arXiv:2303.04791v2 [cs.LG] 6 Jun 2023

Ewa基于 ld 的分子图长程消息传递

Arthur Kosmala [1 2 3] **Johannes Gasteiger** [4 2] **Nicholas Gao** [2] **Stephan G¨unnemann** [2]

摘要

近年来，从分子数据中学习势能面的神经网络架构取得了快速进展。这一成功的关键推动因素是消息传递神经网络（MPNN）范式。其随系统规模扩展时的良好效率，部分依赖于对消息施加空间距离限制。虽然这种对局部性的关注是一种有用的归纳偏置，但也阻碍了模型学习静电和范德华力等长程相互作用。为解决这一缺点，我们提出了 Ewald 消息传递：一种非局部的傅里叶空间方案，通过频率截断而非距离截断来限制相互作用，并在理论上以埃瓦尔德求和方法为依据。它可以作为现有 MPNN 架构的增强模块，因为计算开销低，且不依赖具体的架构细节。我们在包含多样化周期性结构（OC20）和非周期性结构（OE62）的两个数据集上，使用四种基线模型测试了该方法。我们观察到，该方法在所有模型和数据集上都稳定地降低了能量平均绝对误差，OC20 上平均降低 10%，OE62 上平均降低 16%。分析显示，这些改进对真实能量中长程贡献较大的结构影响尤为显著。

1. 引言

图神经网络（GNNs）在利用量子化学参考数据学习分子性质方面展现出巨大潜力，能将推理时间缩短数个数量级（Gilmer 等人，2017年），同时通常能够保持与其真实值方法相当的准确性。然而，

[1]马克斯·普朗克光科学研究所与埃尔朗根-纽伦堡弗里德里希-亚历山大大学，德国 [2]慕尼黑工业大学计算机科学系与慕尼黑数据科学研究所，德国 [3]慕尼黑路德维希-马克西米利安大学物理系，德国 [4]Google Research。通讯作者：Arthur Kosmala < arthur.kosmala@tum.de>。

《 40^{th} 国际机器学习会议，美国夏威夷州檀香山。PMLR 202，2023。版权所有 2023，归作者所有。

尽管近期在理论上理解通用模型方面取得了进展（Gasteiger 等人，2021年；Dym& Maron，2021年），计算限制仍妨碍图神经网络学习原子之间的长程物理相互作用：标准的分子图概念（若未使用人工构造的键连接信息）将原子视为顶点，并在它们的三维距离小于某个截断超参数时，以边连接这些原子（Sch¨utt 等人，2018年）。如果较远处的原子相互作用会影响目标，这就可能造成问题：这些相互作用在图中同样是长程的，可能关联相距许多条边的顶点。大量实证和理论证据表明，在这种情况下图神经网络的性能有限，例如可能是由于过度压缩（Alon & Yahav，2021）。

图神经网络现状与分子动力学中的实际情况形成鲜明对比，在分子动力学中，长程相互作用是基本要素。该领域（尤其是本研究）的核心任务之一，是根据分子中原子的位置 $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$预测其势能。对此映射的手工参数化形式称为经验力场，它们区分成键能项和非成键能项，$E = E^{\mathrm{b}} + E^{\mathrm{nb}}$ (Leach，2001）。成键项刻画共价键的短程特性及其复杂但高度局部化的相互作用。然而，非成键项则捕捉随距离呈重尾（通常为幂律）衰减的相互作用。静电相互作用就是此类项的一个例子，

$$E^{\mathrm{es}} = \sum_{i<j} q_i q_j \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{-1}, \tag{1}$$

其中部分电荷 q_i 量化了原子周围的净电荷 i。它们之间的吸引或排斥会随距离的增加而反比衰减，而能量则是所有原子对相互作用之和。由于这种无尺度衰减，简单地使用距离截断来截断此类幂律，即使截断距离非常大，也可能在可测的热化学预测中引入严重伪影（Wells&Chaffee，2015年）。因此，必须将所有原子对的相互作用纳入求和。埃瓦尔德求和技术能够高效地实现这一点。它将实际的物理相互作用分解为短程部分和长程部分。短程部分可以通过距离截断直接求和；长程部分虽然随距离缓慢衰减，但其傅里叶变换衰减很快，因此可以通过设置频率截断，以傅里叶空间求和的方式高效计算（参见图 1）。

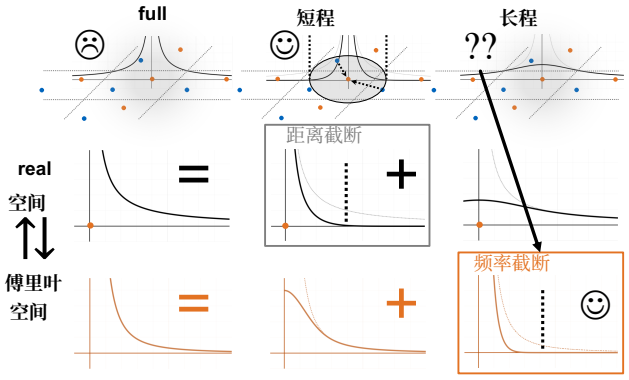


图 1。将一种相互作用分解为短程部分和长程部分；在此示例中，该相互作用呈 $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{-1}$ 衰减特性。

本文构建了埃瓦尔德求和与当前图神经网络模型之间的桥梁。我们提出的*Ewald* 消息传递（Ewald MP）是一种通用框架，其作用是补充现有的图神经网络层，这类似于在埃瓦尔德求和中，经过频率截断的长程部分补充经过距离截断的短程部分。该方法与架构无关且计算高效；我们通过在现有图神经网络模型之上进行修改并加以测试，展示了这一点。在 OC20 数据集的周期性结构和 OE62 数据集的非周期性结构上开展的实验表明，该方法能够稳定改进能量平均绝对误差，在 OC20 和 OE62 上的平均改进幅度分别为 10% 和 16%。此外，我们研究了这些改进与运行时间及参数数量之间的权衡，发现任何经过测试的基线设置都无法稳定复现这一效果。通过进一步考察 OE62 的结果，我们观察到，Ewald 消息传递能够复现 DFT-D3 的效果；DFT-D3 是一种手工设计的长程校正项。这支持了我们的主张，即该方法特别针对长程相互作用的学习。总而言之，我们的贡献包括：

- 全新的**Ewald 消息传递**框架，可轻松集成到现有图神经网络模型中。
- 关于其**成本—性能** 权衡的见解，并将其与更广泛的图神经网络设计格局进行比较。
- 一项研究强调了 Ewald 消息传递作为**长程校正方案**的实用性。

2. 背景

消息传递神经网络。 MPNN 由 Gilmer 等人 (2017年) 提出，作为一种涵盖多种分子图神经网络架构的概念框架。分子表示为由原子构成的图 i，原子作为顶点。若原子相邻，$j \in \mathcal{N}(i)$，则 $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| < c_x$，其中 c_x 是距离截断。MPNN 首先仅根据局部原子属性，将原子转换为嵌入 $h_i^0 \in \mathbb{R}^F$。接下来，它通过迭代方式对复杂的原子相互作用进行建模

在消息传递和更新 步骤中更新嵌入，

$$M_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} f_{\text{int}}\left(h_i^{(l)}, h_j^{(l)}, e_{(ij)}\right), \tag{2}$$

$$h_i^{(l+1)} = f_{\text{upd}}\left(h_i^{(l)}, M_i^{(l+1)}\right), \tag{3}$$

其中，消息求和 $M_i^{(l+1)}$ 从邻域 $\mathcal{N}(i)$ 中收集信息，并使用一个函数 f_{int}，该函数取决于当前嵌入和边特征 $e_{(ij)}$（例如，成对距离）。接下来，一个可学习函数 f_{update} 处理消息和。随后，读出步骤从嵌入中生成输出（例如，原子级能量贡献）。我们围绕一种特定类型的消息步骤——连续滤波卷积，构建 Ewald 消息传递的核心框架（Sch¨等人，2018年）：

$$M_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} h_j^{(l)} \cdot \Phi^{(l)}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \tag{4}$$

其中，$\cdot$ 在特征维度上逐分量作用，而$\Phi^{(l)}(\|\cdot\|)$：$\mathbb{R}^3 \longrightarrow \mathbb{R}^F$ 是由 F 个学习得到的径向滤波器组成的堆栈。在每个通道 $1 \le d \le F$中，$\Phi_d^{(l)}(\|\cdot\|)$学习关注原子周围模糊的距离范围，从而能够针对连续变化的分子几何结构执行回归任务。

周期性系统。 对于一般相互作用，距离截断并不充分；这一点在材料中尤为明显，因为原子间距离没有上界。

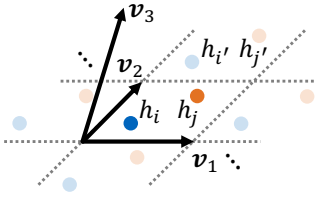


图 2。周期性边界条件的示意图。

周期性边界条件（PBC）可用于将无限系统近似为由有限原子集合重复排列而成的系统（参见图 2），在由晶格矢量（$\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_3$）$\in \mathbb{R}^{3\times3}$张成的超胞[1]中铺满空间。设$\mathcal{S} = \{(h_1, \boldsymbol{x}_1) \text{ ..... } (h_N, \boldsymbol{x}_N)\}$ 表示所选一个超胞中的原子嵌入和位置，并设 $\boldsymbol{t} \in \Lambda$ 为超晶胞晶格 Λ 上的任意整数位移：$= \{\lambda_1\boldsymbol{v}_1 + \lambda_2\boldsymbol{v}_2 + \lambda_3\boldsymbol{v}_3 \mid \boldsymbol{\lambda} \in \mathbb{Z}^3\}$。位于 $\boldsymbol{x}_{i'} := \boldsymbol{x}_i + \boldsymbol{t}$ 的原子 i' 称为 $i \in \mathcal{S}$ 的周期图像，周期性边界条件规定 $h_{i'} \equiv h_i$。

在周期性边界条件下，通用的消息求和会变成无穷级数，因为每个周期图像都可能贡献一条消息。以往针对材料的 MPNN 研究通过采用距离截断缓解了这一问题，同时使用多重图表示，避免在不同顶点上存储冗余 $h_{i'}$（Xie&Grossman，2018年；Chanussot 等人，2021年；Cheng 等人，2021年；Xie 等人，2021年）。

[1]超胞不应与晶体材料的晶胞混淆。非平衡结构中，晶胞平移对称性会被破坏，但仍会施加周期性边界条件（PBC），典型的超胞包含多个晶胞。

3. 埃瓦尔德消息传递

与静电学的类比。埃瓦尔德求和能够应用于 MPNN 的原因，在于静电相互作用与连续滤波器卷积之间存在形式上的对应关系。以往研究指出，这种类比可作为一种有用的归纳偏置（Anderson 等人，2019年）。如果我们将静电能的方程（1） E^{es} 改用一种新记法表示，以呼应方程（4）中的连续滤波器卷积，这种联系就会变得清晰：

$$E^{\text{es}} = \frac{1}{2}\sum_{i=1}^{N} q_i \cdot V_i^{\text{es}}(\boldsymbol{x}_i), \tag{5}$$

$$V_i^{\text{es}}(\boldsymbol{x}_i) = \sum_{j\neq i} q_j \cdot \Phi^{\text{es}}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \quad \Phi^{\text{es}}(r) = \frac{1}{r}, \tag{6}$$

其中 N 是原子总数（$N:=\infty$ 在材料中），我们引入了电势 $\{V_i^{\text{es}} \mid i=1、\ldots\ldots、N\}$ 以及相互作用核 Φ^{es}。将 V_{esi} 的表达式与式（4）中的消息求和进行比较：令截断 $c_x \to \infty$（使得 $\mathcal{N}(i) = \{1,\ldots,N\}\backslash\{i\}$），并代入

$$q_j \leftrightarrow h_j^{(l)}, \quad \Phi^{\text{es}} \leftrightarrow \Phi^{(l)}, \quad V_i^{\text{es}}(\boldsymbol{x}_i) \longleftrightarrow M_i^{(l+1)} \tag{7}$$

将 F维嵌入向量作为原子描述符，取代标量部分电荷，并用一组 $1/r$ 学习得到的卷积滤波器取代简单且基于领域知识的核函数。这些滤波器在形式上充当相互作用核。任何连续滤波器消息求和 M_i（从现在起省略层上标（l））都可以自然地识别为势函数 V_i（$\boldsymbol{x}_i$）在源原子位置处的取值 $\boldsymbol{x}_i$。仅经过一个消息传递步骤后对原子输出进行聚合，会使模型归纳偏向于具有与 E^{es} 类似形式的二体相互作用能。原则上，多个消息与更新步骤能够建模 n体项（Batatia 等人，2022年）。

概念概要。 现在，我们概述埃瓦尔德求和，并基于上述类比说明它与我们的学习方法——埃瓦尔德消息传递——之间的联系。

埃瓦尔德求和 在给定周期性边界条件晶格、通用相互作用核 Φ: $\mathbb{R}^+ \to \mathbb{R}^F$ 以及通用原子 “ 电荷 ”$h_i \in \mathbb{R}^F$ 的情况下，计算势（等价地说，消息求和） $V_i(\boldsymbol{x}_i) = M_i$ 。如果 Φ 是长程的，我们就采用埃瓦尔德求和。

在这种情况下，对于实际尺寸的 c_x 值，朴素的距离截断近似 $M_i \approx \sum_{j\neq i,\ \|\boldsymbol{x}_i-\boldsymbol{x}_j\|<c_x} h_j \cdot \Phi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|)$ 会变得不准确。 Φ 随距离缓慢衰减，这意味着即使 c_x 很大，其尾部仍有相当一部分位于 c_x 之外。这些长程信息随后会因截断而丢失。埃瓦尔德求和预设了已知的分解 $\Phi(r) = \Phi^{\text{sr}}(r) + \Phi^{\text{lr}}(r)$，将其分为短程 核函数和长程核函数，其中

- Φ_{sr} 随距离快速衰减，
- $\Phi_{\text{lr}}(0)$ 定义良好，而 $\hat{}\Phi_{\text{lr}}(\|\cdot\|)$（ $\Phi_{\text{lr}}(\|\cdot\|)$ 的三维傅里叶变换）随频率快速衰减。

根据构造，对距离截断的短程核 $\Phi^{\text{sr}}_{[0,c_x]} := \Phi^{\text{sr}} \cdot \mathbf{1}_{[0,c_x]}$（其中 $\mathbf{1}_{[0,c_x]}$ 是 $[0,c_x]$上的示性函数）进行原子求和时，随着c_x增大，求和会迅速收敛，因此可以高效地应用距离截断。但对于 Φ^{lr}，也就是 Φ剩余的肥尾部分，情况并非如此。埃瓦尔德求和通过将关于 Φ^{lr} 的长程求和改写为对傅里叶变换 $\hat{}\Phi^{\text{lr}}$ 的求和来处理这一问题。将 $\hat{}\Phi^{\text{lr}}$ 替换为 $\hat{}\,\Phi^{\text{lr}}_{[0,c_k]} := \hat{}\Phi^{\text{lr}} \cdot \mathbf{1}_{[0,c_k]}$，即采用频率截断 c_k进行截断，便可高效计算长程求和。将短程与长程求和相加，即可还原对完整核 Φ的求和。图 1 概述了这一方法。Nijboer 与 De Wette（1957）为所有幂律核提供了满足上述性质的便捷分解。

Ewald 消息传递本质上颠倒了埃瓦尔德求和背后的思路：我们不再从已知的物理核 Φ 出发，寻找具有上述性质的分解，而是尝试参数化一个非短程滤波器 Φ 。不过，我们要求它能够通过对由截断支持的短程核和长程核进行埃瓦尔德求和来拟合。这些核彼此独立地学习，$\hat{}\,\Phi^{\text{lr}}_{[0,c_k]}$ 则直接在傅里叶空间中学习。

本节将关键的长程求和改写为对傅里叶频率求和，从而得到一种可以进行频率截断的形式。这为后续深入讨论 Ewald 消息传递及其在现有图神经网络模型中的实现做好了准备。

频率截断。
从这里开始，我们假设长程求和遍历所有原子，包括目标原子i，因此将其记为 $M^{\text{lr}}(\boldsymbol{x}_i)$（不带下标-$i$，但强调其对源原子位置 $\boldsymbol{x}_i$的依赖）。请注意，这一微小的重新定义使映射$\boldsymbol{x} \mapsto M^{\text{lr}}(\boldsymbol{x})$ 成为周期函数。我们检查



图 4。傅里叶模态 $\exp(i\boldsymbol{k}^T\boldsymbol{x})$ 在较小 $\boldsymbol{k} \in \Lambda'$ 下的等相位面示意图。

通过将其表述为对一个超胞中的原子 $\mathcal{S}$ 和超晶胞晶格 Λ 的双重求和：

$$\begin{aligned} M^{\text{lr}}(\boldsymbol{x}) &= \sum_{\tilde{\boldsymbol{t}}\in\Lambda}\sum_{j\in\mathcal{S}} h_j \cdot \Phi^{\text{lr}}(\|\boldsymbol{x} - (\boldsymbol{x}_j + \tilde{\boldsymbol{t}})\|) \\ &= M^{\text{lr}}(\boldsymbol{x} + \tilde{\boldsymbol{t}}), \qquad \forall \tilde{\boldsymbol{t}} \in \Lambda. \end{aligned} \tag{8}$$

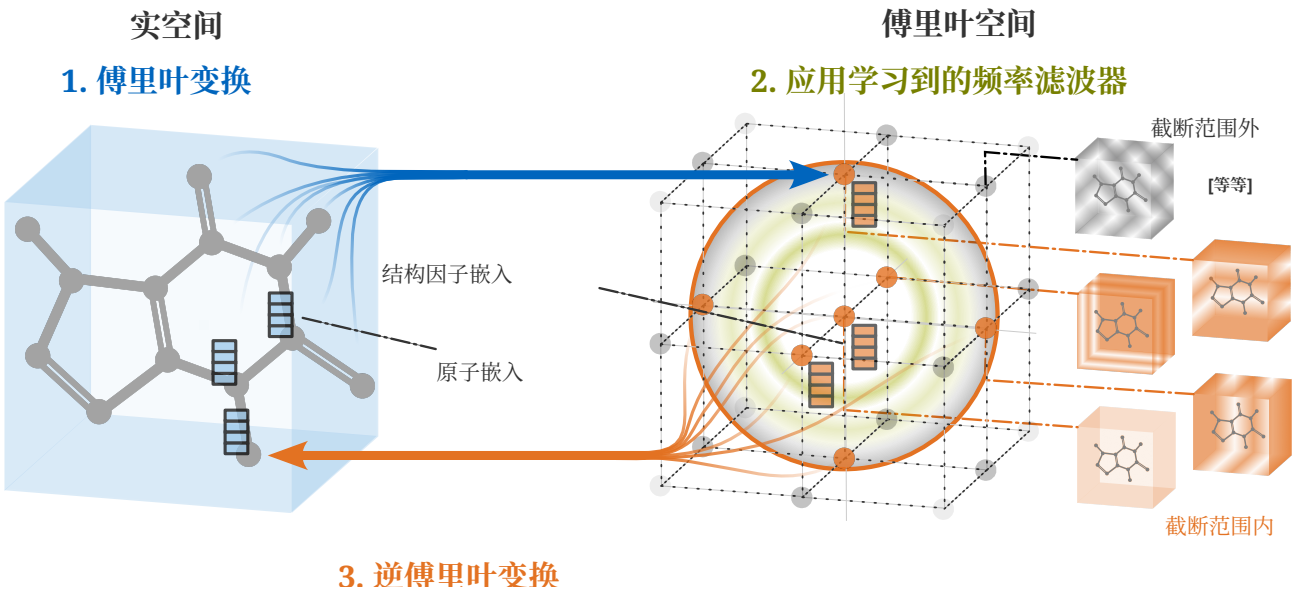


图3. 长程消息传递的三部分结构：首先，对原子 $\sum_{j\in\mathcal{S}}\exp(-i\boldsymbol{k}\boldsymbol{x}_j)[\cdot]$ 求傅里叶和，得到结构因子嵌入 $\{s_{\boldsymbol{k}}\}$ 并将其存储。接着，将这些嵌入与学习到的滤波值按 $\boldsymbol{k}$ 逐项相乘。最后，通过逆傅里叶和 $\sum_{\boldsymbol{k}}\exp(i\boldsymbol{k}\boldsymbol{x}_i)[\cdot]$ 将消息散射回每个原子 i。请注意，该方案具有非局部的全对一特性。如果明确指定了周期性边界条件（PBC），傅里叶域由 $\{\boldsymbol{k}\in\Lambda' \mid \|\boldsymbol{k}\| < c_k\}$ 给出，频率截断为 c_k。如果没有指定，则需要对其进行适当离散化。

因此，M^{lr} 可展开为傅里叶级数（为简洁起见，我们使用复指数作为傅里叶模态）：

$$M^{\text{lr}}(\boldsymbol{x}_i) = \sum_{\boldsymbol{k}\in\Lambda'} \hat{M}^{\text{lr}}_{\boldsymbol{k}} \exp\left(i\boldsymbol{k}^T\boldsymbol{x}_i\right) \tag{9}$$

其傅里叶系数 $\hat{} M^{\text{lr}}_{\boldsymbol{k}} \in \mathbb{C}^F$（将在下文确定）。正如一维傅里叶级数的频率必须是基频的整数倍（即构成一维晶格）一样，三维空间频率 $\boldsymbol{k} \in \Lambda' \subset \mathbb{R}^3$ 必须是三个空间基频（$\boldsymbol{w}_1$、$\boldsymbol{w}_2$、$\boldsymbol{w}_3$）的整数线性组合，$\in \mathbb{R}^{3\times3}$ 这些基频张成倒易晶格Λ'：
$= \{\lambda'_1\boldsymbol{w}_1 + \lambda'_2\boldsymbol{w}_2 + \lambda'_3\boldsymbol{w}_3 \mid \boldsymbol{\lambda'} \in \mathbb{Z}^3\}$。每个基频对应沿一个超胞方向的单周期相移，并在其他方向保持相位不变，即$\boldsymbol{w}_i^T\boldsymbol{v}_j = 1$ 若 $i = j$，否则为 0（参见图 4）。因此，超胞晶格 Λ 确定了倒易晶格 Λ'，且关系式 $\boldsymbol{w}_1 = 2\pi(\boldsymbol{v}_2\times\boldsymbol{v}_3)\Omega^{-1}$、$\boldsymbol{w}_2 = 2\pi(\boldsymbol{v}_3\times\boldsymbol{v}_1)\Omega^{-1}$、$\boldsymbol{w}_3 = 2\pi(\boldsymbol{v}_1\times\boldsymbol{v}_2)\Omega^{-1}$（其中叉积为 $\times$，超胞体积 Ω 为 $= \boldsymbol{v}_1^T(\boldsymbol{v}_2\times\boldsymbol{v}_3)$）很容易验证。

作为引出频率截断的最后一个要素，我们将傅里叶系数 $\hat{} M^{\text{lr}}_{\boldsymbol{k}}$ 与三维傅里叶变换 $\hat{}\Phi^{\text{lr}}(\|\cdot\|)$：$\mathbb{R}^3 \to \mathbb{R}^F$ 联系起来，其中后者对应 $\Phi^{\text{lr}}(\|\cdot\|)$——该核函数不像 $M^{\text{lr}}(\cdot)$ 那样具有周期性，这意味着其频谱是关于 $\boldsymbol{k}\in\mathbb{R}^3$ 的连续径向函数，而非离散的晶格系数集合。在附录 A.5 中，我们推导出恒等式 $\hat{} M^{\text{lr}}_{\boldsymbol{k}} = {}^1\Omega\ \sum_{j\in\mathcal{S}} h_j\exp(-i\boldsymbol{k}^T\boldsymbol{x}_j)\ \hat{}\Phi^{\text{lr}}(\|\boldsymbol{k}\|)$。本质上，式（8）通过在 Λ 上加上 $\sum_{j\in\mathcal{S}} h_j\cdot\Phi^{\text{lr}}(\|\cdot - \boldsymbol{x}_j\|)$ 的所有晶格平移，使其周期化；在傅里叶空间中，这种周期化将 $\hat{}\Phi^{\text{lr}}(\|\cdot\|)$ 映射为离散频谱，只记录其在 Λ' 中符合周期性边界条件的频率处的取值。将其代回式（9），并通过重新定义 ${}^1\Omega\hat{}\Phi^{\text{lr}}(\|\boldsymbol{k}\|) \mapsto \hat{}\Phi^{\text{lr}}(\|\boldsymbol{k}\|)$，将前置因子 ${}^1\Omega$ 吸收到我们的可学习的 滤波器中，便得到埃瓦尔德消息传递所需的长程分量，即对傅里叶频率进行消息求和：

$$M^{\text{lr}}(\boldsymbol{x}_i) = \sum_{\boldsymbol{k}\in\Lambda'} \exp(i\boldsymbol{k}^T\boldsymbol{x}_i)\cdot s_{\boldsymbol{k}}\cdot\hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|), \tag{10}$$

$$s_{\boldsymbol{k}} = \sum_{j\in\mathcal{S}} h_j\exp(-i\boldsymbol{k}^T\boldsymbol{x}_j), \tag{11}$$

与 $\{s_{\boldsymbol{k}}\in\mathbb{C}^F \mid \boldsymbol{k}\in\Lambda'\}$，我们称之为 结构因子嵌入，其灵感源自类似的晶体学量。若将 $\hat{}\Phi^{\text{lr}}$ 替换为截断傅里叶空间核函数 $\hat{}\ \Phi^{\text{lr}}_{[0,c_k]}$，我们便可以限制外层求和 $\sum_{\boldsymbol{k}\in\Lambda'} \mapsto \sum_{\boldsymbol{k}\in\Lambda',\ \|\boldsymbol{k}\|<c_k}$，而计算$s_{\boldsymbol{k}}$ 的内层求和仅遍历单个超胞中的原子。由此得到的截断核函数与常规连续滤波器（式（4））互为补充：连续滤波器不受距离限制，而频率截断则限制了滤波器在空间中的变化速度。更具体地说，其空间分辨率与逆频率截断成正比。

结构因子嵌入。 没有 k截断时，元组集合 $\{(s_{\boldsymbol{k}},\boldsymbol{k}) \mid \boldsymbol{k}\in\Lambda'\}$ 所包含的信息等价于集合 $\{(h_i,\boldsymbol{x}_i) \mid i\in\mathcal{S}\}$，也就是说，它们能够完整量化模型在给定输入结构上的隐藏状态。从另一种角度来看，$\{(h_i,\boldsymbol{x}_i) \mid i\in\mathcal{S}\}$ 等价于一种“嵌入密

度分布”，$\rho(\boldsymbol{x}) = \sum_{\boldsymbol{t}\in\Lambda}\sum_{i\in\mathcal{S}} h_i\delta(\boldsymbol{x}-\boldsymbol{x}_i-\boldsymbol{t})$，其中 δ 表示狄拉克δ分布。其傅里叶变换（按分布意义）为 $\hat{\rho}(\boldsymbol{k}) = \sum_{\boldsymbol{k}'\in\Lambda'}\left[\sum_{j\in\mathcal{S}} h_i \exp(-i\boldsymbol{k}'^T\boldsymbol{x}_j)\right]\delta(\boldsymbol{k}-\boldsymbol{k}') = \sum_{\boldsymbol{k}'\in\Lambda'} s_{\boldsymbol{k}'}\delta(\boldsymbol{k}-\boldsymbol{k}')$，对应于成对的 $\{(s_{\boldsymbol{k}},\boldsymbol{k})\}$。忽略截断范围之外的所有 $s_{\boldsymbol{k}}$，会丢失尺度低于 $\sim c_k^{-1}$ 的原子编码分布的细粒度分辨率信息。从这个意义上说，埃瓦尔德消息传递中引入的结构因子表示对隐藏模型状态进行了粗粒化。此外，嵌入信息以非局部方式表示，因为每个$s_{\boldsymbol{k}}$ 都是所有超胞嵌入的加权和。这种从多到一的关系也可以用来理解埃瓦尔德消息传递的长程特性。

埃瓦尔德消息传递。 通过任意图神经网络层进行常规短程消息传递，并结合公式（10）和（11）中的长程消息传递，这种组合称为埃瓦尔德消息传递。这种组合使埃瓦尔德消息传递能够表达任一有限类别之外的相互作用。我们只需在每个消息传递步骤中将两种消息传递方案的嵌入更新相加，即可将二者结合起来。详情见公式（12）。在我们的实验中，这种集成方式几乎开箱即用。

非周期系统。许多包含重要长程相互作用的结构数据集并未采用周期性边界条件。在这种情况下，没有可用于离散化傅里叶域的倒易晶格，因此结构因子嵌入 $s_{\boldsymbol{k}}$ 变为定义在连续域上的函数 $s(\boldsymbol{k})$：$\mathbb{R}^3 \mapsto \mathbb{C}^F$。非周期情形也可以理解为超胞尺寸趋于无穷大的极限，此时随着基频向量$\hat{\boldsymbol{w}}_i$ 趋近于零，倒易晶格相应地趋于连续极限。由于我们只能用有限个数值表示 $s(\cdot)$ 和 $\Phi(\|\cdot\|)$，因此需要自行施加适当的离散化方案。在我们的实现中，我们用立方体体素网格铺满截断区域，并将所有连续函数替换为它们在每个体素内的平均值。离散化分辨率由体素边长 Δ控制，这是一个与非周期情形相关的新超参数。在主要实验中，我们为每个测试模型采用相同的分辨率 $\Delta = 0.2\ \text{Å}^{-1}$。采用其他分辨率的消融研究见附录 A.3。请注意，只有当体素网格位于随输入结构旋转的坐标系中时，这种方法才能保持旋转不变性。为此，我们采用通过奇异值分解得到的正交归一基，做法与 Gao 和 G¨unnemann（2022年）类似。附录 A.6 概述了用于离散化结构因子的更一般数学框架，提供了与实现相关的体素方法细节，并讨论了该方法与其他可能离散化方案之间的关系。

频率滤波器：非周期情形。到目前为止，我们尚未说明如何学习频率滤波器 $\Phi(\|\cdot\|)$。在非周期情形下，我们将滤波器参数化为 S径向基函数 $\{\hat{\psi}_i(\|\cdot\|)$ 的线性组合：$\mathbb{R}^3 \to \mathbb{R} \mid i = 1、\ldots\ldots、S\}$：$\hat{\Phi}(\|\boldsymbol{k}\|) = \boldsymbol{W}\hat{\Psi}(\boldsymbol{k})$，其中 $\hat{\Psi}(\boldsymbol{k}) = (\psi_1(\|\boldsymbol{k}\|)\ldots\psi_S(\|\boldsymbol{k}\|))$，而 $\boldsymbol{W}\in\mathbb{R}^{F\times S}$ 为学习得到的权重矩阵。所有基函数均在 k截断半径 c_k范围内有定义。我们使用高斯径向基函数（Sch¨utt 等，2018年）。

频率滤波器：周期情形。 在 OC20 上，我们采用一种非径向滤波策略，以便更灵活地利用给定输入结构的周期性边界条件框架所提供的信息（$\boldsymbol{v}_1$，$\boldsymbol{v}_2$，$\boldsymbol{v}_3$）$\in\mathbb{R}^{3\times3}$。这些信息可能很有意义：例如，给定一个材料表面薄片，模型显然应该“知道”哪个方向垂直于表面。这不会破坏旋转不变性，因为周期性边界条件框架会随结构一起旋转。

由于倒易晶格 Λ' 会随输入周期性边界条件框架而变化，因此任意固定截断体积内的频率数量也会变化。受我们的实证发现（附录 A.3）启发，我们改为固定每个倒易基方向上的频率数量（$\boldsymbol{w}_1$、$\boldsymbol{w}_2$、$\boldsymbol{w}_3$）$\in\mathbb{R}^{3\times3}$，不再采用直接频率截断的概念。形式上，我们定义正整数 N_x、N_y、N_z，并考虑乘积集合 $I = \bigtimes_{j=x、y、z}\{-N_j,\ldots,N_j\} \subset \mathbb{Z}^3$。我们通过仅纳入倒格点位置属于 I 的频率，来限制式（10）中傅里叶消息求和的范围。随后，模型会为每个 $\boldsymbol{\lambda'} \in I$ 学习一个滤波器权重 $\hat{\Phi}_{\boldsymbol{\lambda'}} \in \mathbb{R}^F$。由于一般的超晶胞晶格至少具有 ˆ ˆ 点对称性，我们还令 $\Phi_{\boldsymbol{\lambda'}} \equiv \Phi_{-\boldsymbol{\lambda'}}$。这不同于前文讨论的严格频率滤波器概念：滤波器权重 $\Phi_{\boldsymbol{\lambda'}}$ 并非对应于傅里叶空间中的固定点 $\boldsymbol{k}$，而是对应于倒易晶格上的一个位置。与该位置对应的频率 $\boldsymbol{k}_{\boldsymbol{\lambda'}} = \lambda'_1\boldsymbol{w}_1 + \lambda'_2\boldsymbol{w}_2 + \lambda'_3\boldsymbol{w}_3$ 会随每个输入结构的基频率而变化。请注意，这种滤波方法不具有“超晶胞对称性”不变性；该对称性会在保持原子位置不变的同时，沿一个或多个晶格方向将任意超胞映射为其倍数。我们发现，这不会影响 OC20 上的实证性能。不过，对于超胞大小变化较大的数据集，采用对该对称性不变的滤波方法（例如径向基滤波）或许更为合适。

计算复杂度。如图所示，我们的长程消息传递方案由三个部分组成，3，其计算复杂度为 $\mathcal{O}(N_{\text{at}}N_{\text{k}})$，其中 N_{at} 是结构中的原子数（在周期性边界条件下则为超胞中的原子数），而 N_{k} 是长程求和中包含的频率数。由于可以近似假设空间中的原子密度恒定，因此在周期情形下，原子数随超胞体积Ω 线性增长。根据倒易基的定义$(\boldsymbol{w}_1,\boldsymbol{w}_2,\boldsymbol{w}_3) \in \mathbb{R}^{3\times3}$，

固定频率截断体积内包含的频率数 $\boldsymbol{k} \in \Lambda'$ 因此随原子数线性增长 N_{at}。总的来说，如果两个截断值都固定，计算复杂度将达到 $\mathcal{O}(N_{\text{at}}^2)$级。在非周期情形下，若将体素分辨率 Δ 调整至 s的最精细特征，则同样成立；根据傅里叶互易性，这些特征与结构尺度成反比。然而，与标准埃瓦尔德求和不同（Wells &Chaffee，2015年），Ewald MP 并未对 c_x 和 c_k之间设定固定关系。若要达到 $\mathcal{O}(N_{\text{at}})$的计算复杂度，我们可以改为固定 N_{k}，这实际上会改变 $c_k \propto N_{\text{at}}^{-1}$。需要注意的是，这种方法可能会留下一个中程相互作用的“空隙”，既不属于 Ewald MP 的短程分量，也不属于其长程分量。实现次二次方扩展的另一种方法，是将更高效的埃瓦尔德求和变体用于 MPNN（参见第4节）。

4. 相关工作

手工设计的长程校正。通过长程校正来增强短程模型的方法在计算化学中十分常见。这通常涉及对密度泛函理论（DFT）的校正；密度泛函理论是 OC20（Chanussot 等，2021）、OE62（Stuke 等，2020）以及许多其他机器学习基准测试背后的方法。密度泛函理论与消息传递神经网络势面临同类问题：其标准泛函，如 LDA（Jones 与 Gunnarsson，1989）或 PBE（Perdew 等，1996），无法正确表达诸如伦敦色散（Jones，2015）等长程效应。已有大量原子间色散校正方法（Grimme，2006；Tkatchenko 与 Scheffler，2009；Grimme 等，2010；Caldeweyher 等，2019），可适配于许多密度泛函理论泛函。从这一历史视角来看，Ewald 消息传递在 MPNN 的背景下满足了对一种低成本、开箱即用校正的同样需求。其他研究将具有各种基于物理的函数形式的长程项与短程模型相结合，包括完全手工设计的项（Staacke 等，2021），以及采用消息传递神经网络预测的部分电荷的静电项（Unke 等，2021）。相比之下，我们的方法完全处于 MPNN 范式之内，专注于对长程效应进行完全数据驱动的建模。

非局部学习方案。 Ewald 消息传递属于更广泛的一类学习方法，旨在处理结构中的非局部相互作用。注意力机制（Vaswani 等人，2017年）是一个影响力远超量子化学领域的典型例子。早期的图领域方法仅将注意力机制应用于局部节点邻域（Veli˘ckovi´c 等人，2018年；Dwivedi 和 Bresson，2021年），而其他方法很快将其扩展到了全连接图（Ying 等人，2021年； Kreuzer 等人，2021年； Mialon 等人，2021年）。这类标准的全连接图注意力机制的一个缺点是其 $\mathcal{O}(N^2)$ 扩展性会随结构规模而变化。不过，已有研究提出了采用标准注意力次二次复杂度变体的模型，包括一种 $\mathcal{O}(N\log(N))$ 非局部聚合方案（Liu 等人，2022年），以及适用于化学领域（Unke 等人，2021年； Frank 等人，2022年）和通用图应用（Ramp´a˘sek 等人，2022年）的线性扩展改进注意力层。基于高斯过程的 sGDML（Chmiela 等人，2018年）等浅层模型通常也采用局域性近似，这促使近期研究致力于将其扩展至非局部情形，相关工作由Chmiela 等人 （2023）提出。这类方法大多与 MPNN 研究路线正交，也不像 Ewald MP 那样直接借鉴计算物理技术。

长程求和的使用。一些基于 MPNN 的研究与我们采用了相同的方法，即计算多个消息和，这类似于物理学中将手工构造的能量分为成键项和非成键项。它们同样采用了更复杂的短程部分，其中包含键等高阶几何信息（Zhang 等人，2020年），甚至二面角（Gasteiger 等人，2022年），以及更简单、纯径向的原子间长程部分。然而，这两项研究中的长程部分都是在更大的原子邻域内进行标准的距离截断消息求和，因此无法获得埃瓦尔德求和所能实现的扩展性优势和物理归纳偏置。此外，还有两项同期研究将不同形式的埃瓦尔德求和与 MPNN 模型相结合。Lin 等人（2023）以传统方式使用它来计算固定形式的物理势，供下游任务使用，作为具有物理信息的边特征，而非作为学习连续滤波器消息的自由形式假设。Yu 等人（2022年）与我们一样采用可学习的傅里叶空间滤波器，但计算的是标量结构因子，而非结构因子嵌入。更一般地说，Yu 等人（2022年）并未将埃瓦尔德部分用作图神经网络的消息传递步骤，也不会将其反馈至基础图神经网络嵌入；相反，他们只应用一次，然后在最后一层之后将其与短程图神经网络输出相结合。这两项与埃瓦尔德相关的研究都只考虑了周期情形。

埃瓦尔德求和。埃瓦尔德求和是一种用于计算长程相互作用总和的技术。其概念可追溯至一个多世纪前的理论物理学研究（Ewald，1921），此后经过多次改进，以涵盖更一般的相互作用（Nijboer 和 De Wette，1957；Laino 和 Hutter，2008；Mazars，2010）。后续开发主要聚焦于高效数值计算长程求和。尽管采用经过优化的距离和频率截断进行直接埃瓦尔德求和时，其复杂度随原子数呈 $N^{\frac{3}{2}}$ 增长（Wells 和 Chaffee，2015），但采用快速傅里叶变换的方法（Darden 等人，1993；York 和 Yang，1994；Essmann 等人，1995）可将复杂度降至 $N\log(N)$。在本研究中，我们着重介绍埃瓦尔德消息传递这一通用方法，并将数值改进视为未来研究的一个有趣方向。

5. 实验

数据集。 OC20 数据集（Chanussot 等人，2021年）包含约 2.65 亿个结构的吸附能和原子力，每个结构都是在三维周期性边界条件下，吸附物-催化剂组合经密度泛函理论计算所得弛豫轨迹中的一个快照。我们使用 OC20-2M 子集训练模型；经实证，该子集可有效代表完整训练集（Gasteiger 等人，2022年）。OE62 数据集（Stuke 等人，2020年）包含多种目标，其中包括约 62,000 个大型有机分子的密度泛函理论计算能量（单位为 电子伏特）。这些能量计入了长程伦敦色散相互作用，该相互作用通过 DFT-D3 色散校正计算得出（Grimme 等人，2010年）。这两个基准测试都包含原子数相当多的结构（$N > 100$）。OC20 具有周期性边界条件，而 OE62 中的结构虽为非周期性，但空间范围可能很大（> 20 Å）。因此，这两个数据集非常适合用于检验埃瓦尔德消息传递所能带来的长程改进。

基线图神经网络模型。 训练和基线超参数概览见附录 A.2。我们通过修改四种模型来测试 Ewald 消息传递：SchNet（Schütt 等人，2017年）、PaiNN（Schütt 等人，2021年）、DimeNet^{++}（Gasteiger 等人，2020年）和 GemNet-T（Gasteiger 等人，2021年）。这些模型旨在代表现代图神经网络领域中的一部分，但绝非穷尽。参考Chanussot 等人（2021年）采用的设置，大多数基线模型的原子嵌入维度和截断半径都大于原文报告的数值，以适应我们所用结构的规模和化学多样性。

Ewald 修正。 我们的目标是以尽可能精简的方式升级每种基线架构，以便单独研究 Ewald 修正的作用。除 DimeNet^{++}外，所有模型都会在每个交互块中通过跳跃连接更新原子嵌入。我们只需将长程消息传递作为另一项贡献加入，使用 M_i^{lr}来自公式（10）和（11）的内容，

$$h_i^{(l+1)} = \frac{1}{\sqrt{2}}\left[h_i^{(l)} + f_{\text{upd}}^{\text{sr}}\left(M_i^{\text{sr}}\right)\right]$$
$$h_i^{(l+1)} = \frac{1}{\sqrt{3}}\left[h_i^{(l)} + f_{\text{upd}}^{\text{sr}}\left(M_i^{\text{sr}}\right) + f_{\text{upd}}^{\text{lr}}\left(M_i^{\text{lr}}\right)\right], \tag{12}$$

其中，短程消息和 M_i^{sr}和更新函数f_{srupd}针对短程模型而定，而 $f_{\text{upd}}^{\text{lr}}$在不同模型中属于同一类型（架构细节见附录 A.1）。对于同样采用等变向量嵌入的 PaiNN，我们只升级其标量嵌入之间的消息传递。由于 DimeNet^{++} 默认不在原子嵌入之间传递消息，我们使用取自 GemNet-T 的默认嵌入块将其加入模型，并仅通过长程消息传递对其进行更新，采用跳跃连接，$h_i^{(l+1)} = \frac{1}{\sqrt{2}}\left[h_i^{(l)} + f_{\text{upd}}\left(M_i^{\text{lr}}\right)\right]$。我们沿特征维度拼接这些原子嵌入与按原子聚合的 DimeNet^{++} 边嵌入，然后将其输入每个输出块中的逐原子全连接层。我们的 Ewald 修正之间其余的差异都体现在超参数层面。为了实现帕累托高效的基线改进，我们根据各基线的计算成本，选择适当数量的纳入 k- 向量（见附录 A.3）。

比较研究。我们检验在基线中加入 Ewald 消息传递，是否能显著且稳定地降低 OC20 和 OE62 上的能量平均绝对误差（能量平均绝对误差）。我们还测量每种模型配置和数据集的每个结构的平均运行时间（详情见附录A.8），以研究 Ewald 消息传递的优势与其运行时间开销之间的权衡。此外，为了建立一个基线，评估仅通过改变超参数所能实现的长程改进，我们在不使用 Ewald MP、但提高距离截断 c_x 和消息传递邻居最大数量 N_{max} 的配置下，重复测量能量平均绝对误差和运行时间（详情见A.4）。理想情况下，恰当的比较应将这些设置提高到成本大致与 Ewald 改动相当的程度。对于 OC20 的周期性结构，这始终可行。对于 OE62，设置为 $c_x = 12$ Å 和$N_{\text{max}} = 100$ 通常就能覆盖整个分子，这意味着超过此设置后无法显著增加成本。此外，我们将 Ewald 改进与另一种测试设置进行比较：将方程（12）中的 Ewald 模块 $_{\text{lr}_i}$ ）替换为距离截断增大的常规成对消息传递（与第4节讨论的相关工作类似）。具体而言，我们用一个 SchNet 交互模块替换它。与前述增大截断的变体不同，这次我们仅扩展该模块的截断范围。最后，我们考察 Ewald MP 的优势是否显著超出仅增加自由参数所带来的效果。在 OE62 上，我们测量增加原子嵌入通道的基线变体的能量平均绝对误差和运行时间，以使参数数量与 Ewald MP 相当。对于 OC20，我们引用 Gasteiger 等人（2022年a）开展的大量超参数研究；研究结果表明，将嵌入维度调整本工作涉及的小幅度，对 OC20 的能量平均绝对误差几乎没有影响。

首先，我们在 OE62（图 5）和 OC20（图 6）上的比较研究，凸显了在图神经网络设计格局中，短程模型选择所起的主要作用（例如，DimeNet^{++} 在 OC20 上被帕累托支配）。与此同时，Ewald 消息传递带来的准确率提升始终如一，这有力地表明应将其与任何特定的短程模型搭配使用。其运行时间开销至少与基线模型之间的运行时间差异相当，甚至更小。此外，在 OE62 上，所有模型的平均 Ewald

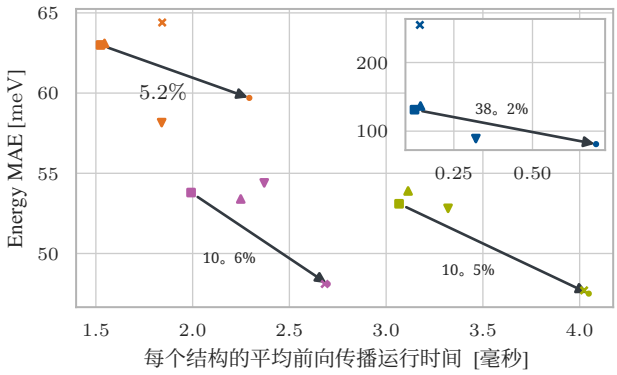


图 *5.* OE62 上的计算成本与能量平均绝对误差。Ewald 消息传递使模型性能提升了 5 % 至 38 %。增大截断距离对 DimeNet^{++} 和 GemNet-T 的效果相近，但对 SchNet 和 PaiNN 甚至有害。用 SchNet 替代 Ewald 长程模块对 SchNet 和 PaiNN 效果良好，但对 DimeNet^{++} 和 GemNet-T 没有影响。增大原子嵌入维度的影响甚微。

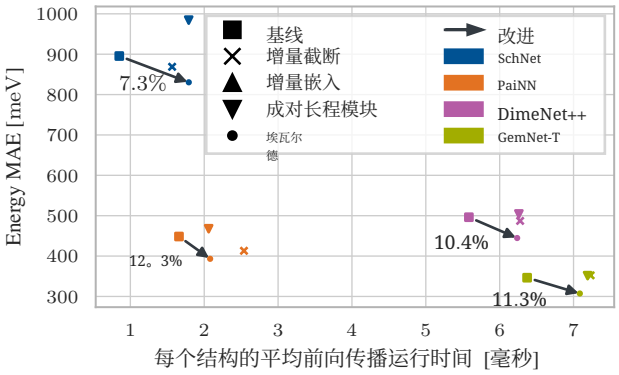


图 *6.*OC20 上的计算成本与能量平均绝对误差（对各测试划分取平均）。Ewald 消息传递始终能提升所有模型的性能，并显著优于两种基线设置：增大全局截断距离，以及用增大距离截断的标准成对模块（此处为 SchNet）替代 Ewald 长程模块。

改进（在 OC20 上还涵盖所有测试划分）为 16.1%（OE62）和 10.3%（OC20）。在 OC20 的所有测试实例中，Ewald 消息传递的成本效益权衡优于较大的截断设置。有趣的是，在 OE62 上，对于使用边嵌入的两个基线模型，较大的截断设置与 Ewald 消息传递的权衡非常相近；但对于仅使用原子嵌入的两个模型，则会产生负面影响。若在 OE62 上将 Ewald 模块替换为距离截断更大的 SchNet 模块，情况则截然相反：这种做法对 PaiNN 和 SchNet 有效，但对表现更好的 DimeNet++和 GemNet-T 基线模型则毫无改进。由于新增的 SchNet 模块已使图连通性接近全连接，其成本无法再进一步增加。在 OC20 上，这种做法始终没有效果或会产生负面影响。与前两种策略不同，Ewald 消息传递带来的改进似乎 稳健 ——在任何 经过测试的模型配置中都未观察到负面影响。最后，将 Ewald 消息传递的自由参数数量与更大的原子嵌入相匹配，效果微乎其微，甚至可能产生负面影响。除这些实证比较之外，我们还指出，一般而言，在全连接图上进行成对消息传递（例如我们在 OE62 上采用的 SchNet LR 模块设置），其计算量必然随结构尺寸呈二次增长。相比之下，Ewald 消息传递能够以低于二次的复杂度传播长程信息，同时还提供基于物理原理的归纳偏置（参见第 3节）。

我们的结果为今后的应用提供了若干设计建议。Ewald 消息传递在 OC20 上似乎效率更高：以约一半的相对成本，便能持续取得相近的相对改进。这表明周期性结构可能是更有吸引力的应用领域。鉴于我们将 Ewald 消息传递与短程图神经网络结合的方式，当模型采用数量较少但计算开销较大的消息传递模块时（在我们的案例中为 DimeNet^{++} 和 GemNet-T），为每个模块添加长程消息求和所带来的相对开销会更低。除能量外，我们还测试了 OC20 上力的平均绝对误差是否有所改善，结果显示虽有改善，但幅度较小。这在意料之中，因为长程消息传递的频率截断方案会去除势能面中快速变化的高频贡献，从而限制其对能量梯度（力）的影响。更多讨论和数值结果见附录 A.7 和 A.11。

对长程相互作用的影响。为了检验我们的假设——Ewald 消息传递尤其能够学习长程效应——我们利用了 OE62 中的能量目标包含一个加性项这一事实：DFT-D3（Grimme 等人，2010年），用于校正长程伦敦色散相互作用。这使我们能够单独研究长程分量。为评估 DFT-D3 项，我们采用针对 Becke-Johnson 阻尼和 PBE0 密度泛函推荐的系数组合，该组合取自 Reckien（2017年）。接下来，我们根据每个分子的 DFT-D3 校正值，将`OE62-val` 数据集中的全部 6000 个分子划分到 15 个大小相同的区间中。由于最外侧的两个区间包含明显的离群值，我们将其排除。在每个区间中，我们记录基线和 Ewald 模型变体的能量平均绝对误差统计量（自助法均值和68% 置信区间，基于 10000 次重采样）。图 7 显示，对于长程能量贡献较高的结构，Ewald 消息传递带来了显著更大的相对改进。作为额外验证，我们还收集了此前所述的相同统计量，针对一种不采用 Ewald 修正、但在训练时将 DFT-D3 校正显式加到模型输出中的模型。这样便得到一个“作弊”模型，它只需学习短程分量，而长程分量

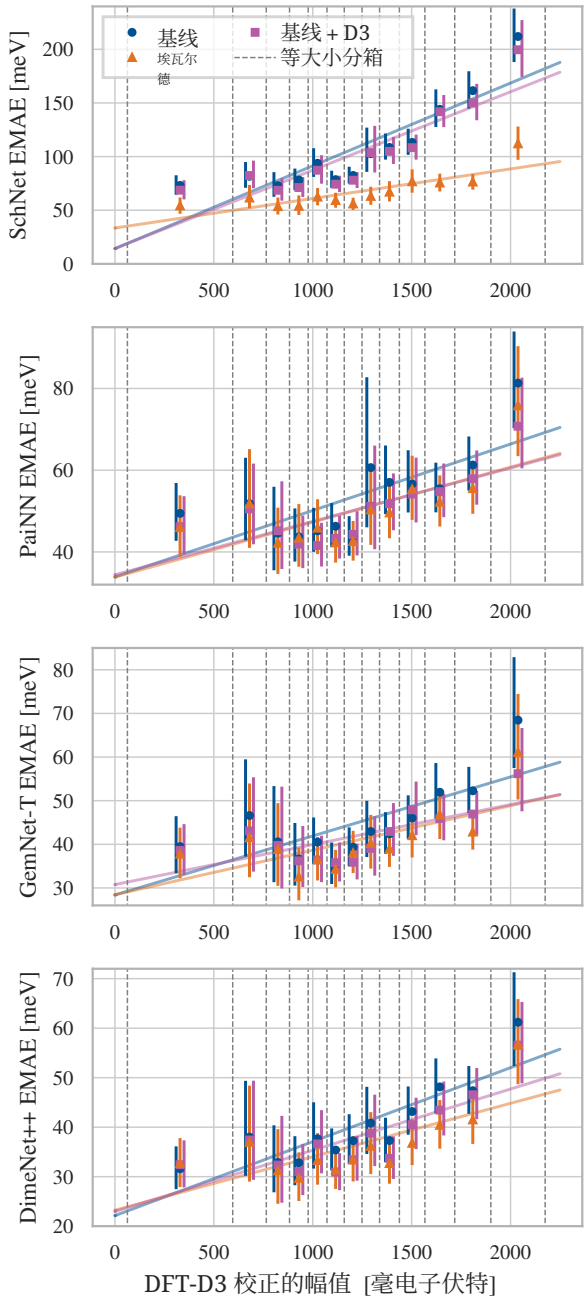


图 7. 按 DFT-D3 长程项分箱的 OE62-val 结构上的能量平均绝对误差。对于长程分量较大的分子，Ewald 消息传递的影响尤为显著，其效果可恢复甚至超过一种“作弊”基线：该基线直接将 D3 真实值加到预测值上（“基线 + D3”）。

可以被完美重构。对于所有模型，Ewald 消息传递的表现都不逊于这一完美基线；对于 SchNet 和 DimeNet^{++}，其表现甚至更优。这项分析表明，Ewald 消息传递无需任何人工设计，就能在 OE62 上恢复色散校正的效果。

6. 结论

在这项工作中，我们展示了图神经网络消息之和与物理相互作用之间的类比如何引出 Ewald 消息传递——一种基于埃瓦尔德求和的通用消息传递框架。该框架计算高效且模型无关，可作为即插即用的扩展，为许多现有图神经网络模型带来价值。我们的结果表明，在不同模型和数据集上，能量目标均得到稳健改进。此外，结果还显示，对于长程能量贡献较大的结构，该方法的效果尤为显著。我们建议在包含多种原子的较大结构或周期性结构中尤其使用 Ewald 消息传递。

致谢

我们衷心感谢 Abhishek Das 迅速协助我们在 OC20-test 数据划分上评估模型，也感谢匿名审稿人提出富有启发性的建议。

本研究由慕尼黑量子谷资助，资金来自联邦教育与研究部（BMBF）的拨款、巴伐利亚高科技议程 Plus 的资金，以及联邦教育与研究部和巴伐利亚自由州根据联邦政府与各州卓越战略提供的资金。

参考文献

Alon, U. and Yahav, E. 关于图神经网络的瓶颈及其实际影响。在*ICLR*，2021年。网址：https://openreview。net/forum?id=i80OPhOCVH2。

Anderson, B. M., Hy, T.-S., 和 Kondor, R. Cormorant：具协变性的分子神经网络。在 *NeurIPS*，2019年。网址 http://arxiv。org/abs/1906。04015。

Batatia, I., Batzner, S.,Kov´acs, D. P., Musaelian, A., Si mm, G. N. C., Drautz, R., Ortner, C., Kozinsky, B., 和 Cs´ anyi, G. e(3) 等变原子中心原子间势的设计空间，2022年。网址https://arxiv。org/abs/2205。06643。

Caldeweyher, E., Ehlert, S., Hansen, A., Neugebauer, H., Spicher, S., Bannwarth, C. 和 Grimme, S. 一种普遍适用的原子电荷依赖的伦敦色散校正。《化学物理学杂志》，150(15): 154122，2019年。doi: 10.1063/1.5090222。网址 https://doi.org/10.1063/1.5090222.

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., Palizhati, A., Sriram, A., Wood, B., Yoon, J., Parikh, D., Zitnick, C. L. 和 Ulissi, Z. Open Catalyst 2020（OC20）数据集与社区挑战赛。《*ACS*催化》，11(10):6059–6072，2021年5月。doi:

10.1021/acscatal.0c04525。网址 https://doi.org/10.1021/acscatal.0c04525。出版方：美国化学会。

Cheng, J., Zhang, C. 和 Dong, L. 一种用于预测材料性质的几何信息增强晶体图网络。 *Communications Materials*, 2 (1):92，2021年9月。ISSN2662-4443。doi: 10。1038/s43246- 021-00194-3。URLhttps://doi。org/10。1038/s43246-021-00194-3。

Chmiela, S., Sauceda, H.E., M¨uller, K.-R. 和 Tkatchenko, A. 利用机器学习力场实现精确分子动力学模拟。 *Nature Communications*，9(1):1–10，2018年9月。ISSN 2041-1723。doi: 10。1038/s41467-018-06169- 2。URLhttps://www。nature。com/articles/s41467-018-06169-2。

Chmiela, S., Vassilev-Galindo, V., Unke, O.T., Kabylda,A., Sauceda, H. E., Tkatchenko, A., andM¨uller, K.- R. 针对含数百个原子的分子的精确全局机器学习力场。 *ScienceAdvances*，9(2):eadf0873，2023。doi: 10。1126/sciadv。adf0873。URLhttps://www。science。org/doi/abs/10。1126/sciadv。adf0873。

Darden, T., York, D., and Pedersen, L. 粒子网格 Ewald：一种用于大型系统中 Ewald 求和的 nlog(n) 方法。《化学物理学杂志》，98(12): 10089–10092，1993。doi: 10。1063/1。464397。URLhttps://doi。org/10。1063/1。464397。

Dwivedi, V. P. 和 Bresson, X. 将 Transformer 的一种泛化方法应用于图，2021年。

Dym, N. 和 Maron, H. 关于旋转等变点云网络的普适性。在*ICLR*，2021年。网址https://openreview。net/forum?id=6NFBvWlRXaG。

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. 和 Pedersen, L. G. 一种平滑粒子网格 Ewald 方法。《化学物理学杂志》，103 (19):8577–8593，1995年。doi: 10。1063/1。470117。网址https://doi。org/10。1063/1。470117。

Ewald, P. P. 光学和静电晶格势的计算。物理学年鉴，369(3):253–287，1921年1月。doi: 10。1002/andp。19213690304。

Frank, T., Unke, O. T., 和 Muller, K. R. So3krates：用于分子系统中任意长度尺度相互作用的等变注意力机制。载于 Oh, A. H., Agarwal, A., Belgrave, D., 和 Cho, K.（编），神经信息处理系统进展，2022年。网址：https://openreview。net/forum?id=tlUnxtAmcJq。

Freysoldt, C., Neugebauer, J., Tan, A. M. Z., 和 Hennig, R. G. 经验性超胞外推在计算体相、表面和二维材料中的点缺陷时的局限性。《物理评论*B*》，105：014103，2022年1月。doi：10。1103/PhysRevB。105。014103。网址：https://link。aps。org/doi/10。1103/PhysRevB。105。014103。

Gao, N. 和 G¨unnemann, S. 将图神经网络与神经波函数相结合，构建从头算势能面。载于国际学习表征会议，2022年4月。Gasteiger, J., Giri, S., Margraf, J. T., 和 G¨unnemann, S. 面向非平衡分子的快速且具备不确定性感知能力的方向性消息传递。载于分子机器学习研讨会，*NeurIPS*，2020年12月。网址：http://arxiv。org/abs/2011。14115。arXiv：2011.14115。

Gasteiger, J., Becker, F., andG¨unnemann, S. Gem- Net：用于分子的通用方向性图神经网络。发表于 *NeurIPS*，2021年11月。网址：http://arxiv.org/abs/2106.08903。arXiv：2106.08903。

Gasteiger, J., Shuaibi, M., Sriram, A., G¨unnemann, S., Ulissi, Z., Zitnick, C. L., and Das, A. 图网络如何推广到大型且多样化的分子系统? *arXiv*，2204.02782，2022年4月。网址：http://arxiv.org/abs/2204.02782。arXiv：2204.02782。

Gasteiger, J., Shuaibi, M., Sriram, A., G¨unnemann, S., Ulissi, Z., Zitnick, C. L., and Das, A. Gemnet-oc：为大型且多样化的分子模拟数据集开发图神经网络，2022年。网址：https://arxiv.org/abs/2204.02782.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., 和 Dahl, G. E. 用于量子化学的神经消息传递。载于 Precup, D. 和 Teh, Y. W.（编），国际机器学习会议，机器学习研究论文集第 70 卷机器学习研究论文集，第 1263–1272 页。PMLR，2017年。网址http://proceedings。mlr。press/v70/gilmer17a。html。

Grimme, S. 采用长程色散校正构建的半经验 GGA 型密度泛函。计算化学期刊，27(15):1787–1799，2006年。doi: https://doi。org/10。1002/jcc。20495。网址https://onlinelibrary。wiley。com/doi/abs/10。1002/jcc。20495。

Grimme, S., Antony, J., Ehrlich, S., 和 Krieg, H. 针对 94 种元素 h-pu 的密度泛函色散校正（密度泛函理论色散校正）提出一致且准确的从头算参数化。《化学物理学杂志》，132 (15):154104，2010年。doi: 10。1063/1。3382344。URLhttps://doi。org/10。1063/1。3382344。

Jones, R. O. 密度泛函理论：其起源、崛起及未来。《现代物理评论》，87: 897–923，2015年8月。doi: 10。1103/RevModPhys。87。897。URL https://link。aps。org/doi/10。1103/RevModPhys。87。897。

Jones, R. O. 和 Gunnarsson, O. 密度泛函形式体系、其应用与前景。《现代物理评论》，61:689–746，1989年7月。doi: 10.1103/RevModPhys.61.689。网址 https://link.aps.org/doi/10.1103/RevModPhys.61.689.

Kreuzer, D.、Beaini, D.、Hamilton, W. L.、L´etourneau, V. 和 Tossou, P. 重新思考具有谱注意力机制的图 Transformer。载于 Beygelzimer, A.、Dauphin, Y.、Liang, P. 和 Vaughan, J. W.（编），《神经信息处理系统进展》，2021年。网址：https: //openreview.net/forum?id=huAdB-Tj4yG。Laino, T. 和 Hutter, J. 关于“埃瓦尔德求和法计算最高至四极阶的静电多极相互作用”的注释 [j. chem. phys. 119, 7471 (2003)]。《化学物理学杂志》，129(7):074102，2008年。doi: 10.1063/1.2970887。网址：https://doi.org/10.1063/1.2970887。Leach, A.《分子建模：原理与应用》。普伦蒂斯·霍尔，哈洛，英格兰；纽约，第2版，2001年1月。ISBN 978-0-582-38210-7。Lin, Y.、Yan, K.、Luo, Y.、Liu, Y.、Qian, X. 和 Ji, S. 用于晶体性质预测的完整原子间势的高效近似，2023年。网址：https: //openreview.net/forum?id=KaKXygtEGK。Liu, M.、Wang, Z. 和 Ji, S. 非局部图神经网络。《IEEE 模式分析与机器智能汇刊》，44(12):10270–10276，2022年。doi: 10.1109/TPAMI.2021.3134200。Mallat, S. 第7章——小波基。载于 St´ephane, M.（编），《信号处理小波之旅》（第三版），第263–376页。学术出版社，波士顿，第三版，2009年。ISBN 978-0-12-374370-1。doi: https: //doi.org/10.1016/B978-0-12-374370-1.00011-2。网址：https://www.sciencedirect.com/science/article/pii/B9780123743701000112。Mazars, M. 三维及准二维系统中逆幂律相互作用的埃瓦尔德方法。《物理学A：数学与理论》，43(42):425002，2010年9月。doi: 10.1088/1751-8113/43/42/425002。网址：https://dx.doi.org/ 10.1088/1751-8113/43/42/425002。Mialon, G.、Chen, D.、Selosse, M. 和 Mairal, J. Graphit：在 Transformer 中编码图结构，2021年。

Nijboer, B. 和 De Wette, F. 关于晶格求和的计算。《Physica》，23(1):309–321，1957年。ISSN 0031-8914。doi: https://doi.org/10.1016/S0031-8914(57)92124-9。URL https://www.sciencedirect.com/science/article/pii/S0031891457921249。Perdew, J. P.、Burke, K. 和 Ernzerhof, M. 简化的广义梯度近似。《Phys. Rev. Lett.》，77:3865–3868，1996年10月。doi：10.1103/PhysRevLett.77.3865。URL https://link.aps.org/doi/10.1103/PhysRevLett.77.3865。Ramp´a˘sek, L.、Galkin, M.、Dwivedi, V. P.、Luu, A. T.、Wolf, G. 和 Beaini, D. 通用、高效且可扩展的图转换器构建方法。载于 Koyejo, S.、Mohamed, S.、Agarwal, A.、Belgrave, D.、Cho, K. 和 Oh, A.（编），《神经信息处理系统进展》，第35卷，第14501–14515页。Curran Associates, Inc.，2022年。URL https://proceedings.neurips.cc/ paper files/paper/2022/file/5d4834a159f1547b267a05a4e2b7cf5e-Paper-Conference.pdf。Reckien, W. BJ 阻尼的泛函及系数列表，2017年。URL https: //www.chemie.uni-bonn.de/pctc/mulliken-center/software/dft-d3/functionalsbj。Sch¨utt, K.、Unke, O. 和 Gastegger, M. 用于预测张量性质和分子光谱的等变消息传递。载于 Meila, M. 和 Zhang, T.（编），《第38届国际机器学习会议论文集》，《机器学习研究论文集》第139卷，第9377–9388页。PMLR，2021年7月18–24日。URL https://proceedings.mlr.press/v139/schutt21a.html。Sch¨utt, K.、Kindermans, P.-J.、Sauceda Felix, H. E.、Chmiela, S.、Tkatchenko, A. 和 M¨uller, K.-R. SchNet：一种用于建模量子相互作用的连续滤波卷积神经网络。载于 Guyon, I.、Luxburg, U. V.、Bengio, S.、Wallach, H.、Fergus, R.、Vishwanathan, S. 和 Garnett, R.（编），NeurIPS，第991–1001页。Curran Associates, Inc.，2017年。URL http://papers.nips.cc/paper/6700-schnet-a-continuous-filter-convolutional-neural-network-for-modeling-quantum-interactions.pdf。Sch¨utt, K. T.、Sauceda, H. E.、Kindermans, P.-J.、Tkatchenko, A. 和 M¨uller, K.-R. SchNet——一种用于分子和材料的深度学习架构。《化学物理学杂志》，148(24):241722，2018年3月。ISSN 0021-9606。doi：10.1063/1.5019779。URL https://aip.scitation.org/doi/abs/10.1063/1.5019779。

Staacke, C. G.、Heenen, H. H.、Scheurer, C.、Cs´anyi, G.、Reuter, K. 和 Margraf, J. T. 长程静电作用在复杂电池材料的机器学习原子间势中的作用。《*ACS Applied Energy Materials*》，4(11):12562–12569，2021年。doi：10.1021/acsaem.1c02363。URLhttps://doi.org/10.1021/acsaem.1c02363.

Stuke, A., Kunkel, C., Golze, D., Todorovi´c, M., Margraf, J. T., Reuter, K., Rinke, P. 和 Oberhofer, H. 61,489种形成晶体的有机分子的原子结构和轨道能量。《科学数据》，7(1):58，2020年。

Tkatchenko, A. 和 Scheffler, M. 基于基态电子密度和自由原子参考数据，准确计算分子范德华相互作用。物理评论快报，102:073005，2009年2月。doi: 10。1103/PhysRevLett。102。073005。网址 https://link。aps。org/doi/10。1103/PhysRevLett。102。073005。

Unke, O. T., Chmiela, S., Gastegger, M., Sch¨utt, K. T., Sauceda, H. E. 和 M¨uller, K.-R. SpookyNet：利用电子自由度和非局域效应学习力场。*NatureCommunications*，12(1):7273，2021年12月。ISSN 2041-1723。doi: 10。1038/s41467- 021-27504-0。网址https://www。nature。com/articles/s41467-021-27504-0。编号：1。出版方：自然出版集团。

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. 和 Polosukhin, I. 注意力机制即一切所需。在 Guyon, I., Luxburg, U. v., Bengio, S., Wallach, H. M., Fergus, R., Vish- wanathan, S. V. N. 和 Garnett, R.（编），*NeurIPS*，第 5998–6008 页，2017年。网址：http://papers。nips。cc/paper/7181-attention-is-all-you-need。

Veli˘ckovi´c, P., Cucurull, G., Casanova, A., Romero, A., Li`o, P., 和 Bengio, Y. 图注意力网络。载于 ICLR，2018年。网址 http://arxiv.org/abs/1710.10903。arXiv：1710.10903。Wells, B. A. 和 Chaffee, A. L. 分子模拟中的埃瓦尔德求和。《Journal of Chemical Theory and Computation》，11(8):3684–3695，2015年。doi：10.1021/acs.jctc.5b00093。网址 https://doi.org/10.1021/acs.jctc.5b00093。PMID：26574452。Xie, T. 和 Grossman, J. C. 用于准确且可解释地预测材料性质的晶体图卷积神经网络。《Physical Review Letters》，120(14):145301，2018年4月。doi：10.1103/PhysRevLett.120.145301。网址 https://link.aps.org/doi/10.1103/PhysRevLett.120.145301。

Xie, T., Fu, X., Ganea, O.-E., Barzilay, R., 和 Jaakkola, T. 用于周期性材料生成的晶体扩散变分自编码器，2021年。网址 https://arxiv.org/abs/2110.06197.

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y. 和 Liu, T.-Y. Transformer 在图表示任务上的表现真的很差吗? 载于*NeurIPS*，2021年5月。网址：https://openreview。net/forum?id=OeWooOxFwDa。

York, D. 和 Yang, W. 用于计算埃瓦尔德求和的快速傅里叶泊松方法。《化学物理学杂志》，101(4):3298–3300，1994年。doi: 10.1063/1.467576。URLhttps://doi.org/10.1063/1.467576。

Yu, H.、Hong, L.、Chen, S.、Gong, X. 和 Xiang, H. 使用倒易空间神经网络捕捉长程相互作用，2022年。

Zhang, S.、Liu, Y. 和 Xie, L. 使用多重图构建分子力学驱动的图神经网络，用于分子结构建模。发表于 分子机器学习研讨会，*NeurIPS*，2020年11月。网址：http://arxiv.org/abs/2011.07457。arXiv：2011.07457。

A. 附录

A.1. 架构与实现细节

图 8 展示了用于得出表达式 $f_{\text{upd}}^{\text{lr}}\left(M_{\text{lr}_i}\right)$（如公式（12）所示）的详细计算步骤；在每个交互步骤中，我们将其作为跳跃连接的一部分，添加到图神经网络特有的短程嵌入更新中（公式（12））。黄色高亮部分计算的是长程求和M_i^{lr}，我们将在第 3 节讨论该求和。索引 i 表示目标原子，而源索引 j 遍历结构中的所有原子（在周期性边界条件下，则遍历单个超胞中的原子）。

实部和虚部。虽然我们使用复指数表示法来简化本文的概念部分，但在实现中，我们分别处理结构因子的实部和虚部。式（11）给出 $s_{\boldsymbol{k}} = \sum_{j\in\mathcal{S}} h_j \cos(\boldsymbol{k}^T\boldsymbol{x}_j)$，$\text{Im}(s_{\boldsymbol{k}}) = -\sum_{j\in\mathcal{S}} h_j \sin(\boldsymbol{k}^T\boldsymbol{x}_j)$。将其代入式（10），并分别考虑 M_i^{lr}的实部和虚部，可得

$$\begin{aligned}\text{Re}(M_i^{\text{lr}}) &= \sum_n \cos(\boldsymbol{k}_n^T\boldsymbol{x}_i)\sum_{j\in\mathcal{S}} h_j\cos(\boldsymbol{k}_n^T\boldsymbol{x}_j)\hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n)\\ &\quad+\sum_n \sin(\boldsymbol{k}_n^T\boldsymbol{x}_i)\sum_{j\in\mathcal{S}} h_j\sin(\boldsymbol{k}_n^T\boldsymbol{x}_j)\hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n),\\ \text{Im}(M_i^{\text{lr}}) &= 0.\end{aligned} \tag{13}$$

这一处理方式导出了图 8 所示的步骤。在采用直接截断进行径向滤波的非周期情形中，索引 n 枚举截断范围内的所有频率 $\{\boldsymbol{k}_n \in \Lambda', \|\boldsymbol{k}_n\| \le c_k\}$，并且有 $\hat{}\,\hat{}\,\Phi(\boldsymbol{k}_n) \equiv \Phi(\|\boldsymbol{k}_n\|)$。在周期情形中，我们舍弃 Φ 的径向对称性以及直接截断的概念，转而确定 $n \equiv \lambda' \in I$，其索引集合 I枚举第 3 节定义的靠近原点的倒易晶格点。为得到上述结果（尤其是 $\text{Im}(M_i^{\text{lr}}) = 0$），我们隐式地使用了傅里叶滤波器系数 $\hat{}\Phi^{\text{lr}}(\boldsymbol{k}_n)$ 为实数这一性质；然而，仅凭位置空间势 Φ^{lr} 为实数，只能保证 $\hat{}\Phi^{\text{lr}}(-\boldsymbol{k}_n) = \hat{}\Phi^{\text{lr}}(\boldsymbol{k}_n)^*$，其中 $*$ 表示复共轭。所有系数 $\hat{}\Phi^{\text{lr}}(\boldsymbol{k}_n)$ 均为实数这一性质，则源于我们对傅里叶空间滤波器施加的额外点对称约束。在非周期滤波情形中，点对称性直接来自径向对称性：$\hat{}\Phi^{\text{lr}}(\|-\boldsymbol{k}_n\|) = \hat{}\Phi^{\text{lr}}(\|\boldsymbol{k}_n\|)$。在周期情形中，我们通过 $\Phi_{\boldsymbol{\lambda}'} \equiv \Phi_{-\boldsymbol{\lambda}'}$ 对所有 $\boldsymbol{\lambda}' \in I$施加对称性。施加此约束的原因是，由任意基张成的 $\mathbb{R}^3$ 晶格总是关于原点具有点对称性——因此，对于两个仅在各自周期性边界条件坐标系上存在点反射差异、而原子位置相同的结构，模型应产生相同的输出。

阻尼值。在非周期情形中，式（13）中的所有三角函数都乘以一个阻尼值$f_n(\boldsymbol{x})$，该值取决于位置 $\boldsymbol{x}$ 以及频率 $\boldsymbol{k}_n$，并且会因结构因子的离散化方案而异（附录 A.6）。在我们测试的体素平均离散化情形中，体素分辨率为 Δ，我们有

$$f_n(\boldsymbol{x}) = \prod_{c\in\{x,y,z\}} \text{sinc}\left(\frac{k_n^c x^c \Delta}{2}\right), \tag{14}$$

其中，上标 c 用于枚举向量分量。

滤波方案的实现。除阻尼值外，周期情形与非周期情形在实现方面的第二个差异，源于我们对傅里叶空间滤波方案的选择（第 3 节）。在非周期情形中，矩阵乘积 $\boldsymbol{W}_\uparrow\boldsymbol{W}_\downarrow$ 得到一个权重矩阵 $\boldsymbol{W}$，我们将其应用于径向基函数值集合，以得到滤波值 $\{\hat{}\Phi^{\text{lr}}(\|\boldsymbol{k}_n\|)$ $\boldsymbol{k}_n \in \Lambda', \boldsymbol{k}_n \le c_k\}$（如第 3 节非周期滤波段落所述）。在周期情形中，我们不将乘积 $\boldsymbol{W}_\uparrow\boldsymbol{W}_\downarrow$ 直接应用于向量，而是根据当前这对权重矩阵（$\boldsymbol{W}_\uparrow, \boldsymbol{W}_\downarrow$）计算并存储所得矩阵。我们将该矩阵的转置 $(\boldsymbol{W}_\uparrow\boldsymbol{W}_\downarrow)^T$ 视为滤波系数 $\{\hat{}\Phi_{\boldsymbol{\lambda}'} = \hat{}\Phi^{\text{lr}}(\boldsymbol{k}_{\boldsymbol{\lambda}'}) = \hat{}\Phi^{\text{lr}}(\boldsymbol{k}_n)\ \boldsymbol{\lambda}' \in I\}$（回顾一下，我们将 n 对应于多重指标 $\boldsymbol{\lambda}'$）。无论是周期情形还是非周期情形，最终都会得到一个形状为 (N_{k}, F) 的矩阵，其中包含 F 维的滤波值堆栈 $\hat{}\Phi^{\text{lr}}(\boldsymbol{k}_n)$；我们按照式（13）中的规定，将其与结构因子逐元素相乘。受 Gasteiger 等人（2021）的启发，我们使用下投影 $\boldsymbol{W}_\downarrow$ 将特征映射到维度为 $N_\downarrow$ 的瓶颈层（所有交互块 l 共享权重值），以促进泛化。因此，$\boldsymbol{W}_\uparrow$ 的形状为 $(F, N_\downarrow)$，而 $\boldsymbol{W}_\downarrow$ 的形状为 $(N_\downarrow, N_{\text{k}})$（周期情形）或 $(N_\downarrow, N_{\text{RBF}})$（非周期情形）。对于 GemNet-T，我们设定 $N_\downarrow = 16$；其他情况则设定 $N_\downarrow = 8$。

更新函数。作为更新函数 $f_{\text{upd}}^{\text{lr}}$，所有模型均使用一个全连接层，后接 N_{hidden} 个残差层，具体设置参见 Gasteiger 等人（2021年）在 GemNet 架构的原子更新块中的做法。尽管作者报告使用了 $N_{\text{hidden}} = 2$ 个残差块，但我们对 $N_{\text{hidden}} = 3$进行如下设置：除 PaiNN 外，所有模型均使用 $N_{\text{hidden}} = 0$，以减少相对于 PaiNN 架构中低成本短程块的额外开销。

A.2. 训练与超参数

基线超参数，除非在此明确列出，否则与（Chanussot 等人，2021年）中基线模型所使用的值相同。与这些原始 OCP 超参数设置的所有差异，要么是为了更公平地评估埃瓦尔德修正而提升基线性能（尤其是主要由埃瓦尔德消息传递处理的能量平均绝对误差指标），

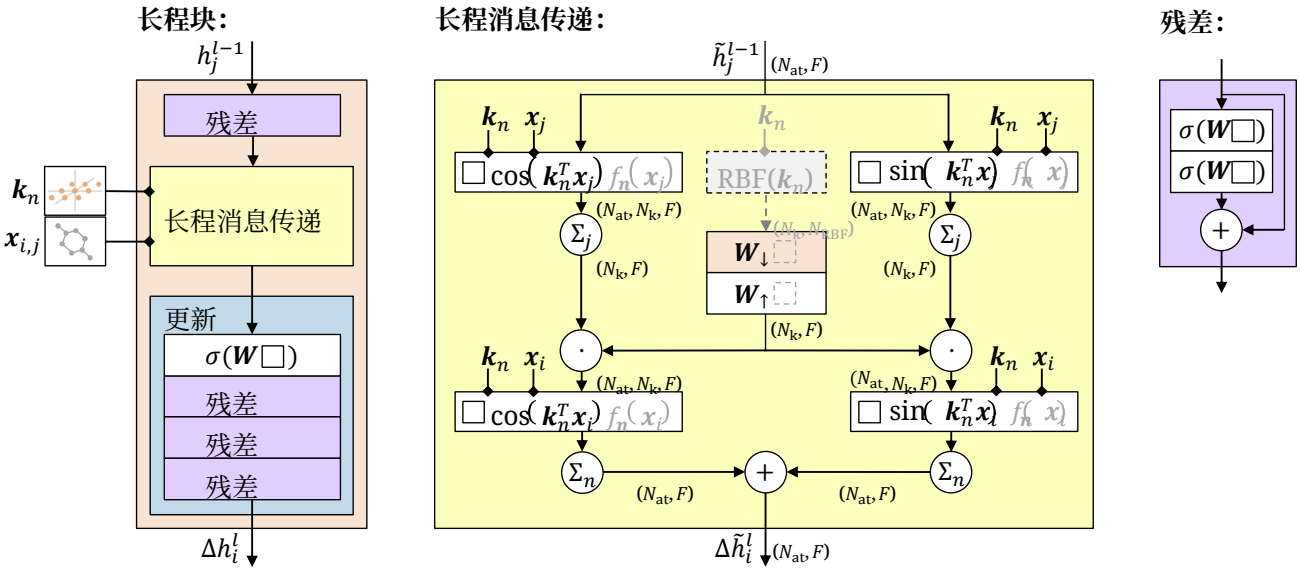


图 8. 埃瓦尔德消息传递中长程部分的架构。左侧的长程块对应表达式 $f_{\text{upd}}^{\text{lr}}\left(M_i^{\text{lr}}\right)$。□ 表示一层的输入，圆圈标出的操作按分量进行。三角形箭头表示信息在各层间的流动，菱形箭头表示参数依赖关系。方括号中标注的是计算各个部分所处理的张量形状。灰色和虚线标记的步骤仅适用于非周期情形。

要么是为了让我们的实验设置能够实现更实际的训练时间。除埃瓦尔德消息传递引入的新超参数外，基线版本和埃瓦尔德版本采用完全相同的模型设置。在 OE62 上，PaiNN 和 SchNet 模型的嵌入维度设为 512，GemNet-T 和 DimeNet^{++} 模型的嵌入维度设为 256。在 OC20 上，由于内存限制，我们不得不将 DimeNet^{++} 滤波器尺寸减小到 192。所有基线模型的距离截断值均为 6 Å，邻居阈值为 N_{max} 50。我们的 SchNet 和 PaiNN 模型使用 4 个交互块，DimeNet^{++} 和 GemNet-T 使用 3 个交互块。

OE62 上的训练超参数。我们发现，在 OE62 上，除 SchNet 外，所有基线模型使用较小的批次大小都能提升性能。因此，我们为 OE62 采用以下批次大小：SchNet 为 64，其他模型为 8。除 GemNet-T （ $1 \cdot 10^{-4}$ ）外，所有模型的初始学习率均为 $5 \cdot 10^{-4}$。与 PaiNN 的参考实现（Sch¨utt 等人，2021年）一致，我们使用 Adam优化器，权重衰减为 $\lambda = 0.01$，并搭配平台调度器（耐心值为 10，衰减因子为 0.5）。对于 DimeNet^{++} 和 GemNet-T，我们使用里程碑调度器，预热期为 250000 步（预热因子为 0.2），并在 750000、1125000 和 1500000 步时将衰减因子设为0。1。对于 SchNet，我们同样使用里程碑调度器，预热 30000 步（预热因子为 0.2），并在 60000、120000 和 180000 步时将衰减因子设为 0。1。PaiNN 训练至最大 200 轮后收敛，SchNet 和 DimeNet^{++}训练至 250 轮后收敛，GemNet-T 则训练至 300 轮后收敛。

OC20 上的训练超参数。为缩短训练时间，我们增大了 OC20 上的批次大小。GemNet-T 和 DimeNet^{++} 的批次大小为 16，PaiNN 为 48，SchNet 为 64。对于 PaiNN，我们使用耐心值为 3、因子为 0.8 的平台调度器，训练 25 个最大训练轮数。GemNet-T 使用相同的等效平台调度器。我们使用里程碑调度器训练 SchNet（预热 90000 步，预热因子为 0.2；在 150000、250000、310000 步时将里程碑因子设为 0.1），训练 25 轮；对于 DimeNet^{++}，预热 250000 步（预热因子为 0.2），并在 750000、1125000 和 1500000 步时将里程碑因子设为 0.1，训练 15 轮。在损失函数中，我们为 SchNet、PaiNN、GemNet-T 和 DimeNet^{++}分别使用 33。3、100、100 和 16。7 的力乘数。

损失函数。 与 Chanussot 等人 (2021年) 一样，我们使用能量和力平均绝对误差的线性组合。使用索引 $\mathcal{S}$ 遍历集合中的 $|\mathcal{D}|$ 原子结构 $\mathcal{D}$ (实际上即为一个小批量)，并使用i 遍历 $|\mathcal{S}|$ 结构中的单个原子 $\mathcal{S}$，我们的损失函数形式如下：

$$\begin{aligned}&\frac{1}{|\mathcal{D}|}\sum_{\mathcal{S}\in\mathcal{D}}\left|E^{(\mathcal{S})}-E_{\text{t}}^{(\mathcal{S})}\right|\\&+\frac{\lambda}{|\mathcal{D}|}\sum_{\mathcal{S}\in\mathcal{D}}\frac{1}{|\mathcal{S}|}\left\|\boldsymbol{F}_i^{(\mathcal{S})}-\boldsymbol{F}_{i;\text{t}}^{(\mathcal{S})}\right\|_p,\end{aligned}\tag{15}$$

预测/真实目标能量 $E^{(\mathcal{S})}/E^{(\mathcal{S})}$ t 、预测/目标力 $\boldsymbol{F}_i^{(\mathcal{S})}$ / $\boldsymbol{F}_i^{(\mathcal{S})}$ ；密度泛函理论、力乘数

$\lambda \in \mathbb{R}_+$ 定义能量与力精度之间所需的权衡，$\|\cdot\|_p$ 表示欧几里得 p-范数。根据模型文献，我们对 SchNet 和 DimeNet^{++}使用 $p = 1$，对 PaiNN 和 GemNet-T 使用 $p = 2$。对于仅包含能量目标的 OE62，我们设定 $\lambda = 0$。

A.3. 埃瓦尔德超参数与消融研究

非周期情形。 在非周期情形下，Ewald 消息传递引入了两个主要的新超参数：频率截断 c_k 和体素分辨率 Δ。频率滤波器的参数化方式是另一个超参数来源。在我们的实现中，需要确定高斯径向基函数的数量 N_{RBF}。为了通过 Ewald 消息传递实现帕累托高效的改进，我们根据各个短程基线图神经网络的相对计算成本调整这些超参数。主要实验设置列于表 1。

	$c_k[\,^{\circ}\text{A}^{-1}]$	$\Delta[\,^{\circ}\text{A}^{-1}]$	N_{RBF}
SchNet	0.4	0.2	48
PaiNN	0.6	0.2	128
DimeNet^{++}	0.8	0.2	128
GemNet-T	1.0	0.2	128

表 *1.* 非周期性埃瓦尔德超参数设置。

周期情形。 在周期情形中，Ewald 消息传递的主要超参数是数字 N_x, N_y, N_z，它们定义了长程求和中包含的频率。原则上，这些参数都可以独立变化，但在我们的工作中，我们通过下文说明的步骤将它们合并为一个独立的有效超参数。

请回想一下，在我们的周期滤波设置中，我们会在不同结构间保持纳入的频率数量不变；从经验上看，这样能获得更好的性能。因此，不能简单地纳入固定截断体积内的所有倒易晶格频率，因为对于倒易晶格超胞不同的输入结构，这些频率的总数会有所不同。不过，我们会确保纳入频率的范围在三个倒易晶格方向上都不会延伸到远得多的距离，这与球形截断体积的情形相同。在我们的设置中，对于任意输入结构都无法做到这一点，因为纳入的倒易晶格索引集合 I (在第3节中定义，并通过我们的超参数固定）会通过 $\boldsymbol{k}_{\boldsymbol{\lambda}'} = \lambda_1' \boldsymbol{w}_1 + \lambda_2' \boldsymbol{w}_2 + \lambda_3' \boldsymbol{w}_3$映射到不同结构的傅里叶空间中的不同点。不过，我们可以确保对于平均结构，纳入的频率沿所有三个晶格方向覆盖相同的傅里叶空间距离。例如，如果一个方向上的倒易晶格点平均而言比另一方向密集两倍，我们就将该方向的 N_i 加倍，使两个轴上由频率覆盖的平均傅里叶空间距离相同。这个单一距离实际上起到了“平均截断”的作用，同时纳入频率的数量保持对所有结构都固定。在 OC20 上，我们观察到，平均倒易空间晶胞沿 z方向（表面法线）的宽度约为沿 x 和 y 方向宽度的三分之一。应用这一超参数绑定约束，并考虑不同的基线模型运行时间后，我们得到表2所列的主要设置。

	N_x	N_y	N_z
SchNet	1	1	3
PaiNN	1	1	3
DimeNet^{++}	2	2	5
GemNet-T	2	2	5

表 *2.* 周期性Ewald超参数设置。

额外超参数。 在周期和非周期情形下，Ewald 消息传递引入的额外超参数包括瓶颈维度 $N_{\downarrow}$ 以及 N_{hidden} 作为更新函数一部分所使用的残差层数量（参见附录A.1）。

消融研究：非周期情形。 我们测试了 DimeNet^{++}-Ewald 模型在不同 c_k 和 Δ 设置下的 OE62 能量平均绝对误差，以探究我们的方法在超参数空间中的稳健性。对于 ($c_k = 0.6\ ^{\circ}\text{A}^{-1}$, $\Delta = 0.2\ ^{\circ}\text{A}^{-1}$)，我们得到的能量平均绝对误差为 46.9毫电子伏特；对于 (c_k =0.4 $^{\circ}\text{A}^{-1}$, $\Delta = 0.1\ ^{\circ-1}$A)，为 47.2 毫电子伏特；对于 ($c_k = 1.6\ ^{\circ}\text{A}^{-1}$,$\Delta = 0.4\ ^{\circ}\text{A}^{-1}$)，为 47.3 毫电子伏特；这些结果均一致优于 51.2 毫电子伏特的基线平均绝对误差。

我们还测试了 SchNet 模型的类似变体，得到的能量平均绝对误差分别为：($c_k = 0.4\ ^{\circ}\text{A}^{-1}$, Δ =0.4 $^{\circ}\text{A}^{-1}$) 时为 83.0 毫电子伏特，($c_k = 0.2\ ^{\circ}\text{A}^{-1}$, $\Delta = 0.2\ ^{\circ}\text{A}^{-1}$) 时为 85.9 毫电子伏特，以及 ($c_k = 0.8\ ^{\circ}\text{A}^{-1}, \Delta = 0.4\ ^{\circ}\text{A}^{-1}$) 时为 82.8 毫电子伏特。这些结果再次与我们的主要设置结果十分接近，并显著优于 133.5 毫电子伏特的基线平均绝对误差。

消融研究：周期情形。我们测试了 SchNet-Ewald 模型在不同设置下的 OC20 能量平均绝对误差（对全部四个测试划分取平均），考察我们的超参数绑定约束（见上文）是否是一个敏感的选择。我们得到的能量平均绝对误差分别为 836 meV（1,1,1）和 828 meV（1,1, 2）。同样，我们发现，相较于 895 meV 的基线平均绝对误差，我们的改进在不同设置下均稳健。

我们在周期性边界条件设置下不使用径向滤波器，因为从经验上看，其表现较差。我们在缩短时间的设置下训练 GemNet- T-Ewald：最多使用 30 个消息传递邻居，并在 OC20-2M 上训练 12 个 epoch。我们记录了 OC20验证集划分中四个划分上的能量平均绝对误差。我们的标准过滤方法使用 (N_x, N_y, N_z) = (2, 2, 5)，包含 137 个滤波器权重（由于第 3 节中提到的镜像对称化，

设置	基线	标准	径向	长方体
ID	305	263	264	278
OOD	455	416	447	436

表3. OC20 上过滤变体的能量平均绝对误差 [毫电子伏特]。

设置	SchNet	PaiNN	DimeNet^{++}	GemNet-T
埃瓦尔德 1220万	1570万	4.8M		16.1M
嵌入 14.4M	1570万	5.4M		15.8M

表 4。Ewald 模型和增大嵌入变体在 OE62 上的参数数量（单位：百万参数）。

）。我们将其与一种使用径向滤波的埃瓦尔德变体进行比较，该变体采用 128 个高斯径向基函数。此外，我们还测试了另一种过滤方法，它使用协方差矩阵为 $\Sigma = \mathrm{diag}(\Delta_x^2, \Delta_y^2, \Delta_z^2)$、中心位于长方体晶格 $\Lambda_c' = \{\lambda_x' \Delta_x (1,0,0)^T + \lambda_y' \Delta_y (0,1,0)^T + \lambda_z' \Delta_z (0,0,1)^T \mid \boldsymbol{\lambda}' \in I\}$各点上的三维高斯基函数 $\mathcal{G}(\Sigma, \mu_i)$，其中 $I = \times_{j=x,y,z} \{-N_j, \ldots, N_j\} \subset \mathbb{Z}^3$。我们测试了 $(N_x, N_y, N_z) = (3,3,6)$ 和 $(\Delta_x, \Delta_y, \Delta_z) = (0.333, 0.333, 0.167)$ $\mathrm{\AA}^{-1}$。如上所示，所有变体都优于基线，但径向和长方体过滤变体在 OOD-both 划分上的泛化能力较差；在该划分中，催化剂和吸附物的组成均为分布外（表 3）。

A.4. 比较研究的超参数

增大截断设置。对于 OC20 中空间上无限的结构，我们假设给定截断下的消息传递邻居数量大致满足 $N \sim c_x^3$，并据此进行相应缩放。对于所有基线，给定起始值 $c_x = 6$ $\mathrm{\AA}$ 和 $N_{\max} = 50$，我们发现，修改后的 OC20 设置——GemNet-T 使用 (6.25 $\mathrm{\AA}$，55)，DimeNet^{++}、PaiNN 使用 (7 $\mathrm{\AA}$，80)，SchNet 使用 (8.5 $\mathrm{\AA}$，140)—— 其成本大致与相应的 Ewald 模型变体相当。在 OE62 上，我们将邻居阈值从 50 加倍至 100（超过大多数结构的总规模），GemNet-T 使用增大的 9 $\mathrm{\AA}$ 截断，其他所有基线使用 12 $\mathrm{\AA}$ 截断。在 OE62 上，我们将邻居阈值从 50 加倍至 100（超过大多数结构的总规模），GemNet-T 使用增大的8 $\mathrm{\AA}$ 截断，其他所有基线使用 12 $\mathrm{\AA}$ 截断。

增大嵌入尺寸设置。在 OE62 增大嵌入尺寸研究中，我们将 SchNet 的尺寸增至 1536，PaiNN 增至 576，DimeNet 增至 384^{++}，GemNet-T 增至 320，以大致匹配埃瓦尔德变体的自由参数数量（表 4）。

成对长程块设置。我们将 Ewald 模块（$f_{\text{upd}}^{\text{lr}}(M_{\text{lr}_i})$ 见公式（12））替换为一个标准SchNet块，其中包含200个高斯函数和256个滤波器。在 OE62 上，当 12 $\mathrm{\AA}$ 且最大邻居数为100时，计算成本趋于饱和，此时分子图接近完全连通。在 OC20 上，长程SchNet块无法与 Ewald计算成本相抗衡：对于 PaiNN，在 6.25 $\mathrm{\AA}$ 时 Ewald开销已与之持平；对于 SchNet 和 DimeNet++，这一点出现在 6.75 $\mathrm{\AA}$；对于 GemNet-T，则出现在 7.25 $\mathrm{\AA}$。此处我们按照增加距离截断时的做法提高邻居阈值。在 OC20 上，增大距离截断的成本更高，这是因为其原子密度更大。

A.5. $\hat{M}_{\boldsymbol{k}}^{\text{lr}}$ 与 $\hat{\Phi}^{\text{lr}}(\|\cdot\|)$之间的关系

为推导出所需的恒等式（与标准埃瓦尔德求和中的做法相同），我们从傅里叶系数 $\hat{M}^{\text{lr}}\boldsymbol{k}$ 开始，

$$\hat{M}_{\boldsymbol{k}}^{\text{lr}} = \frac{1}{\Omega} \int_{\Omega(\mathcal{S})} M^{\text{lr}}(\boldsymbol{x}) \exp(-i\boldsymbol{k}^T\boldsymbol{x}) d^3\boldsymbol{x}, \tag{16}$$

其中 $\Omega(\mathcal{S}) \subset \mathbb{R}^3$ 表示一个超胞平行六面体。现在将式（8）代入 $M^{\text{lr}}(\boldsymbol{x})$，得到

$$\hat{M}_{\boldsymbol{k}}^{\text{lr}} = \frac{1}{\Omega} \int_{\Omega(\mathcal{S})} \sum_{\boldsymbol{t}\in\Lambda} \sum_{j\in\mathcal{S}} h_j \Phi^{\text{lr}}(\|\boldsymbol{x} - \boldsymbol{x}_j - \boldsymbol{t}\|) \cdot \exp(-i\boldsymbol{k}^T\boldsymbol{x}) d^3\boldsymbol{x}. \tag{17}$$

假设 关于 Λ 的晶格求和绝对收敛，我们可以将其与超胞积分交换，从而得到

$$\begin{aligned} &\int_{\Omega(\mathcal{S})} \sum_{\boldsymbol{t}\in\Lambda} \Phi^{\text{lr}}(\|\boldsymbol{y}_j - \boldsymbol{t}\|) \exp(-i\boldsymbol{k}^T\boldsymbol{y}_j) d^3\boldsymbol{y}_j \\ &= \sum_{\boldsymbol{t}\in\Lambda} \int_{\Omega(\mathcal{S})} \Phi^{\text{lr}}(\|\boldsymbol{y}_j - \boldsymbol{t}\|) \exp(-i\boldsymbol{k}^T\boldsymbol{y}_j) d^3\boldsymbol{y}_j \\ &= \int_{\mathbb{R}^3} \Phi^{\text{lr}}(\|\boldsymbol{y}_j\|) \exp(-i\boldsymbol{k}^T\boldsymbol{y}_j) d^3\boldsymbol{y}_j \end{aligned} \tag{18}$$

成立（因为超胞构成了 $\mathbb{R}^3$的镶嵌），其中我们令 $\boldsymbol{y}_j := \boldsymbol{x} - \boldsymbol{x}_j$。现在将所得的关于 $\mathbb{R}^3$ 的积分识别为傅里叶变换 $\hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|)$：

$$\begin{aligned} \hat{M}_{\boldsymbol{k}}^{\text{lr}} &= \frac{1}{\Omega} \sum_{j\in\mathcal{S}} h_j \left[\int_{\mathbb{R}^3} \Phi^{\text{lr}}(\|\boldsymbol{y}_j\|) \exp(-i\boldsymbol{k}^T\boldsymbol{y}_j) d^3\boldsymbol{y}_j \right] \\ &\cdot \exp(-i\boldsymbol{k}^T\boldsymbol{x}_j) = \frac{1}{\Omega} \sum_{j\in\mathcal{S}} h_j \exp(-i\boldsymbol{k}^T\boldsymbol{x}_j) \hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|). \end{aligned} \tag{19}$$

A.6. 结构因子离散化

离散化不存在唯一的方法，我们相信，我们的应用为未来采用更复杂的方案留下了空间，例如基于小波理论的方法（Mallat，2009）。为推动此类工作，我们首先给出一个用于离散化结构因子的相当通用的框架。随后，我们概述具体的实现方式，并说明它如何作为一个特例由该框架导出。设有一个有限集合，其中包含 N_{k} 平方可积的复值基函数 $\{f_i \in L^2(D) \mid 1 \le i \le N_{\text{k}}\}$，定义在一个有界域 D

该有界域包含截断域 $\mathcal{B}_c = \{\boldsymbol{k} \in \mathbb{R}^3 \mid k \le c_k\}$，即 $\mathcal{B}_c \subset D \subset \mathbb{R}^3$。进一步假设其具有正交归一性$\langle f_j, f_k\rangle_{L^2} = \int_D f_j(\boldsymbol{k})^* f_k(\boldsymbol{k})\, d^3\boldsymbol{k} = \delta_{jk}$，其中 * 表示复共轭。结构因子 $s(\boldsymbol{k})$ 可以投影到该基组上，从而得到一个离散的近似表示（式（20））， 其形式由基组系数$\{s_i \mid 1 \le i \le N_\mathrm{k}\}$给出:

$$s(\boldsymbol{k}) \approx \sum_{i=1}^{N_\mathrm{k}} \langle f_i, s\rangle f_i(\boldsymbol{k}) =: \sum_{i=1}^{N_\mathrm{k}} s_i f_i(\boldsymbol{k}) \tag{20}$$

对这些标量积进行数值积分，需要知道 $s(\boldsymbol{k})$ 的高分辨率表示。相反，式（21）利用了其中出现的傅里叶变换:

$$\begin{aligned} s_i = \langle f_i, s\rangle &= \sum_{b\in\mathcal{S}} h_b \int_D f_i(\boldsymbol{k})^* \exp(-i\boldsymbol{k}\boldsymbol{x}_b) d^3\boldsymbol{k} \\ &= \sum_{b\in\mathcal{S}} h_b \widehat{[f_i^*]}(\boldsymbol{x}_b) = \sum_{b\in\mathcal{S}} h_b \hat{f}_i^*(-\boldsymbol{x}_b). \end{aligned} \tag{21}$$

$\hat{}$ 注意这与 $s_{\boldsymbol{k}}$ 类似，只是用 $f_i^*(-\boldsymbol{x}_b)$ 替代了 $\exp(-i\boldsymbol{k}\boldsymbol{x}_b)$。我们用两个例子对此进行说明。

(a) 辅助超胞。 文献中的一种标准方法（Freysoldt 等人, 2022年）是将非周期结构近似为超胞的一部分，并假设该超胞是有限的，但足够大，从而可以忽略虚假的相互作用。这样便重新引入了倒易晶格 Λ'，在其上离散的$s_{\boldsymbol{k}_i} := s(\boldsymbol{k}_i)$ 可以进行计算。由于狄拉克 δ 分布的定义性质，$\langle \delta(\cdot - \boldsymbol{k}_i) s\rangle = s(\boldsymbol{k}_i)$，这属于我们形式体系的一个特殊（即极限）情形，其特征是采用分布形式的"基"和傅里叶变换,

$$\begin{aligned} &\{f_i^{\mathrm{SC}}(\boldsymbol{k}) = \delta(\boldsymbol{k} - \boldsymbol{k}_i) \mid \boldsymbol{k}_i \in \Lambda' \cap \mathcal{B}_c\}, \\ &\hat{f}_i^{\mathrm{SC}*}(-\boldsymbol{x}_b) = \exp(-i\boldsymbol{k}_i\boldsymbol{x}_b), \end{aligned} \tag{22}$$

如上面的公式（22）所示，系数s_i 在公式（21）中的定义确实在这一特殊情况下化简为 $s_{\boldsymbol{k}}$ 的 PBC 定义。

(b) 体素基。 在我们对非周期埃瓦尔德消息传递的实现中，我们通过对三维体素网格 $\Lambda'_\Delta = \{\lambda'_1\Delta(1,0,0)^T + \lambda'_2\Delta(0,1,0)^T + \lambda'_3\Delta(0,0,1)^T \mid \lambda' \in \mathbb{Z}^3\}$ 中覆盖截断范围的部分 $\Lambda'_\Delta \cap \mathcal{B}_c$ 取体素平均值 s_i，对 $s(\boldsymbol{k})$ 进行离散化 s_o。体素边长为 Δ。元素 $\boldsymbol{k}_i \in \Lambda'_\Delta \cap \mathcal{B}_c$ 是体素中心位置。在我们的通用框架中，这对应于一个基组

$$\begin{aligned} &\{f_i^{\mathrm{V}} = \Delta^{-3}\mathcal{X}_{\boldsymbol{k}_i} \mid \boldsymbol{k}_i \in \Lambda'_\Delta \cap \mathcal{B}_c\}, \\ &\hat{f}_i^{\mathrm{V}*}(-\boldsymbol{x}_b) = \exp(-i\boldsymbol{k}_i\boldsymbol{x}_b) \prod_{j\in\{xyz\}} \mathrm{sinc}\left(\frac{k_i^j x_b^j \Delta}{2}\right), \end{aligned} \tag{23}$$

其中， $\mathcal{X}_{\boldsymbol{k}_i}$ 在以 $\boldsymbol{k}_i$为中心的立方体 $\prod_{j\in\{x,y,z\}}\left[k_i^j - \frac{\Delta}{2}, k_i^j + \frac{\Delta}{2}\right]$ 内取值为 1，其他位置则为 0。超参数 Δ 控制离散化分辨率。取标量积,

$$\langle f_i^{\mathrm{V}}, s\rangle = \Delta^{-3} \int_{k_i^x-\frac{\Delta}{2}}^{k_i^x+\frac{\Delta}{2}} \int_{k_i^y-\frac{\Delta}{2}}^{k_i^y+\frac{\Delta}{2}} \int_{k_i^z-\frac{\Delta}{2}}^{k_i^z+\frac{\Delta}{2}} s(\boldsymbol{k}) d^3\boldsymbol{k}, \tag{24}$$

计算 $s(\boldsymbol{k})$ 在 $\boldsymbol{k}_i$周围体素立方体内的平均值。利用公式（23）中的傅里叶变换结果，我们看到 s_i 与 PBC 情况的唯一区别是 $s_{\boldsymbol{k}}$ 多了一个 sinc 阻尼函数。令 $\Delta \to 0$，即可恢复连续极限：一致地，$\mathrm{sinc}\left(\frac{x_b^j\Delta}{2}\right)$ 在该极限下逐点收敛到1 。

到目前为止，我们已经概述了如何通过离散基系数 s_i 表示结构因子。为了在体素实现中计算长程消息，我们按如下方式进行。连续滤波器 $\hat{\Phi}(k)$ 采用相同类型的体素表示 $\sum_{i=1}^{N_\mathrm{k}} \hat{\Phi}_i f_i^{\mathrm{V}}$，其形式由系数 $\hat{\Phi}_i = \langle f_i^{\mathrm{V}}, \hat{\Phi}\rangle$ 给出。这些系数通过在离散网格的体素中心点 $\Lambda'_\Delta \cap \mathcal{B}_c$ 上读取所学习的径向基组合 $\hat{\Phi}(k) = \boldsymbol{W}\hat{\Psi}(k)$ 获得。现在，消息求和的完整表达式可近似写成

$$\begin{aligned} M_a^{\mathrm{lr}} &= \int_D \exp(i\boldsymbol{k}\boldsymbol{x}_a) \left[\sum_{i=1}^{N_\mathrm{k}} s_i \hat{\Phi}_i f_i^{\mathrm{V}}(\boldsymbol{k})\right] d^3\boldsymbol{k} \\ &= \sum_{i=1}^{N_\mathrm{k}} \hat{f}_i^{\mathrm{V}}(-\boldsymbol{x}_a) \left[\sum_b h_b \hat{f}_i^{\mathrm{V}*}(-\boldsymbol{x}_b)\right] \hat{\Phi}_i, \end{aligned} \tag{25}$$

其中在最后一步，我们代入了s_i 的表达式，该表达式来自方程（21）。请注意，这与周期性边界条件消息传递公式方程（10）类似，只是其中的$\exp(i\boldsymbol{k}\boldsymbol{x}_a)$ 和 $\exp(-i\boldsymbol{k}\boldsymbol{x}_b)$分别替换为 $f_i^{\mathrm{V}}(-\boldsymbol{x}_a)$ 和 $f_i^{\mathrm{V}*}(-\boldsymbol{x}_b)$。由于二者的区别仅在于体素情况下额外的 sinc 衰减因子，因此我们的埃瓦尔德消息传递周期性实现和非周期性实现所用代码仅有细微差别。

A.7 力预测的影响

埃瓦尔德消息传递主要是一种用于修正能量而非力目标的方案。原子 i 上的力是能量梯度 $\boldsymbol{F}(\boldsymbol{x}_i) = -\nabla_i E(\boldsymbol{x}_1, \ldots \boldsymbol{x}_N)$ 的负值。由于式（10）中采用了波长截断，埃瓦尔德消息传递只能为能量预测贡献在空间中的变化速度慢于其逆频率截止尺度的项。为了定量地说明这一点，考虑函数 $V_i(\boldsymbol{x})$ 的傅里叶级数拟设：$= E(\boldsymbol{x}_1, \ldots \boldsymbol{x}_{i-1}, \boldsymbol{x}, \boldsymbol{x}_{i+1}, \boldsymbol{x}_N)|_{\boldsymbol{x}_1, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \boldsymbol{x}_N}$，其中 $V_i(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\Lambda'} \hat{V}_{\boldsymbol{k}} \exp(i\boldsymbol{k}^T\boldsymbol{x})$。在这些项中，

$$\boldsymbol{F}(\boldsymbol{x}_i) = -\nabla V(\boldsymbol{x}_i) = -\sum_{\boldsymbol{k}\in\Lambda'} i\boldsymbol{k}\hat{V}_{\boldsymbol{k}} \exp(i\boldsymbol{k}^T\boldsymbol{x}_i). \tag{26}$$

因此，与高频的梯度贡献相比，低频 $\|\boldsymbol{k}\| \leq c_k$ 处的贡献受到抑制。因此，我们假设模型的短程消息和承担了力预测中的主要部分。

A.8. 运行时间测量

我们在 Nvidia A100 GPU 上运行所有模型，并在同一台机器、同一场会话中评估所有模型配置的运行时间。经过 50 个预热批次后，我们对 500 个批次中的每个结构的平均运行时间进行测量，并将该测量重复三次。随后，我们取这三次测量结果中的最小值。

A.9. 进一步的长程分箱结果

作为对本文主体分析的补充，我们提供了类似于图 7 的额外分箱结果。这些结果进一步揭示了各种结构特征与 Ewald 改进之间的相关性。

进一步的 OE62 分析。 我们根据两个结构尺寸指标对结构进行分组：任意一对原子之间的最大距离，以及原子点云沿主轴的标准差 σ。此外，我们还根据公式（6）中的静电能进行分箱；该静电能使用 OE62 数据中的 DFT-Hirshfeld 原子部分电荷计算得到。

进一步的 OC20 分析。 由于周期性边界条件，在 OC20 中不存在明显的尺寸指标。不过，对于 OC20 的 IS2RE-val 划分（其中所有结构都提供了 Bader 部分电荷），我们仍然可以像对 OE62 那样根据静电能进行分箱。

讨论。 我们的补充分析得出了以下结论和使用建议：

- Ewald 消息传递似乎能够处理色散效应。我们希望，我们的方法也能为成本层级中超越 DFT-D3 的参考目标学习这类效应，例如通过非局域 DFT 泛函获得的能量。

- 我们没有观察到静电能越高，相对改进越大的趋势；在所有分箱中，修正幅度都相同。显然，标准图神经网络已经能够在我们的数据上较好地学习静电学。在界面化学或胶体化学等领域的后续基准测试中重复这项研究，或许会很有价值。

- 由于具有良好的扩展性，我们预计 Ewald 消息传递的实用性会随着结构尺寸的增大而提高。然而，我们的分箱图表明，结构尺寸并不能直接代表预期的模型误差或 Ewald 改进程度。

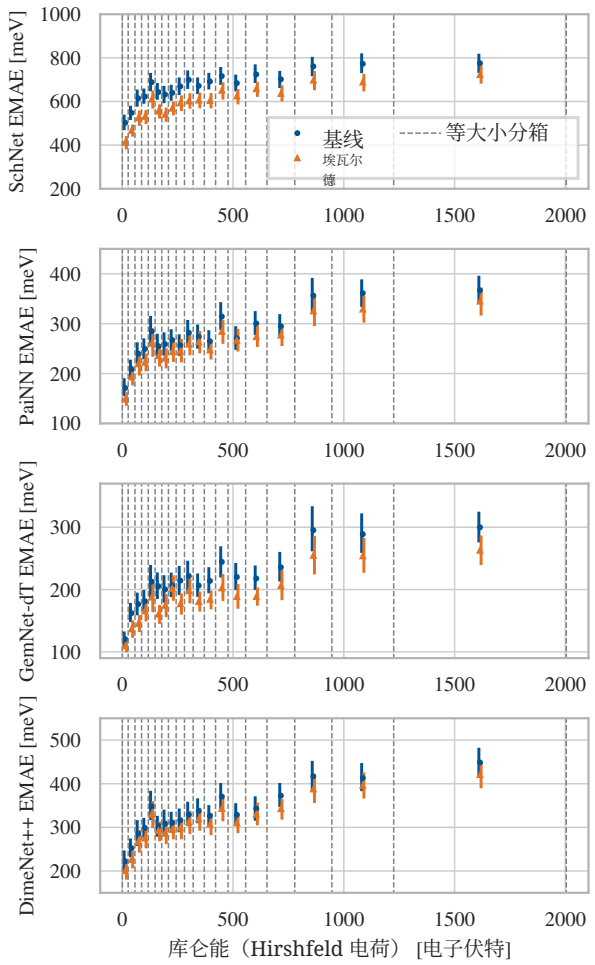


图 9. 按静电能分箱的 OC20-IS2RE-val-id 结构的能量平均绝对误差。对于静电能贡献较高的结构，Ewald 消息传递的影响并不显著。

A.10. OE62 数据集预处理

由于原始 OE62 DFT 能量相对于零值存在较大的平均偏移，我们对所有结构计算了形式为 $\mathcal{S}\ E^{(\mathcal{S})}{}_{\mathrm{t}} = E^{(\mathcal{S})}_{\mathrm{raw}} - \sum_Z C_Z N_{(\mathcal{S})_Z} - C_0$ 的最小二乘拟合目标值，其中 $E^{(\mathcal{S})}{}_{\mathrm{t}}/E^{(\mathcal{S})}_{\mathrm{raw}}$ 分别表示预处理后的目标值 / 原始 DFT 能量，$Z \in \mathbb{N}$ 枚举原子类型，并为每种类型指定回归系数 $C_Z \in \mathbb{R}$ 和偏置系数 C_0，$N_Z^{(\mathcal{S})} \in \mathbb{N}_0$ 表示结构中属于类型 Z 的原子数量 $\mathcal{S}$。模型只学习该拟合的残差；按照构造，这些残差的均值接近零。我们将数据划分为约 50000 个结构作为 OE62-train，并分别划分约 6000 个结构作为 OE62-val 和 OE62-test。

A.11. 能量、力和运行时间数值结果

表。5和6给出了 OC20 和 OE62 的数值结果，我们在第5节中基于这些结果进行可视化。

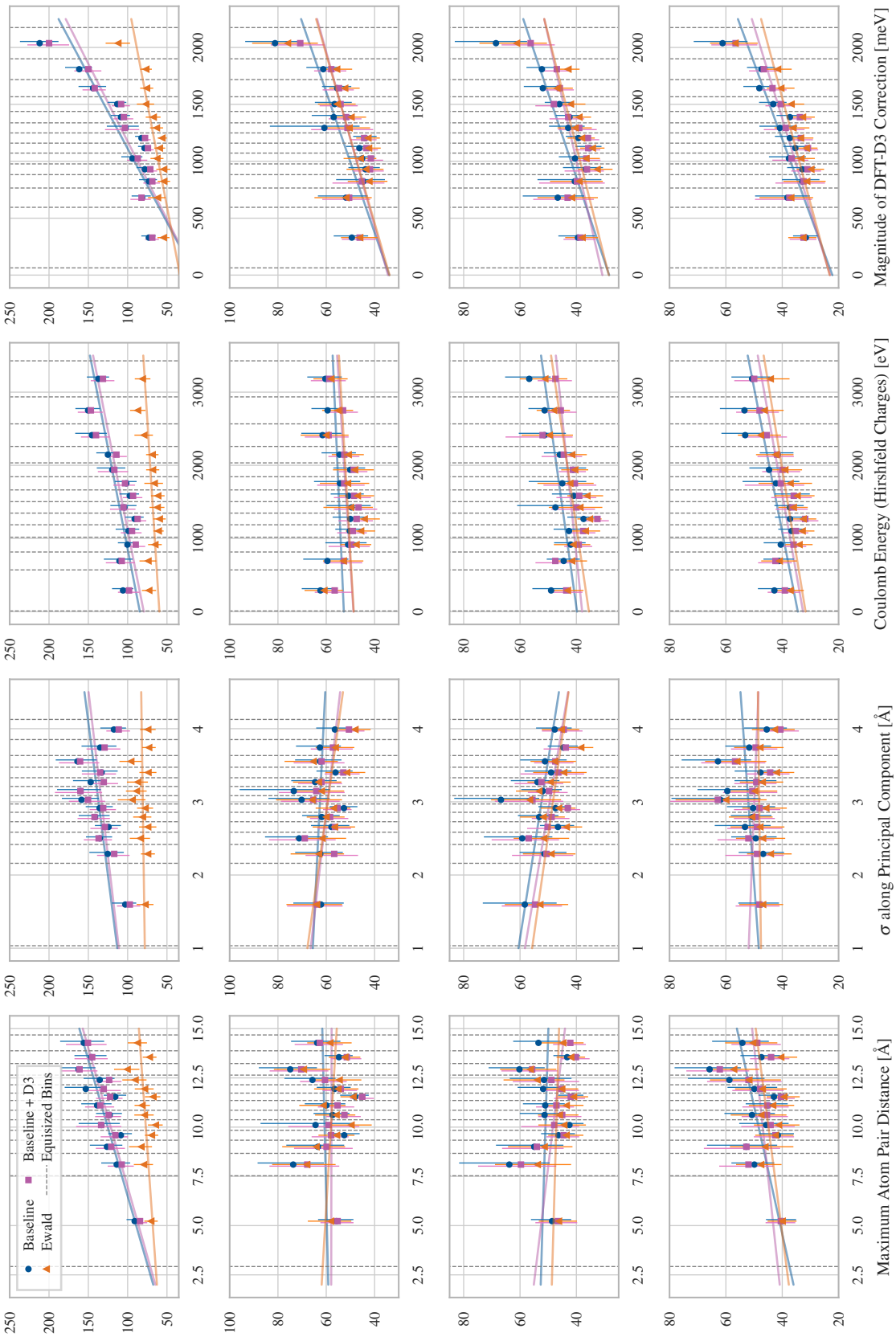


Figure 10. Binning of the energy EMAE on `OE62-val` by two structural size metrics, electrostatic and dispersion energy. Only dispersion is a good predictor for Ewald improvement.

表 5. OE62 上每个输入结构的能量平均绝对误差和运行时间。

		OE62-val		OE62-test		前向传播		前向与反向传播	
模型	变体	MAE 毫电子伏特 ↓	Rel. % ↑	MAE 毫电子伏特 ↓	Rel. % ↑	运行时间 ms/结构。↓	Rel. % ↓	运行时间 ms/结构。↓	Rel. % ↓
SchNet	基线	133.5	-	131.3	-	0.13	-	0.28	-
	嵌入	144.7	-8.4	136.7	-4.1	0.14	15.2	0.33	17.8
	截断	257.4	-92.8	254.8	-94.1	0.14	13.6	0.31	11.6
	SchNet-LR	86.6	35.1	89.2	32.1	0.32	156.0	0.75	171.7
	埃瓦尔德	**79.2**	**40.7**	**81.1**	**38.2**	0.70	461.6	1.03	271.4
PaiNN	基线	61.4	-	63	-	1.52	-	3.16	-
	嵌入	63.5	-3.4	63.1	-0.2	1.54	1.4	3.28	3.8
	截断	65.1	-6.0	64.4	-2.2	1.84	20.9	3.91	23.6
	SchNet-LR	58.3	5.1	58.2	7.7	1.84	20.7	4.21	33.1
	埃瓦尔德	**57.9**	**5.7**	**59.7**	**5.2**	2.29	50.5	4.57	44.4
DimeNet^{++}	基线	51.2	-	53.8	-	1.99	-	4.26	-
	嵌入	50.4	1.6	53.4	0.7	2.25	12.9	4.93	15.8
	截断	48.3	5.7	**48.1**	**10.6**	2.68	34.7	6.10	43.4
	SchNet-LR	51.4	-0.5	54.4	-1.1	2.37	19.0	4.73	11.2
	埃瓦尔德	**46.5**	**9.2**	**48.1**	**10.6**	2.70	35.5	5.93	39.5
GemNet-T	基线	51.5	-	53.1	-	3.07	-	6.96	-
	嵌入	52.7	-2.3	53.9	-1.5	3.11	1.5	6.98	0.4
	截断	47.8	7.2	47.7	10.2	4.02	31.2	8.88	27.7
	SchNet-LR	51.2	0.6	52.8	0.5	3.32	8.3	7.73	11.1
	埃瓦尔德	**47.4**	**8.0**	**47.5**	**10.5**	4.05	32.0	8.86	27.4

表 6。能量和力的平　　均绝对误差，以及每个输入结构的运行时间，在 OC20 测试集的所有四个划分上取平均　　.

		能量		力		前向传播		前向与反向传播	
模型	变体	平均绝对误差 毫电子伏特 ↓	相对误差 . % ↑	MAE $Å^{-1}$ ↓	Rel. % ↑	运行时间 毫秒/结构。↓	Rel. % ↓	运行时间 毫秒/结构。↓	Rel. % ↓
SchNet	基线	895	-	61.1	-	0.85	-	1.83	-
	截断	869	3.0	60.3	1.3	1.57	84.2	3.92	114.2
	SchNet-LR	984	-9.9	65.3	-6.8	1.79	110.6	5.24	186.0
	埃瓦尔德	**830**	**7.3**	**56.7**	**7.2**	1.79	110.7	4.68	155.3
PaiNN	基线	448	-	42.0	-	1.66	-	3.44	-
	截断	413	7.9	40.9	2.6	2.54	53.1	5.06	46.9
	SchNet-LR	468	-4.3	43.3	-3.2	2.06	24.2	5.24	44.3
	埃瓦尔德	**393**	**12.3**	**40.5**	**3.5**	2.08	25.4	4.24	23.1
DimeNet^{++}	基线	496	-	47.1	-	5.58	-	15.68	-
	截断	487	1.8	46.4	1.4	6.28	12.4	17.23	9.9
	SchNet-LR	504	-1.6	48.3	-2.6	6.26	12.1	16.14	2.9
	埃瓦尔德	**445**	**10.4**	**45.5**	**3.3**	6.24	11.7	17.00	8.4
GemNet-dT	基线	346	-	29.9	-	6.37	-	14.77	-
	截断	352	-1.7	30.2	-1.0	7.23	13.5	16.40	11.0
	SchNet-LR	351	-1.5	29.8	0.3	7.19	12.1	16.35	10.7
	埃瓦尔德	**307**	**11.3**	**29.3**	**1.8**	7.09	11.2	16.23	9.9