p_\theta(\Gamma) = S_\theta[\Gamma] - \beta\nabla_\theta U_\theta(\mathbf{r}_0) \tag{S11}$$

相差一个加性常数。在本文所展示的示例中，初始构型取自当前力场的平衡分布，实践中会保留这两项贡献：使用总得分（S11）计算扩散梯度，因此每条轨迹都会贡献其累积路径得分 $S_\theta[\Gamma]$，以及其初始构型的构型得分 $-\beta\,\nabla_\theta U_\theta(\mathbf{r}_0)$。

对于不显式依赖参数的动力学可观测量 $O[\Gamma]$，式 (S2) 的第一项为零，其梯度就是该可观测量与总得分 (S11) 的协方差。对于从初始位置 $\mathbf{r}_{k,\,0}$ 出发的 K 条独立轨迹 $\{\Gamma_k\}_{k=1}^{K}$，我们估计

$$\bar{O} = \frac{1}{K}\sum_{k=1}^{K} O[\Gamma_k], \qquad \nabla_\theta\langle O\rangle_\theta \approx \frac{1}{K-1}\sum_{k=1}^{K}\big(O[\Gamma_k]-\bar{O}\big)\,\Big(S_\theta[\Gamma_k] - \beta\,\nabla_\theta U_\theta(\mathbf{r}_{k,0})\Big). \tag{S12}$$

对于平方误差损失 $\mathcal{L} = (\langle O\rangle_\theta - O_{\mathrm{target}})^2$,

$$\begin{aligned}\nabla_\theta\mathcal{L} &= 2(\langle O\rangle_\theta - O_{\mathrm{target}})\,\nabla_\theta\langle O\rangle_\theta \\ &\approx 2(\bar{O} - O_{\mathrm{target}})\,\frac{1}{K-1}\sum_{k=1}^{K}\big(O[\Gamma_k]-\bar{O}\big)\,\Big(S_\theta[\Gamma_k] - \beta\,\nabla_\theta U_\theta(\mathbf{r}_{k,0})\Big).\end{aligned} \tag{S13}$$

S1.3.2 A-BO 朗之万分裂格式

本文的数值实验采用式 (S6) 的 A-BO 朗之万离散化 [10, 11]。对于粒子 i，将状态 $(\mathbf{r}_i^{(n)},\ \mathbf{v}_i^{(n)})$ 推进一步至 $(\mathbf{r}_i^{(n+1)},\ \mathbf{v}_i^{(n+1)})$ 的一步更新为

$$\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(n)} + \mathbf{v}_i^{(n)}\,\Delta t, \tag{S14}$$

$$\mathbf{v}_i^{(n+1)} = c_0\,\mathbf{v}_i^{(n)} + c_{1,i}\,\boldsymbol{\xi}_i^{(n)} + c_{2,i}\,\mathbf{F}_{i,\theta}(\mathbf{r}^{(n+1)}), \tag{S15}$$

with

$$c_0 = e^{-\gamma\Delta t}, \qquad c_{1,i} = \sqrt{\frac{(1-c_0^2)\,k_BT}{M_i}}, \qquad c_{2,i} = \frac{1-c_0}{\gamma M_i}, \tag{S16}$$

以及 $\boldsymbol{\xi}_i^{(n)} \sim \mathcal{N}(\mathbf{0},\ \mathbf{I})$。该格式先利用当前速度执行完整的位置漂移步，然后使用在新位置处计算的恒定力，通过精确的奥恩斯坦–乌伦贝克传播算子更新速度 [12, 13]。

可以根据存储的位置和速度推断高斯噪声，

$$\boldsymbol{\xi}_i^{(n)} = \frac{1}{c_{1,i}}\left[\mathbf{v}_i^{(n+1)} - c_0\,\mathbf{v}_i^{(n)} - c_{2,i}\,\mathbf{F}_{i,\theta}(\mathbf{r}^{(n+1)})\right]. \tag{S17}$$

从噪声 $\boldsymbol{\xi}_i^{(n)}$ 到新速度的映射是线性的，其行列式为 $c_{1,\,i}^{d}$，其中 d 为空间维度；而在给定旧速度的情况下，位置更新是确定性的。因此，将$p_\theta(\Gamma \mid \mathbf{z}_0)$与高斯噪声分布联系起来的雅可比行列式与参数无关，式（S7）成立。将（S17）代入（S7）得到离散的昂萨格–马赫卢普作用量 [14, 15]

$$p_\theta(\Gamma \mid \mathbf{z}_0) \propto \exp\left[-\sum_i \frac{1}{2c_{1,i}^2}\sum_n \Big(\mathbf{v}_i^{(n+1)} - c_0\,\mathbf{v}_i^{(n)} - c_{2,i}\,\mathbf{F}_{i,\theta}(\mathbf{r}^{(n+1)})\Big)^2\right]. \tag{S18}$$

对数求导得到路径得分

$$
\begin{aligned}
S_\theta[\Gamma] &= \nabla_\theta\Big(-\frac{1}{2}\sum_{i,n}\big|\boldsymbol{\xi}_i^{(n)}\big|^2\Big) \\
&= -\sum_{i,n}\boldsymbol{\xi}_i^{(n)}\cdot\nabla_\theta\boldsymbol{\xi}_i^{(n)} \\
&= \sum_{i,n}\frac{c_{2,i}}{c_{1,i}}\,\boldsymbol{\xi}_i^{(n)}\cdot\nabla_\theta\mathbf{F}_{i,\theta}(\mathbf{r}^{(n+1)}) \\
&= \sum_{i,n}\frac{1}{\gamma}\,\sqrt{\frac{1-e^{-\gamma\Delta t}}{(1+e^{-\gamma\Delta t})\,k_BTM_i}}\;\boldsymbol{\xi}_i^{(n)}\cdot\nabla_\theta\mathbf{F}_{i,\theta}(\mathbf{r}^{(n+1)}).
\end{aligned}
\tag{S19}
$$

在小时间步极限下，$(1-e^{-\gamma\Delta t})/(1+e^{-\gamma\Delta t}) \to \gamma\Delta t/2$ 且前因子化简为$\sqrt{\Delta t/(2\gamma M_i k_B T)}$，从而恢复了正文中给出的得分的 Euler–Maruyama 形式。

S1.4 方差

对动力学可观测量求导的主要难点在于，梯度方差会随轨迹长度增长。在直接通过运动方程求导时，导数会沿轨迹递归传播。在混沌多体系统中，力场的微小变化可能会被迅速放大，导致所得轨迹导数随仿真时间爆炸式增长。

路径得分方法避免了传播这些轨迹导数。为考察其方差，我们考虑方程 S19 的小时间步极限形式。来自时间步长 n 的贡献为

$$
s_n = \sum_i\sqrt{\frac{\Delta t}{2\gamma M_i k_{\mathrm{B}}T}}\,\xi_i^{(n)}\cdot\nabla_\theta F_{i,\theta}\Big(r^{(n+1)}\Big), \qquad S_\theta = \sum_{n=0}^{N_t-1}s_n. \tag{S20}
$$

A-BO积分器的排序在这里十分重要。位置 $r^{(n+1)}$以及在该位置计算的力导数，是在施加当前噪声 $\xi^{(n)}$ 以更新速度之前确定的。因此，对于给定的轨迹历史，每个得分增量的符号和方向都由新抽取的零均值高斯噪声决定。正负贡献在平均意义下相互抵消，从而得到 $\langle s_n\rangle = 0$。

同样的论证也适用于不同时间步之间的相关性。对于 $n<m$，第 m 步的力导数可能取决于较早的轨迹，因而也可能取决于较早的噪声，但它与新抽取的噪声 $\xi^{(m)}$ 独立。对这一新噪声取平均会使 s_n 与 s_m 之间的互相关消失。因此，累积得分表现为步长取决于构型的随机游走，而不是迭代放大的轨迹导数。

因此，得分方差等于其各个增量方差之和。对单位高斯噪声取平均，会将噪声投影的平方转化为力导数范数的平方，从而得到

$$
\mathrm{Var}(S_\theta) \simeq \sum_{n=0}^{N_t-1}\sum_i\frac{\Delta t}{2\gamma M_i k_{\mathrm{B}}T}\left\langle\Big|\nabla_\theta F_{i,\theta}(r^{(n+1)})\Big|^2\right\rangle. \tag{S21}
$$

平衡后，平均力梯度范数近似保持平稳。由于 $t = N_t\Delta t$，式 S21 表明

$$
\mathrm{Var}(S_\theta) \propto t. \tag{S22}
$$

因此，路径得分具有类随机游走涨落，其方差随轨迹长度呈线性增长，而非指数增长。式 S11 中的平衡初始条件贡献增加了一个与时间无关的项，不会改变这种长时间标度。

梯度估计量涉及得分与可观测量的乘积， $\delta O_t S_\theta$。我们假设可观测量通过对粒子和时间起点取平均进行评估，并且轨迹长度显著超过相关时间，这在分子动力学模拟中是标准做法。在这种标准的长轨迹近似下，有效独立样本的数量与 t 成正比，而可观测量的方差则随之降低：

$$\mathrm{Var}(\delta O_t) \propto \frac{1}{t}. \tag{S23}$$

将可观测量和得分的涨落近似为联合高斯分布后，它们乘积的方差满足

$$\mathrm{Var}(\delta O_t S_\theta) = \mathrm{Var}(\delta O_t)\mathrm{Var}(S_\theta) + \mathrm{Cov}(\delta O_t, S_\theta)^2 \leq 2\mathrm{Var}(\delta O_t)\mathrm{Var}(S_\theta). \tag{S24}$$

因此，轨迹长度的影响预计会相互抵消。更长的轨迹会增加累积路径得分的方差，但同时也会改善可观测量的统计估计。我们使用 128 条分别平衡的朗之万轨迹，在 $\gamma = 1$皮秒$^{-1}$ 下评估了初始 NEP 水模型中的这种补偿效应（补充图 S1）。扩散标准差随轨迹长度增加而减小，而累积路径得分的标准差则增大。在这一范围内，由此得到的梯度标准差仅表现出轻微变化。

需要注意的是，方差仍依赖于朗之万摩擦系数，其近似按 $O(1/\gamma)$缩放。由于整个模拟过程中 γ 保持不变，这种依赖性不会影响本文考察的轨迹长度缩放关系。

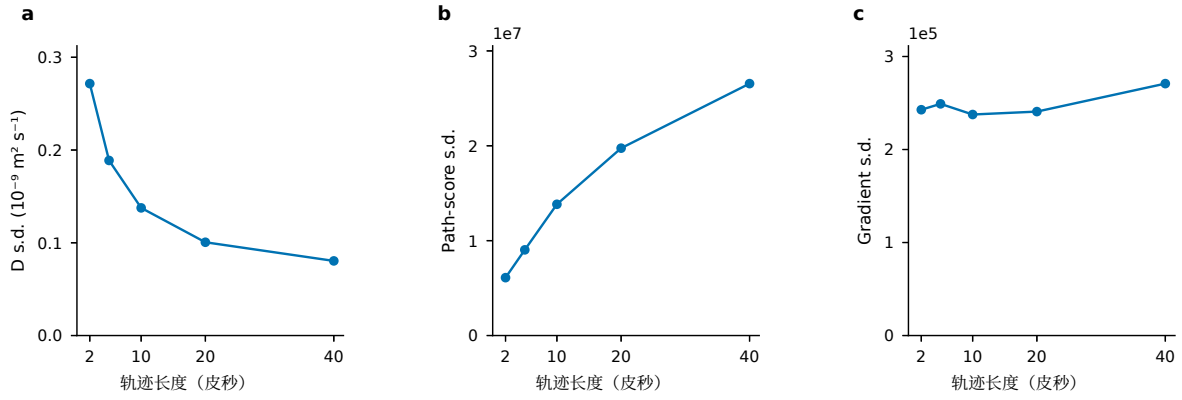


图 S1：水梯度的标准差分解。（a）128 条轨迹中扩散系数的样本标准差。（b）累积动力路径得分的样本协方差迹的平方根。（c）梯度估计的标准差。

S2 验证水势的其他结果。

为验证微调后的水势，我们选取了微调轨迹上迭代次数为 0、200、400 和 680 的四个检查点。此外，我们还在 288.15 和 308.15 开尔文下评估了初始模型（迭代次数为 0）和微调后的模型（迭代次数为 680），以检验温度可迁移性。

对于每个模型和温度，通过按比例缩放分子位置，制备了五个独立的初始构型，使其立方盒长度对应于该温度和大气压下液态水的实验密度 [16], ，同时保持分子内键长和键角不变。所有模拟均使用 512 个水分子（1536 个原子）、 $\Delta t = 1$ fs 的时间步长，并在定容条件下进行 $\gamma = 1$ ps^{-1} 的朗之万动力学模拟。在 288.15 开尔文下，为考虑初始模型中应力自相关衰减较慢的情况，延长了平衡和生产阶段的轨迹。

补充图S2展示了 298.15 开尔文下的关联函数及其提取窗口，这些数据是对每个检查点的五条独立生产阶段轨迹取平均得到的。补充表 S2 列出了沿优化迭代过程四个检查点对应的三个输运系数，补充表S3则比较了初始模型和微调后模型在 288.15、298.15 和 308.15 开尔文下的表现。每个数据项均为五条独立生产阶段轨迹的均值和样本标准差，模拟时长如表 S1 所示。

为量化水动力学的温度依赖性，我们使用阿伦尼乌斯关系分析了这三个动力学可观测量

$$Q(T) = Q_0 \exp\left(\pm \frac{E_a}{RT}\right), \tag{S25}$$

其中， Q 表示自扩散系数 D、剪切黏度 η或分子内 H–H 向量的二阶旋转弛豫时间 τ_2 ； R 为气体常数， E_a 为活化能。负号适用于 D，而正号适用于 η 和 τ_2。

我们使用各温度下的实验密度，对 288.15、298.15 和 308.15 K 时 $\ln(Q/Q_{\mathrm{ref}})$ 与 $1000/T$ 进行了普通最小二乘拟合。所得表观活化能见图S3并汇总于表 S4，用于表征这 20 K 温度区间内的温度依赖性。实验活化能采用与表S3所列参考数据相同的拟合方法获得。微调使三个动力学可观测量的表观活化能均显著接近其实验值，表明模型在训练条件之外的温度依赖性也得到改善。

表 S1：水验证协议。生产阶段时长为每条轨迹的时长；每个可用参数组均使用五条轨迹。

T（开尔文）	密度（$\mathrm{g\,cm^{-3}}$）	平衡	生产阶段
288.15	0.99910	400 皮秒	10 纳秒
298.15	0.99704	200 皮秒	5 ns
308.15	0.99403	200 皮秒	5 ns

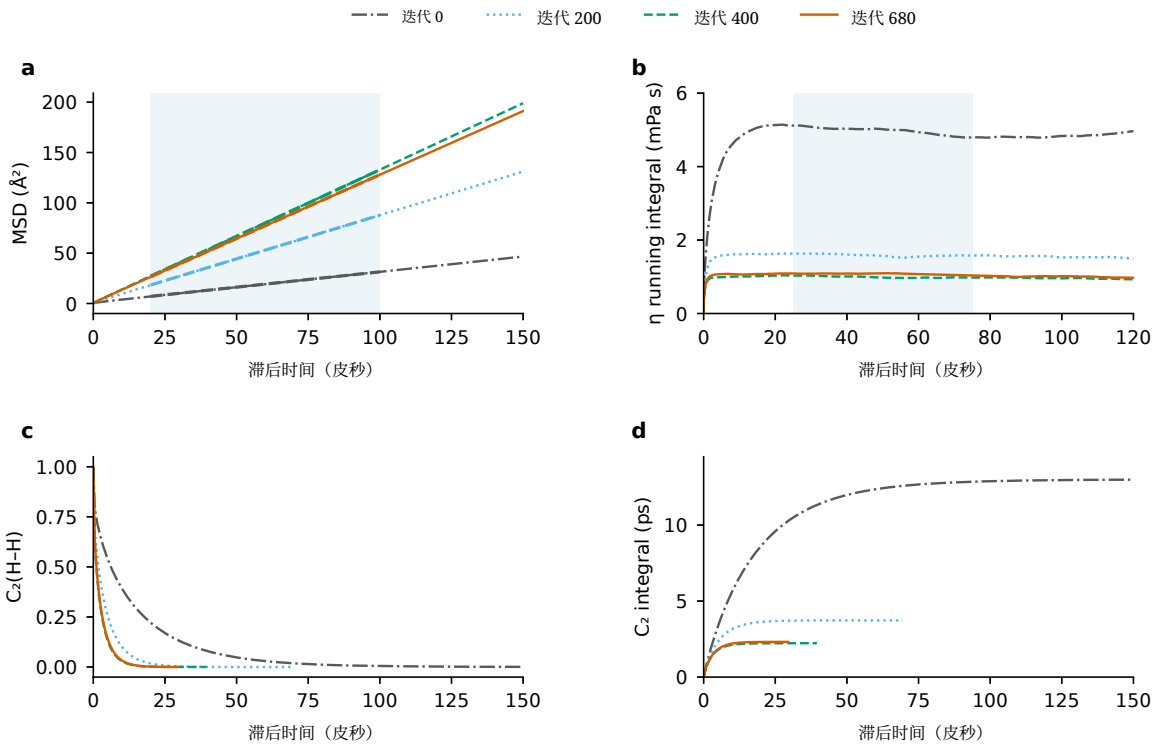


图 S2：水的输运系数的提取。（a）平均均方位移，阴影部分表示 20–100 皮秒的拟合窗口，以及该窗口内的线性拟合。（b）平均黏度运行积分，阴影部分表示 25–75 皮秒的平台窗口。（c）平均 H–H 二阶旋转相关函数，以及（d）在各集合采样窗口内的运行梯形积分。曲线为每个参数组中五条独立轨迹的平均值。

表 S2：输运系数（温度为 298.15 K. $T = 298.15\,\mathrm{K}$, $\rho = 0.99704 \mathrm{g\ cm^{-3}}$。仅将 D 和氧-氧径向分布函数纳入损失函数。数值为五条独立轨迹的均值 ± 标准差。

迭代	D(10^{-9} 分^2 秒$^{-1}$)	η (mPa s)	τ_2 (ps)
0	0.507 ± 0.005	4.98 ± 0.28	12.99 ± 0.21
200	1.446 ± 0.007	1.59 ± 0.08	3.73 ± 0.05
400	2.200 ± 0.025	0.99 ± 0.12	2.23 ± 0.02
680	2.114 ± 0.018	1.08 ± 0.07	2.32 ± 0.01
实验	2.299	0.890	2.36

表S3：跨温度可迁移性。在288.15、298.15和308.15开尔文下的初始值和微调值，各温度下均采用该温度的实验密度。实验参考值： D [17]； η [18]；298.15开尔文下 [19, 20]。

物理量	Set	288.15开尔文	298.15开尔文	308.15开尔文
D(10^{-9} 分^2 秒$^{-1}$)	初始	0.275 ± 0.005	0.507 ± 0.005	0.832 ± 0.011
	微调后	1.602 ± 0.010	2.114 ± 0.018	2.712 ± 0.005
	实验	1.766	2.299	2.895
η (毫帕·秒)	初始	9.897 ± 0.482	4.977 ± 0.283	2.901 ± 0.185
	微调后	1.346 ± 0.079	1.077 ± 0.071	0.861 ± 0.077
	实验	1.138	0.890	0.720
τ_2 (皮秒)	初始	26.51 ± 0.75	12.990 ± 0.214	7.396 ± 0.057
	微调后	3.208 ± 0.014	2.318 ± 0.011	1.763 ± 0.004
	实验	2.93*	2.36	1.81*

*288.15 和 308.15 开尔文时的实验 τ_2 数值，是从 Krynicki [21], 和报告的分子内质子偶极相关时间 τ_d 的拟合温度依赖性中数字化提取的，因此在正文图5c中以空心符号绘出。在同一拟合曲线上读取 298.15 开尔文时的数值为 2.28 皮秒，与 Jonas 等人 [19, 20]的测量所得 2.36 皮秒数值一致。将质子自旋-晶格弛豫率中的分子间贡献分离出来，会给核磁共振对 τ_2 [22] 的估计引入额外的、依赖于模型的不确定性。在极限窄化极限下，从质子弛豫推断出的分子内相关时间对应于 $C_2(t)$ 的时间积分，与此处使用的 τ_2 定义相符；而拉伸指数（Kohlrausch–Williams–Watts）拟合参数和$1/e$ 衰减时间则表征不同的弛豫测量。

表 S4：288.15–308.15 K 范围内水动力学的表观活化能。初始模型、微调后模型和实验的数值，是通过对 288.15、298.15 和 308.15 K 下的数据进行三点阿伦尼乌斯拟合得到的（图S3）。实验数值使用表S3中列出的参考数据进行拟合。所有活化能均以 kJ mol^{-1} 为单位报告。

物理量	初始	微调后	实验
D	40.86	19.44	18.25
η	45.35	16.47	16.94
τ_2 (HH)	47.17	22.11	17.75

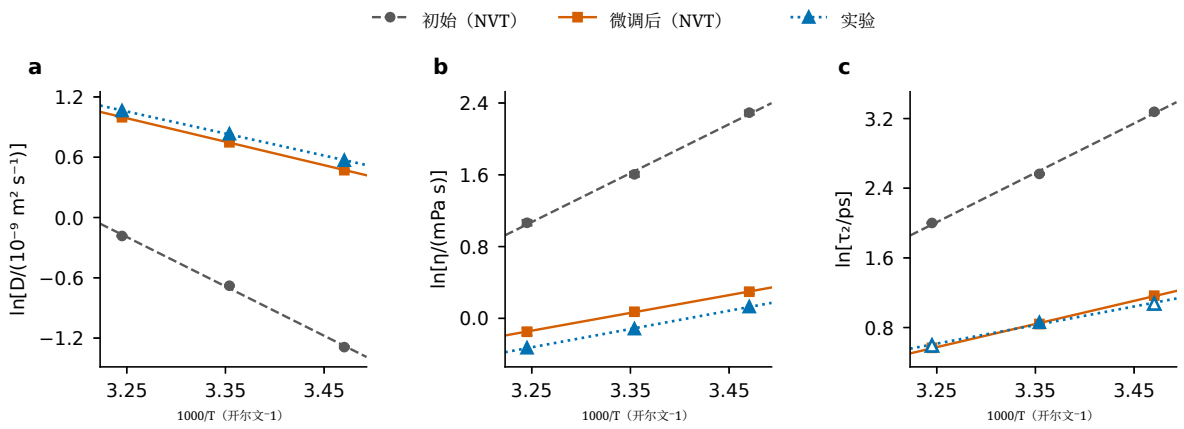


图 S3：288.15–308.15 K 范围内水动力学的阿伦尼乌斯分析。(a) 扩散系数 D、**(b)** 剪切黏度 η 和**(c)** 分子内 H–H 向量的二阶旋转弛豫时间 τ_2，对应初始模型、微调后模型和实验测量。结果以 $\ln(Q/Q_{\mathrm{ref}})$ 对 $1000/T$作图，其中 Q 表示 D、 η或 τ_2。仿真符号表示五条独立正则系综轨迹的均值。相应的标准误差太小，因此图中无法显示误差棒。线条表示线性阿伦尼乌斯拟合，据此确定表 S4 中报告的表观活化能。

表 S5：模拟与优化参数。 t 为用于评估 D的轨迹长度， Δt为时间步长， γ 为朗之万摩擦， T 为恒温器温度， ρ 为密度， N_{atoms} 为模拟的原子数， K 为用于估算梯度的独立轨迹数， α_D和 α_{struct}分别为动力学损失项和结构损失项的学习率。

参数	熔融铜	水
t (皮秒)	1	2
Δt (皮秒)	10^{-3}	10^{-3}
γ (皮秒$^{-1}$)	1	1
T (开尔文)	1500	300
ρ (g cm^{-3})	7.94	0.994
D_{target} (10^{-9} 分^{2}秒$^{-1}$)	4.3†	2.3‡
N	256	1536
K原子	64	64
α_D	10^{-4}	10^{-3}
$\alpha_{\text{结构}}$	10^{-6}	10^{-2}

†参考文献 [23–26]; ‡参考文献 [17, 27, 28]。

参考文献

[1] Glynn, P. W. 随机系统的似然比梯度估计。《ACM通讯》33，75–84（1990）。[2] Williams, R. J. 联结主义强化学习的简单统计梯度跟随算法。机器学习 8，229–256（1992）。[3] Kleijnen, J. P. & Rubinstein, R. Y. 使用得分函数法对计算机模拟模型进行优化与敏感性分析。《欧洲运筹学杂志》88，413–427（1996）。[4] Fu, M. C. 梯度估计。载于 Henderson, S. G. & Nelson, B. L.（编）《模拟》，《运筹学与管理科学手册》第13卷，第19章，575–616页（Elsevier，2006）。[5] Mohamed, S., Rosca, M., Figurnov, M. & Mnih, A. 机器学习中的蒙特卡洛梯度估计。《机器学习研究杂志》21，1–62（2020）。[6] Thaler, S. & Zavadlav, J. 通过可微轨迹重加权从实验数据中学习神经网络势。Nature Communications 12，6884（2021）。[7] Donati, L., Hartmann, C. & Keller, B. G. 路径系综和马尔可夫状态模型的 Girsanov 重加权。The Journal of Chemical Physics 146，244112（2017）。[8] Kieninger, S., Ghysbrecht, S. & Keller, B. G. 欠阻尼朗之万动力学模拟的 Girsanov 重加权。理论。arXiv 预印本 arXiv:2303.14696（2023）。[9] Bolhuis, P. G., Brotzakis, Z. F. & Keller, B. G. 通过路径重加权施加动力学约束来优化分子势模型。The Journal of Chemical Physics 159，074102（2023）。[10] Leimkuhler, B. & Matthews, C. 用于分子采样的随机数值方法的合理构造。《应用数学研究快报》2013，34–56（2013）。[11] Leimkuhler, B. & Matthews, C. 通过朗之万动力学实现稳健高效的构型分子采样。The Journal of Chemical Physics 138（2013）。[12] Paterlini, M. G. & Ferguson, D. M. 使用带速度Verlet积分的朗之万方程进行恒温模拟。Chemical Physics 236，243–252（1998）。[13] Melchionna, S. 朗之万动力学拟辛传播子的设计。The Journal of Chemical Physics 127，044108（2007）。[14] Onsager, L. & Machlup, S. 涨落与不可逆过程。《物理评论》91，1505–1512（1953）。[15] Machlup, S. & Onsager, L. 涨落与不可逆过程。II. 具有动能的系统。《物理评论》91，1512–1515（1953）。[16] Kell, G. S. 液态水的密度、热膨胀率和压缩性：大气压和饱和条件下的相关性与表格，并在1968年温标上进行了审校和表达。《Journal of Chemical & Engineering Data》20，97–105（1975）。

[17] Holz, M., Heil, S. R. & Sacco, A. 水和六种选定分子液体的温度依赖性自扩散系数，用于准确的^{1}H NMR 脉冲场梯度测量校准。*Physical Chemistry Chemical Physics* **2**，4740–4742（2000）。[18] Korson, L., Drost-Hansen, W. & Millero, F. J. 不同温度下水的黏度。*TheJournal of Physical Chemistry* **73**，34–39（1969）。[19] Jonas, J., DeFries, T. & Wilbur, D. J. 压缩水中的分子运动。*TheJournal of Chemical Physics* **65**，582–588（1976）。[20] Kumar, R. & Skinner, J. L. 具有显式三分子相互作用的水模拟模型。*The Journal of Physical Chemistry B* **112**，8311–8318（2008）。[21] Krynicki, K. 纯水在 0°c至 100°c之间的质子自旋-晶格弛豫。*Physica* **32**，167–178（1966）。[22] Lankhorst, D., Schriever, J. & Leyte, J. C. 通过质子核磁共振弛豫测定水在 $H_2^{17}O$ 中的旋转相关时间及一些相关结果。***Berichte der Bunsengesellschaft für physikalische Chemie*** **86**，215–221（1982）。[23] Meyer, A. 准弹性中子散射揭示的液态铜中的自扩散。*PhysicalReview B* **81**，012102（2010）。[24] Meyer, A. 使用准弹性中子散射测量液态金属中的自扩散系数。载于***EPJ*** 会议网络，第83卷，01002（EDP Sciences，2015）。[25] Brillo, J.《多组分液态合金的热物理性质》（De Gruyter Oldenbourg，2016）。[26] Hoyt, J., Sadigh, B., Asta, M. & Foiles, S. 基于原子尺度计算推导的铜–镍体系动力学相场参数。《材料学报》**47**，3181–3187（1999）。URL https://www.sciencedirect.com/science/article/pii/S1359645499001895。[27] Mills, R. 1–45° 范围内正常水和重水的自扩散。*The Journal of PhysicalChemistry* **77**，685–688（1973）。[28] Krynicki, K., Green, C. D. & Sawyer, D. W. 水中自扩散对压力和温度的依赖性。《化学会法拉第讨论》**66**，199–208（1978）。