

周期性材料的多尺度拓扑-量子-图论融合描述符：基于单纯复形与第一性原理的普适性表征框架

摘要 (Abstract)

传统材料描述符在精确量化周期性材料中普遍存在的结构畸变、连续对称性破缺及多体相互作用等复杂物理化学现象时存在显著局限，难以有效解码其深层的结构-性能关系，从而制约了材料科学的理性设计与高通量发现。为突破此瓶颈，本研究提出并构建了一个多尺度拓扑-量子-图论融合描述符框架。该框架将代数拓扑的单纯复形理论、第一性原理计算以及图论路径分析进行深度融合，通过对0-单纯形（原子）、1-单纯形（化学键）、2-单纯形（三体相互作用）直至全局（Global）层次进行系统性表征，并经由统计聚合生成一个固定长度、信息完备且高度可解释的特征向量（材料指纹），实现了对周期性材料基因组的全面数字化表征。

本框架的核心创新体现在三个层面。首先，在微观机制的量化上，本研究首次将李代数理论应用于周期性材料的结构表征，通过`lie_algebra_incompatibility`等特征，实现了对结构畸变引起的对称性破缺及局域应力的精确量化；同时，引入辛几何思想，定义了`log_pseudo_symplectic_fluctuation_volume`等描述符，旨在量化晶格中结构自由度与电子自由度间的内在耦合强度。其次，在多尺度体系的构建上，通过对各单纯形层次局部特征（如`bader_charge`, `bond_distance`）进行统计聚合（如均值、方差），确保了最终全局特征向量的维度固定与信息完整性，使其作为连续可微的数学映射与梯度优化算法高度兼容。最后，为表征非局域关联，本文提出一种基于多物理场信息的统一重要性采样算法，用以识别晶体拓扑图中的关键路径，并沿路径定义了一系列非局域关联函数（如`path_curvature_mean`），为理解长程有序及集体动力学行为提供了量化途径。

本框架的有效性已通过对 CsPbX_3 ($\text{X} = \text{Br}, \text{I}, \text{Cl}$) 钙钛矿这类周期性晶体材料体系的严格案例分析得到验证。其中，对于初始标记为立方 Pm-3m 相的 CsPbCl_3 ，本框架的描述符给出了显著的非零不对称性特征值，与预期不符。这一反常结果促使对输入结构进行了复核，并最终揭示其真实结构为正交 Amm2 相。此发现有力证明了本框架不仅能准确量化已知结构，更能作为一种高灵敏度的结构探测器，识别输入数据中的潜在不一致性或错误标签，展现了其作为数据验证与科学发现工具的独特价值。

综上，本研究的核心贡献在于建立了一套高度可解释、蕴含物理洞察力、且适用于人工智能驱动材料发现的通用描述符框架。这些描述符不仅能显著加速高通量虚拟筛选流程，实现对材料性能的快速预测，

更为关键的是，它们为AI驱动的逆向设计提供了信息丰富且高效的知识驱动搜索空间，有望推动材料科学研究从传统的试错式探索向精确设计、智能发现的模式转变。

第一章：引言——周期性材料复杂性挑战与多尺度描述符范式革新

1.1 背景与问题阐述

周期性材料，作为从无机晶体、金属有机框架（MOFs）、共价有机框架（COFs）到高分子聚合物等多元化体系的统称，是现代科技发展的基石。其宏观功能特性（如导电性、力学强度、热稳定性、光电效应及催化活性）深刻根植于其复杂的微观结构与电子排布。然而，传统材料描述符，如晶格常数、原子类型或简单的局域配位数，在捕捉周期性材料中普遍存在的复杂非局域物理化学现象、由结构畸变引起的连续对称性破缺以及多体相互作用导致的涌现行为时，表现出显著的局限性。这些局限性使得现有方法难以有效解码决定材料性能的深层基因型，从而制约了基于第一性原理计算与高通量筛选的材料理性设计效率。当前，亟需一种能够深刻量化微观细节、有效桥接多尺度信息，并能精准捕获复杂物理化学现象的新型材料描述符体系，以加速功能材料的发现与创新。

1.2 提出的解决方案：一个基于代数拓扑单纯复形的周期性材料描述符框架

为应对上述挑战，本研究提出并构建了一个基于代数拓扑单纯复形的全新周期性材料描述符框架。该框架旨在系统性解码材料基因组，深度融合数学、物理、化学、材料学及计算机科学等多学科前沿理论，实现对周期性材料多尺度、多维度属性的全面、精准且物理真实的量化。解决方案的核心体现在以下几个方面：

首先，在微观机制的量化方面，首次将李代数与辛几何概念系统性地应用于周期性材料表征。通过 $\text{lie_algebra_incompatibility}$ （1-单纯形）和 $\text{hierarchical_stress_flow}$ （2-单纯形）等特征，实现了对结构畸变导致的对称性破缺及局域应力流的精确量化。同时，借助 $\text{log_pseudo_symplectic_fluctuation_volume}$ （全局）等描述符，能够量化并解析材料中结构自由度与电子自由度间的内在耦合强度，为揭示内应力流形与能量传输提供创新的理论工具。

其次，在多尺度可微描述符体系的构建方面，建立了自下而上的层次化特征工程体系，实现了从原子到多体相互作用的连续可微表征。该体系涵盖0-单纯形（原子中心环境）、1-单纯形（化学键）、2-单纯形（三体相互作用）描述符，以及反映材料整体属性的全局描述符。通过对大量局部特征（如 `bader_charge`、`bond_distance`）进行统计聚合（如均值、方差、极值等），最终形成维度固定、信息完备的全局特征向量，确保信息的完整性、层级关联性及其与梯度优化算法的高度兼容性。

最后，在非局域关联解码与拓扑信息捕捉方面，提出基于多物理场信息驱动的统一重要性采样算法，以路径熵、电负性差及最大扭转应力为核心因子，高效搜寻周期性材料拓扑图中的关键路径。沿这些路径，定义并计算了一系列非局域关联函数（如 `path_curvature_mean`、`path_wrapping_norm_mean`、`path_structure_autocorr_mean`），有效解析了材料内部长程有序、声子传播以及电荷输运等集体动力学行为，为理解材料的宏观性能提供了微观关联基础。

1.3 核心论点：严谨验证与发现潜力—— CsPbCl_3 的启示

本研究的核心论点在于：我们所构建的多尺度描述符框架不仅能够准确地捕获周期性材料的结构和电子信息，并且其卓越的敏感性和严谨的物理数学基础使其能够超越常规表征，成为揭示数据中隐藏物理真实的强大分析工具。

我们将通过对 CsPbX_3 ($X = \text{Br}, \text{I}, \text{Cl}$) 钙钛矿体系的严格比较分析来验证这一论点。其中， CsPbBr_3 (Pm-3m) 作为高度对称的参考基准，其不对称性相关特征值应为零； CsPbI_3 (Pnma) 作为已知扭曲的案例，其特征值应能定量反映其畸变程度。而 CsPbCl_3 的案例将作为本研究的关键验证。该材料起初被标注为立方 Pm-3m 相，但我们的描述符却计算出了显著的非零不对称性值。这一“异常”并非算法错误，而是精确揭示了其真实结构为正交 Amm2 相。这一发现强有力地证明了本框架不仅能正确描述已知结构，更能作为一种高灵敏度的结构探测器，自主发现输入数据中的不一致性或错误标签，从而验证了其设计的优越性和在材料科学发现中的巨大潜力。因此，本研究不仅是描述符的构建，更是对其严谨验证过程及其在科学发现中所展现能力的系统性论证。

1.4 学术定位：连接不同范式的桥梁

1.4.1 个人动机与研究背景

本研究的深层动机源于对周期性量子系统中深层对称性原理的持续探索，以及如何将其定量化应用于材料功能设计的实践需求。在前期对钙钛矿材料合成与表征的实践经验中，作者深刻体会到传统材料描述

符在捕捉诸如八面体倾斜、键合畸变等微观结构细节方面的局限性。这些看似微小的结构变化，却对材料的宏观物理化学性能（如光电性质、铁电行为）产生决定性影响。

与此同时，对抽象代数理论，特别是李代数和辛几何如何描述物理现象（例如周期性材料中的声子振动模态或对称性破缺导致的几何畸变）的浓厚兴趣，进一步启发了本研究的理论方向。尤其值得提及的是，在接触基于图神经网络和Transformer架构的端到端模型（如QCFormer）中涉及的商代数概念时，其在处理复杂数据结构中的强大能力令人印象深刻，但同时也引发了对这些“黑箱”模型缺乏物理可解释性的思考。

这种个人化的思考路径，结合对实验现象的深入洞察与对前沿数学工具的理论探索，最终汇聚成本研究的核心目标：即开发一套能够超越传统描述范式，实现对材料内在“几何基因型”进行更精细、更可解释、且具有深层物理意义的量化表征框架。本工作旨在为材料科学提供一个“白箱”工具，不仅能够精确描述已知结构，更能自主揭示数据中隐藏的物理真理，从而加速新材料的发现与设计。

1.4.2 与端到端模型的对比与互补：白盒与黑盒的协同

本研究致力于构建一套显式、理论驱动的特征工程范式，这与当前人工智能领域中流行的隐式、数据驱动的端到端（End-to-End）模型（如QCFormer）形成了鲜明对比，却又构成了一种深刻的互补关系。

本工作的贡献：我们旨在创建一个“白盒”式的、可解释的、具有物理意义的描述符体系。其核心目标是知识发现，即通过定量化材料的结构、电子和拓扑特征，提供一套能够揭示材料“遗传密码”的鲁棒性描述符。这些描述符不仅能用于预测，更关键的是能够回溯其物理起源，从而为材料科学提供深层的物理化学洞察，指导新材料的理性设计，而非仅仅提供一个预测结果。

QCFormer的贡献：作为一种先进的图神经网络模型，QCFormer代表了“黑箱”式的高性能预测器。其主要目标在于实现特定任务（如材料性能预测）的最先进预测精度。通过复杂的网络结构和大量的训练数据，QCFormer能够学习到材料特征与性能之间的隐式、非线性映射关系，但其内部运作机制往往难以直接解释。

协同效应：本研究并非旨在与端到端模型竞争，而是作为一种高度互补的方法。我们的显式描述符体系为QCFormer等黑箱模型提供了高质量、物理真实的输入特征，从而能够潜在地提高其预测精度和泛化能力。更重要的是，本工作提供了端到端模型往往缺乏的可解释性和物理洞察力。通过结合两者的优势，我们可以实现：首先，利用白盒描述符进行知识驱动的材料设计和原理发现；其次，将这些高级描述符作为输入，进一步赋能黑箱模型，以达到卓越的预测性能。这种协同范式有助于构建更全面、更强大的材料智能研究体系。

1.4.3 传统群论的革新与拓展：从定性分类到定量测量

本研究是对经典晶体学中传统群论的一次革命性拓展与深化。传统群论提供了对周期性材料对称性的定性分类（如空间群标签），它能够识别材料的对称操作，并据此将材料归类。然而，它在量化对称性偏

离的程度，以及捕捉真实材料中普遍存在的连续畸变和非理想结构方面存在局限性。

本工作将周期性材料的描述从纯粹的几何对称操作，提升到对几何、量子物理和抽象代数的深度融合，实现了从定性分类到定量测量的转变：

- 从定性分类到定量测量：**传统群论主要提供的是一个离散的对称性标签（例如 Pm-3m 空间群）。而我们的描述符，如 `global_asymmetry_norm` 和 `point_group_reduction_factor`，则提供了对周期性材料对称性及其偏离的连续、可微的量化值。这意味着，即使两个周期性材料属于同一个空间群，我们的描述符也能捕捉到它们之间在对称性破缺程度上的细微差异，为连续对称性破缺的定量研究提供了工具。
- 从几何到物理场的融合：**传统的群论主要关注纯粹的几何对称操作。而本研究则将几何信息与物理场（如电子密度梯度、力场）深度融合。例如，`pseudo_symplectic_coupling` 量化了键方向向量与电子密度梯度之间的耦合，而 `total_torsional_stress` 则关联了结构不对称性与原子受力。这种融合超越了纯几何操作，使得描述符能够捕捉到结构畸变与电子-声子耦合、电荷传输等物理现象之间的深层联系，从而更全面地反映材料的本征属性和功能特性。

通过这种方式，本研究不仅是对传统晶体学工具箱的补充，更是一种范式上的创新，为材料科学提供了前所未有的定量化和可解释性工具，以应对复杂周期性材料的结构-性能挑战。

第二章：方法论：多尺度单纯复形描述符框架

本章系统阐述了我们所提出的多尺度单纯复形描述符框架的构建方法。该框架旨在从多个层次和维度全面刻画周期性材料的“基因组”，从而为材料的理性设计和性能预测提供严谨、可解释且物理真实的特征。我们的方法论遵循单纯复形理论，将周期性材料抽象为由不同维度的单纯形构成的复形，并针对每个维度开发了一系列创新的描述符，最终通过统计聚合形成一个维度固定的全局特征向量。

具体而言，本框架包含以下四个核心层次的描述符，它们共同构成了对周期性材料从微观到宏观、从局部到整体的全面表征：

- 0-单纯形描述符 (第三章)：**聚焦于原子中心环境，旨在捕捉单个原子及其局域配位环境的内在物理化学属性、电子结构特征，以及经典几何与抽象代数特性。这些描述符包含了丰富的量子化学属性，如Bader电荷、静电势、电子密度和电子局域化函数（ELF）。由于这些量子特征以三维网格数据的形式从第一性原理计算中获得，我们采用高精度三线性插值方法，从连续的物理场分布中精确地估算出原子核位点上的局域值，从而弥合了网格计算与原子中心特性描述之间的鸿沟。本章将详细介绍包括 `vectorial_asymmetry_norm_sq`（量化局域环境的几何不对称性）和 `symmetry_breaking_quotient`（从群论角度评估对称性破缺程度）在内的关键特征。

2. **1-单纯形描述符 (第四章)**: 侧重于化学键, 旨在刻画键合过程中原子间的相互作用、信息传递以及键合环境的复杂拓扑特征。本章将重点阐述如 `lie_algebra_incompatibility` (量化键两端结构张量的几何不相容性)、`quotient_algebra_orbit_size` (从商代数角度衡量键的拓扑等价性) 以及 `pseudo_symplectic_coupling` (揭示键的结构-电子耦合强度) 等创新描述符。
3. **2-单纯形描述符 (第五章)**: 关注三体相互作用, 旨在捕捉超越两体相互作用的更复杂、更精细的几何、电子和拓扑信息。本章将引入如 `hierarchical_stress_flow` (描述应力如何在三体系统中传递与耗散) 和 `point_group_reduction_factor` (衡量三体系统的对称性和谐度) 等高级特征, 它们对于理解键角畸变和晶格稳定性至关重要。
4. **全局描述符 (第六章)**: 作为对前三章局部描述符的高阶汇总、聚合与抽象, 旨在量化周期性材料的整体宏观属性。本章将展示如何通过统计聚合所有局部特征的分布, 并结合 `global_asymmetry_norm`、`log_pseudo_symplectic_fluctuation_volume` 以及一系列创新的图论路径分析特征, 为人工智能驱动的材料设计和性能预测提供高级且综合的输入。

每个层次的描述符都经过精心设计, 融合了数学、物理、化学和计算机科学的交叉理论, 并通过严谨的计算逻辑实现, 确保了其数值的严谨性、物理的可靠性与学术的普适性。在接下来的章节中, 我们将对每个层次的描述符进行详细的数学定义、物理化学意义及多学科内涵的阐述。

第三章: 0-单纯形描述符 - 原子中心环境的深度刻画与多维度表征

3.1 引言: 原子作为0-单纯形——周期性材料微观基石的精细解码

在晶体材料的描述体系中, 0-单纯形 (0-Simplex) 是最基本的构成单元, 其在物理实体上直接对应于系统中的单个原子。本章所构建的0-单纯形描述符, 其核心目标是超越原子本身的元素属性, 深入到对其所处局域化学与物理环境的精细化、定量化表征。传统的描述符, 如原子序数或电负性, 仅提供孤立原子的信息, 而无法捕捉该原子在晶格内因与其他原子相互作用而产生的复杂变化。

为此, 我们设计的0-单纯形描述符不仅涵盖了原子的内禀属性, 更重要的是, 它将第一性原理计算得到的一系列局域物理场 (如电子密度、Bader电荷、静电势等) 以及通过代数拓扑方法定义的结构几何信息, 都锚定在每个原子中心上。这种做法使得每个0-单纯形 (原子) 都从一个简单的元素符号, 转变为一个承载了丰富结构-物理耦合信息的高维向量。

本章将详细阐述每一类0-单纯形描述符的数学定义、物理化学内涵及其计算实现。这些描述符共同构成了一个高维度的原子“指纹”, 为后续在更高阶单纯形 (如1-单纯形/化学键、2-单纯形/三体相互作用) 中

构建更复杂的跨尺度特征奠定了坚实的基础。

3.2 基础物理化学描述符：原子内禀属性与经典环境的量化

本节旨在详细阐述一系列原子层面的基础物理化学描述符，它们是定义原子身份及其在晶体材料中基本角色的基石。这些属性不仅是元素周期性行为的核心体现，也是构建更复杂材料特征的起点。这些描述符主要通过标准化学信息库（如Mendeleev、Pymatgen）直接获取，或通过基础晶体化学分析得出。

3.2.1 原子序数 (atomic_number)

原子序数 Z 被严格定义为原子核中质子的数量，是决定元素化学性质和其在周期表中位置的基本量。在物理、化学与材料学层面，原子序数作为最基本的身份标识，直接决定了原子的核电荷和电子排布，进而影响其尺寸和所有化学性质。在晶体材料描述中，它是不可或缺的基本组成特征，为构建组合化学特征和理解原子间相互作用奠定基础。

3.2.2 电负性 (electronegativity)

电负性 χ (Pauling标度) 严格定义为原子在化学键中吸引电子能力的相对量。从物理、化学与材料学的角度看，电负性是判断化学键类型（离子键、共价键、极性共价键）和键合极性的关键指标；原子间电负性差值越大，键的离子性越强。在晶体材料中，电负性影响材料的介电常数、功函数、能带结构和化学反应性。特别是在复合材料中，不同元素的电负性差异会驱动电荷转移，形成偶极子，进而影响材料的极化行为和界面性能，对钙钛矿等极性材料的功能性尤为关键。对于计算机与人工智能学科，电负性作为连续数值特征，可以帮助机器学习模型学习和识别化学键的极性强度，进而预测与电子结构和界面性质相关的材料性能。

3.2.3 第一电离能 (ionization_energy)

第一电离能 I_1 严格定义为从气态孤立基态原子中移除一个电子所需的最小能量，其单位为kJ/mol。在物理、化学与材料学层面，电离能反映了原子对外层电子的束缚能力，能量越高，电子越难被移除。它与原子的金属性、还原性以及参与氧化还原反应的难易程度密切相关；低电离能的原子倾向于形成阳离子并贡献自由电子，从而影响晶体材料的导电性（如金属导体和半导体的电子激发阈值），以及光电效应中光伏材料的激发阈值。在计算机与人工智能学科中，作为连续数值特征，电离能有助于模型理解原子形成阳离子的倾向及其在电子传输、电化学反应中的作用，是预测材料电学和光电性能的重要输入特征。

3.2.4 电子亲和能 (electron_affinity)

电子亲和能 E_{ea} 严格定义为气态孤立基态原子接受一个电子并形成负离子时所释放的能量，其单位为kJ/mol。从物理、化学与材料学的角度看，电子亲和能反映了原子捕获电子的能力，能量越高（释放能

量越多)，原子越容易形成负离子。它与原子的非金属性、氧化性以及参与化学反应的难易程度密切相关；高电子亲和能的原子会影响晶体材料的功函数（对电子注入/提取界面至关重要）、半导体的导电类型（n型或p型掺杂），以及电荷传输和催化反应中的电子捕获能力。对于计算机与人工智能学科，作为连续数值特征，电子亲和能有助于模型理解原子形成阴离子的倾向及其在电子传输中的作用。它与电离能共同为模型提供了原子得失电子能力的全面量化，对于预测半导体、光伏材料和催化剂的性能具有指导意义。

3.2.5 价电子数 (valence_electrons)

价电子数 N_v 严格定义为原子外层电子的数量，这些电子通常参与化学键的形成。在物理、化学与材料学层面，价电子数直接决定了原子的成键能力和其在化学键中的价态，是化学键形成、化合价和晶体材料结构预测的核心参数，影响材料的导电性（价电子形成导带或价带）、光学性质（电子跃迁）以及化学反应活性。对于半导体，价电子数决定了其作为n型或p型掺杂剂的潜力。在计算机与人工智能学科中，它是一个离散的数值特征，反映原子在化学键合中的活跃程度，对于预测材料的电子结构相关性能（如导电性、光学吸收）非常有用。

3.2.6 离子半径 (ionic_radius)

离子半径 r_{ion} 严格定义为离子在晶体材料中有效占据的空间大小，它取决于离子的电荷、配位数和自旋态。在物理、化学与材料学层面，离子半径决定了离子的空间填充效率、晶格常数和晶胞体积，是预测离子型晶体材料结构稳定性（如Goldschmidt容忍因子、鲍林规则）、离子迁移率和固溶体形成的关键因素。它直接影响材料的密度、熔点、硬度以及离子导体中的离子迁移通道大小。在计算机与人工智能学科中，作为一个连续数值特征，离子半径反映离子在晶体堆积中的空间贡献，对于预测材料结构稳定性、可合成性、密度以及在固态离子导体中判断离子传输通道的“开放性”具有重要作用。

3.2.7 共价半径 (covalent_radius)

共价半径 r_{cov} 严格定义为原子在共价键中有效占据的空间大小，单位为埃（Å）。在物理、化学与材料学层面，共价半径决定了原子在共价键中的键长和键角，是预测共价键键长、分子几何形状和共价型晶体材料结构的关键因素。它影响共价材料的密度、硬度以及与共价键相关的电子结构（如能带）。在计算机与人工智能学科中，共价半径是一个连续数值特征，反映原子在共价键合中的空间贡献。它通常与实际键长结合使用，以形成键长畸变等更高阶特征，为机器学习模型提供关于键合环境的精细信息。

3.2.8 配位数 (coordination_number)

配位数 CN 严格定义为晶体材料中一个中心原子周围与其直接相连的近邻原子数量。本框架统一采用 `pymatgen.analysis.local_env.CrystalNN` 算法来确定配位数，`CrystalNN` 算法通过分析原子间的距离、元素性质以及周期性边界条件来识别化学键，并统计与中心原子直接相连的原子数量。这是一个结合了几何判断和化学规则的复杂过程，旨在模拟真实的化学键合环境。在物理、化学与材料学层面，配位数直

接反映了原子在晶体材料晶格中的空间填充和局域相互作用强度。它是判断原子价态、键合类型和局域几何构型的核心指标，直接影响配位多面体的形状和连接方式，进而影响材料的密度、稳定性、电子结构（如晶体场分裂）和许多宏观物理性能（如硬度、熔点、离子迁移率）。配位数的改变常与相变或缺陷形成相关。在计算机与人工智能学科中，作为一个离散的数值特征，配位数反映原子局域环境的结构紧密程度和键合复杂性，可以作为图神经网络中节点特征的一部分，有助于模型理解晶体网络中的拓扑特性。

3.2.9 平均位点价态 (avg_site_valence)

平均位点价态 V_{avg} 严格定义为通过 `pymatgen.analysis.bond_valence.BVAnalyzer` 计算得到的原子形式氧化态 (formal oxidation state)。该分析器尝试根据晶体材料的晶体结构和化学式确定每个原子的最可能氧化态。并且 `BVAnalyzer` 内部包含了一套启发式规则和键价模型，旨在从晶体结构和化学式中推断出最合理的原子氧化态。在物理、化学与材料学层面，该特征反映了原子在晶体内部的平均电荷状态，与电子在原子间的转移和局域化相关。它是化学计量学、氧化还原反应和化合物稳定性预测的重要参数。在材料学中，它影响材料的导电性（价态变化可能导致半导体或绝缘体行为）、磁性（未配对电子）以及催化活性（氧化还原中心）。在计算机与人工智能学科中，平均位点价态反映原子在晶体材料中的氧化还原状态，对预测材料的电化学和催化性能具有重要意义。

3.2.10 键价和 (bond_valence_sum)

键价和 BVS_i 严格定义为中心原子 i 所形成的所有化学键的键价总和。每个键的键价 s_{ij} 是根据键长 d_{ij} 和经验参数 R_0, B 计算的：

$$s_{ij} = \exp\left(\frac{R_0 - d_{ij}}{B}\right)$$

其中 R_0 和 B 是与特定元素对相关的经验常数，通常来源于专门的键价参数数据库（如 `pymatgen` 库内部集成的张弛(Brese)和奥基夫(O'Keeffe)参数）。中心原子 i 的键价和 BVS_i 为：

$$BVS_i = \sum_{j \in \text{NN}(i)} s_{ij}$$

在理想情况下， BVS_i 应接近中心原子 i 的形式氧化态。 BVS 是一个半经验模型，其核心思想是原子键价的总和应等于原子的形式氧化态。每个键的键价 s_{ij} 通过经验公式将键长转化为价键贡献，其中 R_0 和 B 参数通过拟合大量已知结构获得。总和则是对所有近邻键价的累加。在物理、化学与材料学层面，键价和反映了原子局域成键环境的电荷平衡以及是否存在应力（过配位或欠配位）。它是验证假定原子氧化态是否合理、诊断晶体材料结构中不寻常键合状态的重要工具，对几何畸变非常敏感。在材料学中，它常用于判断材料的结构稳定性、离子迁移通道（ BVS 偏离理想值可能指示离子迁移位点）和

催化活性。高BVS偏离度可能意味着局域应变或非晶化倾向。在计算机与人工智能学科中，它是一个连续数值特征，提供了对原子局域电荷平衡和应变状态的量化，能够捕捉结构中的微妙变化，对于预测材料的结构稳定性、相变行为和离子输运性能非常有用。

3.2.11 键长畸变 (bond_length_distortion)

键长畸变 Bond Length Distortion 描述符严格定义为中心原子所有近邻键长的相对标准差，它量化了局域配位多面体中键长的均匀性，反映了几何畸变的程度。其公式定义为：

$$\text{Bond Length Distortion} = \sqrt{\frac{1}{N_{\text{NN}}} \sum_{j=1}^{N_{\text{NN}}} \left(\frac{d_{ij} - \bar{d}_i}{\bar{d}_i} \right)^2} \quad (\text{当 } N_{\text{NN}} > 0 \text{ 时})$$

其中 d_{ij} 是中心原子 i 到其近邻 j 的键长， \bar{d}_i 是中心原子 i 的平均键长， N_{NN} 是配位数（从第4.2.8节获取）。当 $N_{\text{NN}} \leq 1$ 时，畸变指数定义为0。

在物理、化学与材料学层面，键长畸变直接反映了晶体材料中晶体场对离子键长对称性的影响，与晶格应变、局域结构相变和电子局域化相关。它是判断配位多面体畸变程度的直接指标，与Jahn-Teller效应、离子尺寸不匹配和键合的共价/离子性混合程度相关。在材料学中，它影响材料的晶体对称性、电子能带结构、光学性质（如吸收/发射光谱）和输运性质（如离子迁移）。高度畸变的多面体在功能材料（如铁电体、催化剂）中扮演着关键角色。在计算机与人工智能学科中，它是一个连续数值特征，提供了对局域几何畸变的量化，有助于机器学习模型学习和预测与结构不稳定或特定功能特性（如铁电性、压电性）相关的材料行为。

3.3 量子化学核心描述符：电子结构与局域物理场的精细表征

本节描述符直接从密度泛函理论（DFT）第一性原理计算结果中提取，旨在深入揭示原子在晶体材料中的电子结构细节、电荷分布特征、局域磁性以及能量态密度。这些特征是理解晶体材料本征物理化学性质（如导电性、光学性质、催化活性、磁性）的量子力学基础，具有高度的物理真实性和预测能力。

3.3.1 Bader 电荷 (bader_charge)

Bader电荷 q_i 严格定义为原子核和其Bader体积（Bader Volume）内的电子总数之差。Bader体积是基于分子中原子的量子理论（Quantum Theory of Atoms in Molecules, QTAIM）概念，通过电子密度 $\rho(\mathbf{r})$ 的拓扑分析确定的，其核心在于零通量表面（zero-flux surface），即电子密度梯度 $\nabla\rho(\mathbf{r})$ 在该表面上处处为零或垂直于表面。原子 i 的Bader电荷 q_i 严格定义为：

$$q_i = Z_i - \int_{\Omega_i} \rho(\mathbf{r}) d^3\mathbf{r}$$

其中 Z_i 是原子 i 的原子序数（核电荷）， Ω_i 是通过QTAIM划分出的原子 i 的Bader体积， $\int_{\Omega_i} \rho(\mathbf{r}) d^3\mathbf{r}$ 是该体积内积分的电子总数。

在物理、化学与材料学层面，Bader电荷直接量化了原子在晶体材料中的实际电荷状态，反映了电荷在原子间的转移和局域化程度。正值表示原子失去电子（阳离子特性），负值表示原子获得电子（阴离子特性）。它是判断化学键离子性/共价性强弱的定量指标，与化学反应活性、氧化还原能力和分子间相互作用密切相关。对于多价态原子，Bader电荷可以区分其在不同环境下的实际价态。在材料学中，它影响材料的电荷输运（如离子导体和半导体中的载流子）、介电性质（如极化强度）、催化活性（活性位点的电荷状态）和表面性质（如功函数）。在离子型钙钛矿等晶体材料中，Bader电荷的分布对晶格稳定性、铁电性及带隙性质有显著影响。在计算机与人工智能学科中，作为一个连续数值特征，Bader电荷提供了原子电荷状态的高精度、物理真实量化，是建立电子结构-性能预测模型的核心输入。

3.3.2 静电势 (electrostatic_potential)

静电势 $V(\mathbf{r})$ 严格定义为晶体材料中每个原子位置 \mathbf{r} 的静电势值。在DFT中，静电势是电子和原子核在空间中产生的平均电场分布，它可以通过对DFT计算得到的静电势网格数据进行高精度三线性插值得到。

三线性插值是多维插值的一种，它通过对包含目标点的小立方体（由八个最近的网格点定义）内的已知数据点进行三次线性插值来估算任意点的值。假设我们有一个三维网格上的函数 $f(x, y, z)$ ，以及一个待插值的目标点 (x_d, y_d, z_d) 。这个目标点被一个由 (x_1, y_1, z_1) 和 (x_2, y_2, z_2) 定义的立方体所包围，其中 $x_1 \leq x_d \leq x_2, y_1 \leq y_d \leq y_2, z_1 \leq z_d \leq z_2$ 。插值过程可以分解为以下步骤：

首先，沿 x 轴进行两次线性插值，得到四条边上四个点的插值结果：

$$\begin{aligned} f_{11}(x_d) &= f(x_1, y_1, z_1) \frac{x_2 - x_d}{x_2 - x_1} + f(x_2, y_1, z_1) \frac{x_d - x_1}{x_2 - x_1} \\ f_{21}(x_d) &= f(x_1, y_2, z_1) \frac{x_2 - x_d}{x_2 - x_1} + f(x_2, y_2, z_1) \frac{x_d - x_1}{x_2 - x_1} \\ f_{12}(x_d) &= f(x_1, y_1, z_2) \frac{x_2 - x_d}{x_2 - x_1} + f(x_2, y_1, z_2) \frac{x_d - x_1}{x_2 - x_1} \\ f_{22}(x_d) &= f(x_1, y_2, z_2) \frac{x_2 - x_d}{x_2 - x_1} + f(x_2, y_2, z_2) \frac{x_d - x_1}{x_2 - x_1} \end{aligned}$$

接下来，利用这四个点沿 y 轴进行两次线性插值：

$$f_1(x_d, y_d) = f_{11}(x_d) \frac{y_2 - y_d}{y_2 - y_1} + f_{21}(x_d) \frac{y_d - y_1}{y_2 - y_1}$$

$$f_2(x_d, y_d) = f_{12}(x_d) \frac{y_2 - y_d}{y_2 - y_1} + f_{22}(x_d) \frac{y_d - y_1}{y_2 - y_1}$$

最后，将这两个结果沿 z 轴进行一次线性插值，得到最终的目标点值：

$$f(x_d, y_d, z_d) = f_1(x_d, y_d) \frac{z_2 - z_d}{z_2 - z_1} + f_2(x_d, y_d) \frac{z_d - z_1}{z_2 - z_1}$$

这一过程确保了从连续的场分布中，精确地估算出原子核所在位置或其周围关键区域的局域值，从而弥合了网格计算与原子局域特性描述之间的鸿沟，为后续高级特征的构建奠定了数据基础。通过这种精确的插值方法，我们可以进一步定义和计算原子位点处的各类物理量。

原子 i 处的静电势 V_i ，表示在原子 i 的笛卡尔坐标 \mathbf{r}_i 处，通过前述的高精度三线性插值方法（Interpolate）从DFT计算得到的静电势三维网格数据 V_{grid} 中获取的局域值。其严格定义为：

$$V_i = V(\mathbf{r}_i) = \text{Interpolate}(V_{\text{grid}}, \mathbf{r}_i)$$

其中 $V(\mathbf{r}_i)$ 是静电势在空间点 \mathbf{r}_i 处的数值，它正是通过对 V_{grid} （DFT计算得到的整个晶胞的静电势三维网格数据）在 \mathbf{r}_i 处进行三线性插值所获得。

在物理、化学与材料学层面，静电势反映了原子所处局域环境的电场强度和方向。静电势的局部高低直接影响电子在晶格中的迁移路径和能量势垒。它与原子的亲电性/亲核性、路易斯酸碱性以及化学反应活性位点密切相关。例如，高负电势区域通常是亲电试剂的攻击目标。在材料学中，它影响材料的介电常数、功函数、电荷传输（如离子和电子导体）、催化活性和吸附性能。静电势的梯度驱动了电荷载流子的运动。在计算机与人工智能学科中，作为一个连续数值特征，静电势提供了原子局域电势环境的精细量化，有助于模型理解和预测电荷在材料中的行为。

3.3.3 电子密度 (electron_density)

电子密度 $\rho(\mathbf{r})$ 严格定义为在晶体材料中每个原子位置 \mathbf{r} 的电子概率密度值。它量化了在给定空间点找到电子的概率。在DFT中，总电子密度由所有占据轨道的电子密度之和给出。原子 i 处的电子密度 ρ_i 严格定义为：

$$\rho_i = \rho(\mathbf{r}_i) = \text{Interpolate}(\rho_{\text{grid}}, \mathbf{r}_i)$$

其中 ρ_{grid} 是DFT计算得到的整个晶胞的电子密度三维网格数据， \mathbf{r}_i 是原子 i 的笛卡尔坐标。

在物理、化学与材料学层面，电子密度是量子力学中的基本量，其分布决定了晶体材料几乎所有的物理性质。电子密度的局部峰值对应于原子核或共价键区域，低值对应于原子间空间。它与化学键的形成、键级、原子间相互作用强度以及Kato Cusp Condition（电子密度在原子核处呈现尖锐的峰值）相关。在材料学中，它影响材料的键合类型（如共价、离子、金属键）、硬度、弹性模量、声子谱（取决于键合强度）、光学性质（如折射率）、以及Mössbauer和NMR谱学（与原子核周围的电子环境直接相关）。在计算机与人工智能学科中，作为一个连续数值特征，电子密度提供了原子局域电子环境的最直接、最基础的量子力学量化。

3.3.4 电子局域化函数 (ELF) (elf)

电子局域化函数（Electron Localization Function, ELF）是一个量化电子在空间中局域化程度的维度为0到1的函数。ELF值接近1表示电子高度局域化（如在共价键或孤对电子区域），值接近0.5表示电子类自由电子气（如在金属中），值接近0表示电子离域化或原子间区域。ELF的严格定义基于电子对概率密度与自由电子气中均匀电子对概率密度的比较。在DFT计算中，通常通过动能密度来计算。原子 i 处的 ELF 值 ELF_i 严格定义为：

$$ELF_i = ELF(\mathbf{r}_i) = \text{Interpolate}(ELF_{\text{grid}}, \mathbf{r}_i)$$

其中 ELF_{grid} 是通过GPAW的 ELF 模块计算得到的整个晶胞的ELF三维网格数据。

在物理、化学与材料学层面，ELF直观地反映了电子在空间中的“壳层”或“成键”特征，提供了一种对化学键（共价键、离子键、金属键）、孤对电子和核心电子区域的可视化方式。它是判断化学键类型、键强度、原子价态和孤对电子存在的有力工具。高ELF区域对应于共价键的中心或非键合电子对。在材料学中，它影响晶体材料的键合强度、硬度、绝缘性（高ELF区域表示电子不易移动）、以及化学反应中的活性位点。在共价材料和半导体中，ELF的分布对于理解载流子传输机制至关重要。在计算机与人工智能学科中，作为一个连续数值特征，ELF提供了原子局域电子局域化程度的高分辨率量化，有助于模型学习和预测与键合类型和电子行为相关的材料性能。

3.3.5 局域磁矩 (local_magnetic_moment)

局域磁矩 μ_i 严格定义为每个原子所贡献的净磁矩，通常通过自旋极化DFT计算中的原子轨道投影或Wigner-Seitz球积分获得。对于自旋极化体系，它表示向上自旋电子密度与向下自旋电子密度的差值在原子区域的积分。原子 i 的局域磁矩 μ_i 严格定义为：

$$\mu_i = \int_{\Omega_i} (\rho_{\uparrow}(\mathbf{r}) - \rho_{\downarrow}(\mathbf{r})) d^3\mathbf{r}$$

其中 ρ_{\uparrow} 和 ρ_{\downarrow} 分别是向上和向下自旋电子密度， Ω_i 是原子 i 的区域。

在物理、化学与材料学层面，局域磁矩直接反映了原子在晶体材料中的磁性状态（如铁磁性、反铁磁性、顺磁性）。非零磁矩表示原子具有未配对电子。它与过渡金属离子的d轨道电子排布、自旋态以及配体场理论密切相关。在材料学中，它是设计和理解磁性材料、自旋电子学材料、磁存储材料以及磁性催化剂的关键参数。局域磁矩的分布决定了材料的整体磁序和磁各向异性。在计算机与人工智能学科中，作为一个连续数值特征，局域磁矩提供了原子局域磁性状态的精确量化，是建立磁性材料设计模型的核心输入。

3.3.6 费米能级处的局域态密度 (local_dos_fermi)

原子 i 在费米能级 ε_F 处的局域态密度（Local Density of States, LDOS） $\text{LDOS}_i(\varepsilon_F)$ 严格定义为费米能级附近单位能量间隔内原子 i 贡献的电子态数量。它通常通过对投影态密度（Projected Density of States, PDOS）进行插值或在费米能级附近求积分得到。

在物理、化学与材料学层面，local_dos_fermi反映了原子在费米能级附近的电子局域态密度，是判断晶体材料导电性（金属、半导体、绝缘体）的关键指标。高LDOS值通常表示原子对导电性有较大贡献。它与原子在晶体中的成键能力、轨道杂化模式以及化学反应活性位点密切相关。在材料学中，直接影响材料的电导率、半导体带隙特性、肖特基势垒以及光电转换效率。在计算机与人工智能学科中，作为直接反映原子局域电子结构和成键特性的量子描述符，它是构建高精度材料性能预测模型不可或缺的输入。

3.3.7 s轨道电子数 (s_electron_count)

原子 i 的 s 轨道电子数 $N_{i,s}$ 严格定义为该原子 s 轨道投影态密度从价带底到费米能级积分的电子总数：

$$N_{i,s} = \int_{E_{\text{valence}}}^{E_F} \text{PDOS}_{i,s}(E) dE$$

其中 $\text{PDOS}_{i,s}(E)$ 是原子 i 的 s 轨道在能量 E 处的投影态密度。

在物理、化学与材料学层面，s_electron_count揭示了原子 s 轨道上的电子占据情况和键合类型，通常与各向同性的球形对称键合相关。在材料学中，它影响晶体材料的稳定性、机械强度（如通过形成强的方向性共价键）以及功函数。

3.3.8 p轨道电子数 (p_electron_count)

原子 i 的 p 轨道电子数 $N_{i,p}$ 严格定义为该原子 p 轨道投影态密度从价带底到费米能级积分的电子总数：

$$N_{i,p} = \int_{E_{\text{valence}}}^{E_F} \text{PDOS}_{i,p}(E) dE$$

其中 $\text{PDOS}_{i,p}(E)$ 是原子 i 的 p 轨道在能量 E 处的投影态密度。

在物理、化学与材料学层面，`p_electron_count`揭示了原子 p 轨道上的电子占据情况和键合类型，通常与方向性共价键和 π 键相关。在材料学中，它影响晶体材料的导电性（通过形成导电网络）、光学性质（通过 π 电子共轭）以及化学反应活性。

3.3.9 d轨道电子数 (`d_electron_count`)

原子 i 的 d 轨道电子数 $N_{i,d}$ 严格定义为该原子 d 轨道投影态密度从价带底到费米能级积分的电子总数：

$$N_{i,d} = \int_{E_{\text{valence}}}^{E_F} \text{PDOS}_{i,d}(E) dE$$

其中 $\text{PDOS}_{i,d}(E)$ 是原子 i 的 d 轨道在能量 E 处的投影态密度。

在物理、化学与材料学层面，`d_electron_count`揭示了原子 d 轨道上的电子占据情况和键合类型，尤其对过渡金属的磁性、催化活性和强关联效应至关重要。例如，半满d轨道的过渡金属通常具有高催化活性。在材料学中，它直接决定了材料的磁性起源和磁相变、催化活性（与d带中心理论密切相关）以及电子能带结构。本框架在计算该特征时，通过智能判断原子的原子序数和化学环境，并应用适应性阈值，以确保只有当d轨道具有物理相关性时才进行计算，显著提升了特征的稳健性和数据质量。在计算机与人工智能学科中，作为直接反映原子局域电子结构和成键特性的量子描述符，它是构建高精度材料性能预测模型（如导电性、催化活性、热电性能、超导性）不可或缺的输入，其物理驱动的构建方式赋予了模型更强的泛化能力和物理洞察力。

3.4 经典几何与代数描述符：0-单纯形局域几何的抽象量化

本节将介绍一组基于经典几何和抽象代数概念的描述符，旨在从更高级的数学视角量化晶体材料中原子局域环境的形状、对称性偏离以及结构刚性。这些特征超越了简单的键长、键角，提供了对原子周围空间分布的几何不变性和代数结构的深层洞察。

3.4.1 矢量不对称性模方 (vectorial_asymmetry_norm_sq)

矢量不对称性模方 $\|\mathbf{V}_{\text{asymmetry}}\|^2$ 描述符严格定义为中心原子到其所有近邻原子位移向量矢量和的欧几里得范数平方。它直接量化了局域原子环境偏离中心对称的程度。其公式为：

$$\|\mathbf{V}_{\text{asymmetry}}\|^2 = \left\| \sum_{j \in \text{NN}(i)} (\mathbf{r}_j - \mathbf{r}_i) \right\|_2^2$$

其中 \mathbf{r}_i 是中心原子 i 的坐标， \mathbf{r}_j 是其近邻 j 的坐标， $\text{NN}(i)$ 是原子 i 的近邻集合。当局部环境具有中心对称性时，这个矢量和为零，模方也为零。

在物理、化学与材料学层面，该特征直接反映了原子局域环境的几何极性。非零值表示存在净的偶极矩或电荷中心偏离，这与晶体材料的压电性、铁电性以及自发极化相关。它与配位多面体的畸变、键角应变以及立体化学活性密切相关。例如，在ABX₃钙钛矿中，B位阳离子的不对称性与铁电性有直接联系。在材料学中，它影响材料的极性、介电常数、非线性光学效应和力学响应。高模方值可能预示着材料具有较大的自发极化强度或在外部场下容易发生形变。在计算机与人工智能学科中，作为一个连续数值特征，它提供了局域几何极性的精确量化，有助于机器学习模型预测材料的功能特性。

3.4.2 均方邻居距离 (mean_squared_neighbor_distance)

均方邻居距离描述符严格定义为中心原子局域环境结构张量 (Structure Tensor) 的迹 (Trace) 除以其配位数 (近邻数量)。结构张量 $\mathbf{T}_{\text{struct},i}$ 是一个二阶张量，其定义为中心原子到其所有近邻位移向量外积的求和：

$$\mathbf{T}_{\text{struct},i} = \sum_{j \in \text{NN}(i)} \mathbf{v}_{ij} \otimes \mathbf{v}_{ij} = \sum_{j \in \text{NN}(i)} (\mathbf{r}_j - \mathbf{r}_i) \otimes (\mathbf{r}_j - \mathbf{r}_i)$$

其中 \otimes 表示外积， $\mathbf{v}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ 代表从中心原子 i 指向近邻原子 j 的位移向量。该张量的迹 $\text{Tr}(\mathbf{T}_{\text{struct},i})$ 严格定义为：

$$\text{Tr}(\mathbf{T}_{\text{struct},i}) = \sum_{j \in \text{NN}(i)} \|\mathbf{r}_j - \mathbf{r}_i\|_2^2 = \sum_{j \in \text{NN}(i)} d_{ij}^2$$

最终的均方邻居距离计算公式为：

$$\text{Mean Squared Neighbor Distance} = \frac{\text{Tr}(\mathbf{T}_{\text{struct},i})}{N_{\text{NN}}} = \frac{1}{N_{\text{NN}}} \sum_{j \in \text{NN}(i)} d_{ij}^2$$

其中 N_{NN} 是中心原子 i 的配位数。当 $N_{\text{NN}} = 0$ 时，该值为0。

尽管其计算最终等价于所有近邻键长平方的平均值，这种基于结构张量的定义提供了更严谨的数学基础和概念上的普适性。它将局域环境的“尺寸”或“分散程度”量化为一个基础的几何不变量，同时也为未来拓展至表征局域环境各向异性等更复杂的几何属性（通过张量其他不变量）提供了统一的框架。

在物理、化学与材料学层面，该特征直观反映了原子局域环境的空间“尺度”。值越大表示近邻原子分布越稀疏或平均键长越长。它与局域体积、原子堆积密度和声子振动模态等宏观性质密切相关。例如，在离子导体中，较大的均方邻居距离可能暗示了更开放的离子迁移通道。在计算机与人工智能学科中，作为一个连续数值特征，它提供了局域原子分布空间尺度的量化，且其推导过程的严谨性确保了特征的可靠性与可解释性。

3.4.3 局域环境各向异性 (local_environment_anisotropy)

局域环境各向异性 α_i 描述符严格定义为从原子局域环境结构张量（Structure Tensor）的特征值计算出的一个无量纲指数，用于量化局域环境的几何各向异性程度。对于结构张量 $\mathbf{T}_{\text{struct},i}$ ，计算其三个非负实数特征值 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ 。各向异性指数 α_i 严格定义为：

$$\alpha_i = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1 + \lambda_2 + \lambda_3)^2}}$$

为了展示该指数与结构张量基本不变量的内在联系并提升其数学严谨性，我们可以将其平方 α_i^2 进一步推导为结构张量的迹（一阶不变量）和平方迹（二阶不变量）的函数。这一等价形式清晰地揭示了该描述符是基于原子局域环境中键长平方和（由 $\text{Tr}(\mathbf{T}_{\text{struct},i})$ 给出）以及其二阶矩信息（由 $\text{Tr}(\mathbf{T}_{\text{struct},i}^2)$ 给出）的内在几何关系。这种基于张量不变量的推导不仅提供了更深层次的数学见解，也提升了特征的理论完备性和可解释性。

数学推导：

我们已知：

1. 结构张量的迹： $\text{Tr}(\mathbf{T}_{\text{struct},i}) = \lambda_1 + \lambda_2 + \lambda_3$
2. 结构张量平方的迹： $\text{Tr}(\mathbf{T}_{\text{struct},i}^2) = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$

现在展开 α_i^2 的分子部分：

$$\begin{aligned} & (\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2 \\ &= (\lambda_1^2 - 2\lambda_1\lambda_2 + \lambda_2^2) + (\lambda_2^2 - 2\lambda_2\lambda_3 + \lambda_3^2) + (\lambda_3^2 - 2\lambda_3\lambda_1 + \lambda_1^2) \\ &= 2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2) - 2(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1) \end{aligned}$$

我们利用代数恒等式 $(\lambda_1 + \lambda_2 + \lambda_3)^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 + 2(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1)$ ，可以得到：

$$2(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1) = (\lambda_1 + \lambda_2 + \lambda_3)^2 - (\lambda_1^2 + \lambda_2^2 + \lambda_3^2)$$

将此代回分子表达式：

$$\begin{aligned} \text{分子} &= 2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2) - [(\lambda_1 + \lambda_2 + \lambda_3)^2 - (\lambda_1^2 + \lambda_2^2 + \lambda_3^2)] \\ &= 3(\lambda_1^2 + \lambda_2^2 + \lambda_3^2) - (\lambda_1 + \lambda_2 + \lambda_3)^2 \end{aligned}$$

用迹的表示法替换：

$$\text{分子} = 3 \cdot \text{Tr}(\mathbf{T}_{\text{struct},i}^2) - (\text{Tr}(\mathbf{T}_{\text{struct},i}))^2$$

现在，将分子和分母都用迹的表示法代回 α_i^2 的定义式：

$$\begin{aligned} \alpha_i^2 &= \frac{3 \cdot \text{Tr}(\mathbf{T}_{\text{struct},i}^2) - (\text{Tr}(\mathbf{T}_{\text{struct},i}))^2}{2(\text{Tr}(\mathbf{T}_{\text{struct},i}))^2} \\ &= \frac{3}{2} \frac{\text{Tr}(\mathbf{T}_{\text{struct},i}^2)}{(\text{Tr}(\mathbf{T}_{\text{struct},i}))^2} - \frac{1}{2} \end{aligned}$$

因此，局域环境各向异性指数可表示为：

$$\alpha_i = \sqrt{\frac{3}{2} \frac{\text{Tr}(\mathbf{T}_{\text{struct},i}^2)}{(\text{Tr}(\mathbf{T}_{\text{struct},i}))^2} - \frac{1}{2}}$$

此指数的取值范围为 $[0, 1]$ ：当 $\alpha_i = 0$ 时，表示局部环境是完全各向同性的（如完美球形或立方对称）；当 $\alpha_i = 1$ 时，表示局部环境是高度各向异性的（如一维链状结构）。

在物理、化学与材料学层面，该特征反映了原子局域环境的几何畸变程度和空间各向异性。高指数表明局部电场、应变或能量分布可能不均匀，从而影响电荷传输、声子散射等物理过程。它与配位多面体的形状、键角和键长分布的不均匀性相关，有助于理解局部键合对宏观性能的影响。在材料学中，它影响晶体材料的力学性能（如各向异性弹性模量）、热输运（如各向异性热导率）、光学性质（如双折射）和电子输运。高度各向异性的局部环境在功能材料（如铁电体、压电体、热电材料）中扮演着关键角色。在计算机与人工智能学科中，作为一个连续数值特征，它提供了局域环境几何各向异性的高分辨率量化，有助于机器学习模型预测与结构相关的功能特性。

3.4.4 对称性破缺商 (symmetry_breaking_quotient)

对称性破缺商描述符严格定义为原子晶体学位点对称群的阶（Order）与该原子局域环境（包括近邻）点群的阶之比。这一无量纲指标旨在量化原子实际所处对称性相对于其理想晶体学对称性的偏离程度，即由于局域环境的几何或化学不对称性导致的对称性降低。其数学表达式为：

$$\text{Symmetry Breaking Quotient} = \frac{|G_{\text{site}}|}{|G_{\text{local}}|}$$

其中， G_{site} 代表原子位点在晶体周期性下的稳定子子群（Stabilizer Subgroup），其阶 $|G_{\text{site}}|$ 是保持该位点不变的空间群操作的数量； G_{local} 则是以该原子为中心，其局域配位团簇（包括中心原子及其所有近邻）所形成的分子点群（Point Group）。该商值严格限制在 $[0, 1]$ 之间。在理想情况下，若局域环境能够完全保持原子位点的晶体学对称性，则商值为1；任何形式的对称性破缺都将导致商值小于1，且值越小意味着对称性破缺越剧烈。

在物理、化学与材料学层面，对称性破缺商直接反映了晶体材料内部局域对称性破缺的程度。该特征与局域应变、电偶极子形成以及晶格畸变引起的宏观物理效应（如压电效应和铁电效应）密切相关。它能够量化配位多面体的畸变、键角应变以及离子在晶格中“偏离中心”的倾向，从而有助于深入理解化学键合对局域对称性的影响。在材料学中，它直接影响材料的铁电性、压电性、介电常数和光学非线性。高的对称性破缺值可能预示着材料具有较大的自发极化或更容易响应外部电场。例如，在钙钛矿材料中，B位阳离子的对称性破缺是诱发铁电相变的关键驱动力之一。在计算机与人工智能学科中，作为一个连续数值特征，对称性破缺商提供了局域对称性偏离的精确、群论量化，有助于机器学习模型捕捉与对称性相关的功能特性，从而实现对具有特定对称性相关性能材料的预测与设计。

3.5 深度融合量子-代数描述符：量子信息与几何代数的内禀耦合

本节将介绍一组开创性的深度融合描述符，它们突破了传统特征工程中简单拼接或线性组合的局限，实现了量子化学信息（如电子密度、Bader电荷、ELF）与几何代数（如结构张量、李代数）的内禀耦合。这些特征通过在张量构造或数学运算之初就将量子信息作为“权重”或“作用量”引入，从而构建出更具物理深度和抽象严谨性的描述符。这种“物理驱动的融合范式”揭示了原子局域结构与电子行为之间的高阶相互作用，为理解晶体材料的深层“基因型”提供了前所未有的工具。

3.5.1 结构-化学不相容性 (structure_chemistry_incompatibility)

structure_chemistry_incompatibility 描述符严格定义为原子局域环境的结构主方向与原子的扩展化学矢量在李代数 $so(3)$ 中的交换子（Commutator）的弗罗贝尼乌斯范数，旨在量化局域几何结构与原子内

在化学性质之间的“非对易性”或“冲突程度”。该特征的计算首先从原子局域结构张量 $\mathbf{T}_{\text{struct},i}$ （定义见 3.4.2 节）的特征值分解中，选择其最大特征值对应的特征向量作为局域结构的主要伸展方向 $\mathbf{v}_{\text{struct}}$ 。同时，构建一个 7 维的归一化化学矢量 \mathbf{v}_{chem} ，其包含了多种原子内禀化学属性：[electronegativity, covalent_radius, ionization_energy, electron_affinity, atomic_volume, polarizability, effective_charge]。为在李代数框架下进行比较，需将 3D 结构主方向和 7 维化学矢量严格映射到三维旋转群 $SO(3)$ 的李代数 $\mathfrak{so}(3)$ 的生成元。为了精确实现这一映射，我们首先需要深入理解李群 $SO(3)$ 及其李代数 $\mathfrak{so}(3)$ 的数学结构及其生成元的特性。

三维旋转群 $SO(3)$ 是由所有行列式为 1 的正交 3×3 矩阵构成的李群，其元素 \mathbf{R} 满足 $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ 且 $\det(\mathbf{R}) = 1$ 。李代数 $\mathfrak{so}(3)$ 是李群 $SO(3)$ 在单位元处（即恒等矩阵 \mathbf{I} ）的切空间。李代数中的元素 \mathbf{A} 与李群元素 $\mathbf{R}(t)$ 通过指数映射相关联，即 $\mathbf{R}(t) = \exp(t\mathbf{A})$ 。为了明确 $\mathfrak{so}(3)$ 中元素的具体形式，我们考虑李群 $SO(3)$ 中一条通过单位元的平滑路径。

若 $\mathbf{R}(t)$ 是 $SO(3)$ 中的一条路径，且 $\mathbf{R}(0) = \mathbf{I}$ ，则其在 $t = 0$ 处的导数 $\mathbf{A} = \left. \frac{d\mathbf{R}}{dt} \right|_{t=0}$ 属于李代数 $\mathfrak{so}(3)$ 。

由于 $\mathbf{R}(t)\mathbf{R}(t)^T = \mathbf{I}$ ，对 t 求导得：

$$\begin{aligned} \frac{d}{dt}(\mathbf{R}(t)\mathbf{R}(t)^T) &= \frac{d\mathbf{I}}{dt} \\ \frac{d\mathbf{R}}{dt}\mathbf{R}(t)^T + \mathbf{R}(t)\frac{d\mathbf{R}^T}{dt} &= \mathbf{0} \end{aligned}$$

在 $t = 0$ 处， $\mathbf{R}(0) = \mathbf{I}$ ，且 $\left. \frac{d\mathbf{R}}{dt} \right|_{t=0} = \mathbf{A}$ ，则上式变为：

$$\begin{aligned} \mathbf{A}\mathbf{I}^T + \mathbf{I}\mathbf{A}^T &= \mathbf{0} \\ \mathbf{A} + \mathbf{A}^T &= \mathbf{0} \end{aligned}$$

这证明了李代数 $\mathfrak{so}(3)$ 的元素都是反对称矩阵。换言之， $\mathfrak{so}(3)$ 是所有 3×3 反对称实矩阵的集合。

对于任意三维矢量 $\mathbf{a} = (a_x, a_y, a_z)^T$ ，都可以通过叉积（cross product）运算与反对称矩阵建立同构关系。具体而言，矢量 \mathbf{a} 对应的反对称矩阵 $\mathbf{M}_{\mathbf{a}}$ 使得对于任意矢量 \mathbf{v} ， $\mathbf{a} \times \mathbf{v} = \mathbf{M}_{\mathbf{a}}\mathbf{v}$ 。其严格形式为：

$$\mathbf{M}_{\mathbf{a}} = \begin{pmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{pmatrix}$$

李代数 $\mathfrak{so}(3)$ 上的李括号运算定义为矩阵的交换子 $[\mathbf{A}, \mathbf{B}] = \mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A}$ 。如果 \mathbf{A} 和 \mathbf{B} 是反对称矩阵，则它们的交换子 $[\mathbf{A}, \mathbf{B}]$ 仍然是反对称矩阵，因此李代数在李括号运算下是封闭的。

在明确了三维向量与 $\mathfrak{so}(3)$ 中反对称矩阵的对应关系及李代数的性质之后，关键在于如何将更高维度的化学信息（如7维化学向量 \mathbf{v}_{chem} ）有效地映射到这一三维李代数空间中。

对于7维化学向量 \mathbf{v}_{chem} 到3维投影向量 (x, y, z) 的映射，本研究采用了基于物理意义的严格Gram-Schmidt 正交化投影方法。该方法定义了三组具有明确物理含义的初始方向向量（分别侧重电化学特性、空间几何特性和电子响应特性），并通过正交化构建出用于投影的标准正交基 $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ 。最终的3维化学投影分量 (x, y, z) 由7维化学向量 \mathbf{v}_{chem} 与这些正交基的内积给出：

$$x = \mathbf{v}_{\text{chem}} \cdot \mathbf{e}_1$$

$$y = \mathbf{v}_{\text{chem}} \cdot \mathbf{e}_2$$

$$z = \mathbf{v}_{\text{chem}} \cdot \mathbf{e}_3$$

此方法确保了投影的物理可解释性和数学严谨性。最后，计算结构生成元 $\mathbf{M}_{\text{struct}}$ （由 $\mathbf{v}_{\text{struct}}$ 映射）和化学生成元 \mathbf{M}_{chem} （由 \mathbf{v}_{chem} 的3D投影映射）的李代数交换子 $\mathcal{C}(\mathbf{M}_{\text{struct}}, \mathbf{M}_{\text{chem}}) = \mathbf{M}_{\text{struct}}\mathbf{M}_{\text{chem}} - \mathbf{M}_{\text{chem}}\mathbf{M}_{\text{struct}}$ ，最终描述符为此交换子矩阵的弗罗贝尼乌斯范数：

$$\|\mathcal{C}(\mathbf{M}_{\text{struct}}, \mathbf{M}_{\text{chem}})\|_F = \sqrt{\text{Tr}(\mathcal{C}(\mathbf{M}_{\text{struct}}, \mathbf{M}_{\text{chem}})^T \mathcal{C}(\mathbf{M}_{\text{struct}}, \mathbf{M}_{\text{chem}}))}$$

非零的弗罗贝尼乌斯范数表示结构和化学性质在旋转操作下存在“非对易性”或“冲突”，即它们不能通过简单的旋转变换同时对齐。该特征在物理学上量化了晶体中局域几何结构与原子化学本性之间的内在矛盾或驱动力。高不相容性值可能预示着局域应变、结构不稳定或自发畸变，这与晶体材料的相变、极化效应以及缺陷形成能密切相关。它揭示了原子键合环境与原子自身电子亲和力、电离能等化学属性之间的不匹配程度，高不相容性可能促使原子偏离理想晶格位置以适应其化学键合偏好，从而驱动结构重排或相变。在材料学中，直接影响材料的结构稳定性、铁电性、压电性和相变行为。例如，在钙钛矿材料中，B位阳离子的结构-化学不相容性可能是其铁电性的重要来源。该特征为设计具有特定结构畸变或功能响应的材料提供了新的指导原则。在计算机与人工智能学科中，作为一个连续数值特征，它提供了高维度化学信息与几何结构之间深层相互作用的抽象量化，并通过结合李代数和范数理论，能够捕捉传统描述符难以表达的结构-电子耦合信息，是构建可解释性材料设计模型的关键，其物理驱动的构建方式赋予了模型更强的泛化能力和物理洞察力。

3.5.2 化学矢量在结构度量下的范数 (chemical_in_structural_metric)

chemical_in_structural_metric 描述符严格定义为原子的3D化学投影矢量 $\mathbf{v}_{\text{chem,proj}}$ 在其局域环境结构张量 $\mathbf{T}_{\text{struct},i}$ 作为度规张量下的二次型，旨在量化原子化学性质在局域结构空间中的“伸展”或“阻尼”程度。其数学表达式为：

$$\text{Norm}_{\mathbf{T}_{\text{struct},i}}(\mathbf{v}_{\text{chem,proj}}) = \mathbf{v}_{\text{chem,proj}}^T \mathbf{T}_{\text{struct},i} \mathbf{v}_{\text{chem,proj}}$$

其中， $\mathbf{v}_{\text{chem,proj}}$ 是7维化学矢量通过3.5.1节中描述的Gram-Schmidt投影方法获得的3D投影向量，而 $\mathbf{T}_{\text{struct},i}$ 是原子局域环境的结构张量（定义见3.4.2节）。这个二次型提供了一个直观的几何解释：结构张量定义了局域空间中的“距离”或“形状”，而该特征则衡量了化学矢量在这种由原子排列所定义的“扭曲”空间中的“长度”或“能量”。该计算包含了对结构张量进行正则化处理（通过添加小的单位矩阵项）以确保其正定性，从而提高数值稳定性，使其能够可靠地作为度规张量。在物理学上，该特征反映了原子化学本性与局域几何结构之间的“耦合强度”。高值可能意味着原子化学偏好与结构几何高度匹配，或化学性质在结构空间中被“放大”，导致更强的相互作用。它揭示了原子化学势场如何与局域原子排列相互作用。例如，具有特定化学性质的原子可能倾向于占据某种几何构型的位点，使得在此结构度规下其化学矢量的范数最大化，从而实现能量上的稳定。在材料学中，该特征直接影响材料的稳定性、固溶度、原子扩散路径和相变行为。高耦合强度可能意味着原子在该位点上具有更高的结构适应性，它为优化材料组分和结构、预测合金化行为提供了新的指导。在计算机与人工智能学科中，作为一个连续数值特征，它提供了高维度化学信息在结构空间中的“度量”，实现了多物理量深层耦合的特征构建。该特征通过引入结构张量作为度规，使机器学习模型能够更好地理解化学信息在特定几何环境中的表现，是构建多物理场耦合模型的关键。

3.5.3 电荷加权的局域尺寸 (charge_weighted_local_size)

charge_weighted_local_size 描述符严格定义为电荷加权的局域结构张量 $\mathbf{T}_{\text{charge_weighted},i}$ 的迹。这个张量的构造方式与常规结构张量相似，但在计算每个近邻位移向量的外积时，引入了近邻原子的Bader电荷绝对值（取自3.3.1节）作为权重：

$$\mathbf{T}_{\text{charge_weighted},i} = \sum_{j \in \text{NN}(i)} |q_j| (\mathbf{r}_j - \mathbf{r}_i) \otimes (\mathbf{r}_j - \mathbf{r}_i)$$

最终的charge_weighted_local_size为该加权张量的迹：

$$\text{Charge Weighted Local Size} = \text{Tr}(\mathbf{T}_{\text{charge_weighted},i}) = \sum_{j \in \text{NN}(i)} |q_j| \|\mathbf{r}_j - \mathbf{r}_i\|^2$$

这一“深度融合”范式意味着电荷不再是事后的调制因子，而是在张量构造之初就参与进来的权重，代表了一个“电荷相互作用感知下”的局域尺寸度量。在物理学上，该特征反映了原子局域环境的“有效”空间尺寸，其中电荷较大的近邻对尺寸的贡献更大。这与局域电荷分布的不均匀性、离子键强度和静电相互作用范围密切相关，能够更准确地反映电荷对结构空间的影响。它揭示了电荷转移和离子性对局域几何结构的影响。在离子键为主的晶体中，该特征会突出带有高净电荷的离子对局域尺寸的贡献，从而更精确地捕捉电荷-几何的相互作用。在材料学中，直接影响材料的离子迁移率（在固体电解质中，电荷加权的尺寸可能与离子扩散通道的“有效”大小相关）、介电响应和缺陷形成能。例如，较大的电荷加权尺寸可能指示离子间存在较强的库仑斥力，从而影响晶格稳定性。在计算机与人工智能学科中，作为一个连续数

值特征，它通过将量子信息（电荷）内禀融合到几何张量中，实现了多物理量深层耦合的特征构建。这使得模型能够学习到电荷分布对局域结构尺寸的精细影响，提高了特征的物理可解释性和预测能力。

3.5.4 ELF加权的局域不对称性 (elf_weighted_local_anisotropy)

elf_weighted_local_anisotropy 描述符严格定义为原子中心处的电子局域化函数（ELF）值（取自第3.3.4节）与其矢量不对称性模方（取自第3.4.1节）的乘积。它直接将电子结构的局域化特性与原子环境的几何非对称性联系起来。其数学表达式为：

$$\text{ELF Weighted Local Asymmetry} = \text{ELF}_i \times \|\mathbf{V}_{\text{asymmetry},i}\|^2$$

其中 ELF_i 是原子 i 处的ELF值， $\|\mathbf{V}_{\text{asymmetry},i}\|^2$ 是原子 i 的矢量不对称性模方。该特征的物理图像是：在电子高度局域化（高ELF值）的区域，其几何不对称性（由矢量不对称性模方衡量）的“权重”更大。这意味着这种结合了电子局域化信息的几何不对称性，对材料的物理化学性质和功能性的影响可能更为显著。在物理学上，该特征直接量化了电子局域化对局域结构不对称性的强化效应。高值表明在电子高度局域化的区域存在显著的几何不对称性，这可能与局域电偶极子的形成、非中心对称性引起的能带结构变化以及光电响应密切相关。它揭示了电子成键模式（共价性、孤对电子）与局域几何畸变之间的协同作用。例如，在共价性较强的键合环境中，即使是很小的几何不对称性也可能产生显著的局域极化。在材料学中，该特征直接影响晶体材料的压电性、铁电性、非线性光学效应以及光伏性能。高ELF加权不对称性可能预示着材料具有优异的极性相关功能，它为理解和设计具有特定光电、压电性质的功能材料提供了新的视角。在计算机与人工智能学科中，作为一个连续数值特征，它通过将电子结构信息（ELF）与几何信息（矢量不对称性模方）进行深度融合，实现了物理驱动的特征工程。这使得机器学习模型能够更有效地捕捉电子局域化对几何不对称性功能影响的复杂相互作用，从而提高模型对材料功能特性（如极性、光响应）的预测精度和可解释性。

第四章：1-单纯形描述符：化学键的本质量化与相互作用表征

1-单纯形描述符聚焦于晶体材料中的化学键（原子对）及其直接环境，旨在从几何、物理化学和量子力学层面精确量化键的本质属性和它如何促成原子间的相互作用。本章通过对键长、配位数、电荷转移、电子局域化等关键属性的深入分析，构建了一个多维度、高内涵的键级描述体系。这些描述符是理解晶体结构稳定性、构效关系以及预测材料功能特性的关键桥梁。

4.1 几何描述符：键的长度与配位环境

本节主要关注化学键最直接的几何属性——键长，以及键两端原子的局域配位情况。这些特征为理解晶体材料中的空间排列和原子间的基本几何约束提供了基础。

4.1.1 键长 (bond_distance)

键长 d_{ij} 严格定义为连接两个成键原子 i 和 j 的欧几里得距离。若原子 i 的坐标为 \mathbf{r}_i 且原子 j 的坐标为 \mathbf{r}_j ，则键长 d_{ij} 的计算公式为：

$$d_{ij} = \|\mathbf{r}_j - \mathbf{r}_i\|_2 = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}$$

键长作为原子间相互作用的直接空间尺度，是化学键强度和键合类型的重要物理表征；较短的键长通常对应更强的键合。在化学层面，它与共价半径、离子半径以及键合势能密切相关，是判断成键类型（共价、离子、金属）和预测分子稳定性的关键参数，例如，短而强的共价键通常具有较高的键能和方向性。在材料学中，键长直接影响晶体材料的晶格常数、密度、热膨胀系数和机械强度，其微小变化可能导致材料宏观性能的显著差异，例如在半导体中，键长的变化会影响电子能带结构。在计算机与人工智能领域，bond_distance作为最基本的几何特征，是构建复杂高阶描述符的基石，它提供了低维、直接可计算的输入，在材料性质预测的机器学习模型中具有广泛应用，易于理解且无需复杂的特征工程。

4.1.2 键端原子配位数 (site1_coord_num, site2_coord_num)

site1_coord_num 和 site2_coord_num 描述符严格定义为化学键两端原子 i 和 j 的配位数，即原子在晶体结构中直接连接的近邻原子数量。其计算依赖于pymatgen.analysis.local_env.CrystalNN策略，该策略通过基于键距离和化学环境的智能判断来确定配位。配位数的数学表达式为：

$$\text{CoordNum}_k = |\text{NN}(k)|$$

其中 $\text{NN}(k)$ 表示原子 k 的近邻原子集合。这些值从第4.2.8节中预计算的原子配位数列表中获取。CrystalNN算法通过分析原子间的距离、元素性质以及周期性边界条件来识别化学键，并统计与中心原子直接相连的原子数量。这是一个结合了几何判断和化学规则的复杂过程，旨在模拟真实的化学键合环境。在物理、化学与材料学层面，配位数直接反映了原子在晶体材料晶格中的空间填充和局域相互作用强度。它是判断原子价态、键合类型和局域几何构型的核心指标，直接影响配位多面体的形状和连接方式，进而影响材料的密度、稳定性、电子结构（如晶体场分裂）和许多宏观物理性能（如硬度、熔点、离子迁移率）。配位数的改变常与相变或缺陷形成相关。在计算机与人工智能中，作为离散但具有明确物理意义的特征，配位数可以作为图神经网络中节点特征的一部分，或作为传统机器学习模型中的分类或回归特征，为模型理解晶体网络的拓扑特性提供了基本描述。

4.2 派生与关联描述符：基于0-单纯形特征的键级信息整合

本节介绍的描述符通过聚合或差异化化学键两端原子（0-单纯形）的预计算特征而得到，旨在捕捉键合过程中原子属性的相互作用和信息传递。

4.2.1 电负性差 (delta_electronegativity)

delta_electronegativity描述符严格定义为化学键两端原子 i 和 j 电负性 χ 的绝对差值。其数学表达式为：

$$\Delta\chi_{ij} = |\chi_j - \chi_i|$$

其中原子电负性数据来源于第4.2.2节。

电负性差值在物理上是衡量键合离子性程度的直接指标，较大的差值通常预示着更强的离子键特性和电子从一个原子向另一个原子的转移倾向。化学上，它直接与化学键的极性、偶极矩和材料的介电性质相关；在化学反应中，电负性差值高的键通常是反应的活性位点。材料学上，此特征影响材料的极性、介电常数、压电性以及离子导电性，对于离子导体而言，较大的电负性差可以促进离子在晶格中的迁移。在计算机与人工智能领域，作为衡量键合极性的重要特征，它为机器学习模型提供了关于键合本质的直观信息，在预测热力学稳定性、材料带隙或表面活性时具有重要作用。

4.2.2 离子半径差 (delta_ionic_radius)

delta_ionic_radius描述符严格定义为化学键两端原子 i 和 j 离子半径 r^{ion} 的绝对差值。其数学表达式为：

$$\Delta r_{ij}^{\text{ion}} = |r_j^{\text{ion}} - r_i^{\text{ion}}|$$

其中原子离子半径数据来源于第4.2.6节。

离子半径差值在物理上直接反映了键合原子空间尺寸的匹配程度，较大的差值可能导致结构畸变或应力。在化学上，对于离子晶体，离子半径比是判断晶体结构类型和稳定性的关键因素（如鲍林规则），同时它也影响离子键的强度和方向性。材料学上，此特征与晶格畸变、缺陷形成能、相变行为以及离子迁移率密切相关；例如，在钙钛矿晶体材料中，离子半径的适当匹配对于形成稳定的ABX₃结构至关重要。在计算机与人工智能领域，离子半径差提供了关于原子尺寸匹配性的信息，这对于预测材料结构稳定性、可合成性以及某些离子传输性质的模型而言是重要的输入。

4.2.3 共价半径和 (sum_covalent_radii)

sum_covalent_radii描述符严格定义为化学键两端原子 i 和 j 共价半径 $r^{\text{Covalent Radii}}$ 之和。其数学表达式为：

$$\Sigma r_{ij}^{\text{Covalent Radii}} = r_i^{\text{Covalent Radii}} + r_j^{\text{Covalent Radii}}$$

其中原子共价半径数据来源于第4.2.7节。

共价半径和在物理上通常被视为理想共价键键长的近似值，它提供了一个基准，用于评估实际键长是受到压缩还是拉伸。在化学上，它直接与共价键的性质和原子间的重叠积分相关；在共价晶体中，共价半径和可以帮助预测键合的强度和方向性。材料学上，此特征影响晶体材料的晶体结构、机械性能（如硬度、弹性模量）和热膨胀；在共价主导的材料中，共价半径和对于理解键合网络和其稳定性至关重要。在计算机与人工智能领域，sum_covalent_radii可以与实际键长结合使用，以形成键长畸变等更高阶特征，为模型提供关于键合环境的精细信息。

4.2.4 平均Bader电荷 (avg_bader_charge)

avg_bader_charge描述符严格定义为化学键两端原子 i 和 j Bader电荷 Q^{Bader} 的算术平均值。其数学表达式为：

$$\overline{Q}_{ij}^{\text{Bader}} = \frac{Q_i^{\text{Bader}} + Q_j^{\text{Bader}}}{2}$$

其中原子Bader电荷数据来源于第3.3.1节。

平均Bader电荷在物理上反映了键合原子对整体电荷的平均贡献，是衡量键合环境平均电荷状态的指标。在化学上，它与平均氧化态、键合的总电子数以及材料的化学稳定性相关，对于理解一个键合区域的整体电子富集或亏缺状态非常有用。材料学上，此特征影响晶体材料的导电性、光学性质以及化学反应性，例如，在一个键合链上的平均电荷可以指示电子传输路径。在计算机与人工智能领域，avg_bader_charge提供了键合区域整体电子环境的统计信息，有助于模型理解键合的电子性质。

4.2.5 Bader电荷差 (delta_bader_charge)

delta_bader_charge描述符严格定义为化学键两端原子 i 和 j Bader电荷 Q^{Bader} 的绝对差值。其数学表达式为：

$$\Delta Q_{ij}^{\text{Bader}} = |Q_j^{\text{Bader}} - Q_i^{\text{Bader}}|$$

其中原子Bader电荷数据来源同上。

Bader电荷差在物理上直接反映了键合原子间的电荷转移强度和电荷分布的不均匀性，是键合离子性的重要量化指标；较大的差值意味着更显著的电荷极化。在化学上，它与键的极性、偶极矩（尽管自身不直接给出方向）、离子键成分以及化学键的稳定性相关，是理解电子如何在键中重新分布的关键参数。材料学上，此特征影响晶体材料的介电性能、铁电性、压电性以及电子/离子传输路径，在极性材料中，Bader电荷差是衡量其极化强度的重要指标。在计算机与人工智能领域，delta_bader_charge与电负性差互补，提供了更直接的电荷分布不均匀性信息，对于预测极性材料性能以及指导新材料设计具有重要价值。

4.3 量子化学成键描述符：键中电子结构的微观洞察

本节的描述符直接来源于密度泛函理论（DFT）第一性原理计算的电子密度、ELF（电子局域函数）和Bader电荷等量子物理量。它们深入揭示了化学键内部的电子结构细节，是理解晶体材料微观成键机制和宏观量子效应的基石。

4.3.1 键中点电子密度 (bond_midpoint_density)

bond_midpoint_density描述符严格定义为化学键中点 \mathbf{r}_{mid} 处的全电子密度 $\rho(\mathbf{r}_{\text{mid}})$ 。键中点是连接两原子核的线段的几何中点。电子密度 $\rho(\mathbf{r})$ 衡量了在空间某一点找到电子的概率，其值越高表示该区域电子越富集。其数学表达式为：

$$\rho(\mathbf{r}_{\text{mid}}) = \rho\left(\frac{\mathbf{r}_i + \mathbf{r}_j}{2}\right)$$

键中点电子密度在物理上是衡量共价键强度和多重性的直接指标，共价键通常在键合区域（尤其是键中点）表现出电子密度的显著富集。在化学上，它与键级、键的共价性质以及原子间电子共享的程度密切相关，高键中点密度通常意味着强的共价相互作用。材料学上，此特征影响晶体材料的机械强度、硬度、热导率以及电子传输性质，例如，高键中点密度在金刚石等共价晶体中表现出卓越的机械性能。在计算机与人工智能领域，bond_midpoint_density提供了化学键最核心的电子结构信息，是机器学习模型预测共价主导材料性质的关键输入，其连续性也使其适合基于梯度优化的模型。

4.3.2 键中点电子密度拉普拉斯算子 (bond_density_laplacian)

bond_density_laplacian描述符严格定义为化学键中点 \mathbf{r}_{mid} 处的电子密度拉普拉斯算子 $\nabla^2\rho(\mathbf{r}_{\text{mid}})$ 。拉普拉斯算子 $\nabla^2\rho(\mathbf{r})$ 衡量了电子密度在空间某一点的局部汇聚或发散趋势。根据QTAIM理论，

$\nabla^2\rho(\mathbf{r}) < 0$ 表示电子在 \mathbf{r} 处汇聚（键合区域），而 $\nabla^2\rho(\mathbf{r}) > 0$ 表示电子在 \mathbf{r} 处发散（核区域）。其数学表达式为：

$$\nabla^2\rho(\mathbf{r}_{\text{mid}}) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \rho(\mathbf{r}) \Big|_{\mathbf{r}=\mathbf{r}_{\text{mid}}}$$

键中点电子密度拉普拉斯算子在物理上是键合类型判别的强大工具，在共价键中，键临界点（Bond Critical Point, BCP，即 $\nabla\rho = 0$ 的点）处的拉普拉斯算子通常为负值，表示电子在键合区域内共享；而在离子键中，该值可能为正或接近零，表示电荷累积在原子核周围。在化学上，它与键合的共价性/离子性、键路径的存在性以及化学反应活性直接相关，提供了比简单电子密度更丰富的拓扑信息。材料学上，此特征影响晶体材料的键合刚度、机械响应以及电子局域化程度，例如，键中点负的拉普拉斯值指示了更强的共价网络。在计算机与人工智能领域，作为衡量键合拓扑特征的精确量子量，它为机器学习模型提供了更深层次的键合本质信息，有助于模型区分不同键合类型，从而提升材料性能预测的准确性。

4.3.3 键中点ELF (bond_midpoint_elf)

bond_midpoint_elf描述符严格定义为化学键中点 \mathbf{r}_{mid} 处的电子局域函数 ELF(\mathbf{r}_{mid})。ELF是一个介于0和1之间的无量纲函数，量化了电子在该区域的局域化程度。高ELF值（接近1）表示强电子局域化（如共价键或孤对电子），低ELF值（接近0.5）表示非局域化（如金属键），极低ELF值（接近0）表示电子亏缺。其数学表达式为：

$$\text{ELF}(\mathbf{r}) = \frac{1}{1 + (\mathcal{D}(\mathbf{r})/D_0(\mathbf{r}))^2}$$

其中 $\mathcal{D}(\mathbf{r})$ 是电子动能密度， $D_0(\mathbf{r})$ 是均匀电子气的动能密度参考。高ELF值（接近1）表示强电子局域化（如共价键或孤对电子），低ELF值（接近0.5）表示非局域化（如金属键），极低ELF值（接近0）表示电子亏缺。

键中点ELF在物理上直接可视化三体团簇核心区域的电子局域化特性，帮助识别共价键、离子键或金属键在多体环境中的表现。化学上，它用于判断多中心键合的强度和类型，例如在金属间化合物或分子簇中的复杂成键。材料学上，此特征影响材料的局部键合刚度、电子传输特性、介电响应以及声子散射；例如，在半导体中，高ELF值区域通常是电子传输的障碍。在计算机与人工智能领域，它提供了直观且量化的三体区域电子局域化信息，是构建能够区分复杂键合类型、预测材料多体相互作用和功能特性的机器学习模型的关键特征。

4.3.4 键端ELF均值 (bond_elf_from_0simplex)

bond_elf_from_0simplex描述符严格定义为化学键两端原子 i 和 j 的电子局域化函数（ELF）值的算术平均值。该特征直接从0-单纯形特征（参见第3.3.4节）聚合而来，提供了键合区域整体电子局域化程度

的概括性描述。其数学表达式为：

$$\overline{\text{ELF}}_{ij} = \frac{\text{ELF}_i + \text{ELF}_j}{2}$$

其中 ELF_i 和 ELF_j 是原子 i 和 j 在其位点处的ELF值。

在物理、化学与材料学层面，该特征反映了化学键整体电子局域化情况。高均值表示键合区域电子普遍高度局域化，可能对应于强共价键或离子键；低均值则可能指示更离域的电子特性（如金属键）。它有助于理解键的平均成键性质。在计算机与人工智能领域，作为一个连续数值特征，它提供了对键合区域电子局域化程度的简单而有效的量化，可用于预测材料的键合类型、电学性质和稳定性。

4.3.5 键端ELF不对称性 (bond_elf_asymmetry)

bond_elf_asymmetry描述符严格定义为化学键两端原子 i 和 j 的电子局域化函数（ELF）值的绝对差值。该特征从0-单纯形特征（参见第3.3.4节）派生，量化了键两端电子局域化程度的差异。其数学表达式为：

$$\Delta\text{ELF}_{ij} = |\text{ELF}_j - \text{ELF}_i|$$

其中 ELF_i 和 ELF_j 是原子 i 和 j 在其位点处的ELF值。

在物理、化学与材料学层面，该特征反映了化学键两端电子局域化程度的差异性或不均匀性。高不对称性值可能指示键合极性较强，或键两端原子在共享电子方面的能力存在显著差异。它有助于识别键中电子局域化的强度分布。在计算机与人工智能领域，作为一个连续数值特征，它提供了对键合区域电子局域化不均匀性的量化，对于预测极性材料的性质、电荷传输路径和化学反应活性位点具有重要意义。

4.3.6 键端电子密度均值 (bond_density_from_0simplex)

bond_density_from_0simplex描述符严格定义为化学键两端原子 i 和 j 的电子密度值（取自其各自原子位点，参见第3.3.3节）的算术平均值。它提供了键合区域整体电子丰度的概括性度量。其数学表达式为：

$$\bar{\rho}_{ij} = \frac{\rho_i + \rho_j}{2}$$

其中 ρ_i 和 ρ_j 是原子 i 和 j 在其位点处的电子密度值。

在物理、化学与材料学层面，该特征反映了化学键连接的两个原子周围的平均电子丰度。高均值表示键合环境电子普遍富集，可能与较强的成键相互作用有关。它有助于理解键的平均电子环境。在计算机与

人工智能领域，作为一个连续数值特征，它提供了对键合区域电子丰度的简单量化，可用于预测材料的键合强度、密度和电子性质。

4.3.7 键电子密度梯度 (bond_density_gradient)

bond_density_gradient描述符严格定义为化学键两端原子 i 和 j 电子密度值（取自其各自原子位点，参见第3.3.3节）的绝对差值除以键长。这量化了沿着化学键方向的电子密度变化率，反映了键中的电荷极化或共享程度。其数学表达式为：

$$\text{Density Gradient}_{ij} = \frac{|\rho_j - \rho_i|}{d_{ij}}$$

其中 ρ_i 和 ρ_j 是原子 i 和 j 在其位点处的电子密度值， d_{ij} 是键长（参见第4.1.1节）。当键长接近零时，为避免除零错误，分母会加上一个小的容差值。

在物理、化学与材料学层面，该特征反映了沿着化学键的电子密度斜坡或梯度。高梯度表示电荷从一个原子向另一个原子显著转移或键合极性很强。它有助于识别电荷传输路径和极性键的强度。在计算机与人工智能领域，作为一个连续数值特征，它提供了对键合区域电子密度变化率的量化，对于预测材料的电荷输运性能、介电性质和化学反应活性位点具有重要意义。

4.3.8 键有效电荷 (bond_effective_charge)

bond_effective_charge描述符严格定义为化学键两端原子 i 和 j 的Bader电荷值（参见第3.3.1节）的算术平均值。这与第4.2.4节的平均Bader电荷是数值上相同的特征，但在本节中再次强调其作为量子化学分析中对键合体系整体电荷状态的直接反映。其数学表达式为：

$$\overline{Q}_{ij}^{\text{Bader}} = \frac{Q_i^{\text{Bader}} + Q_j^{\text{Bader}}}{2}$$

其中 Q_i^{Bader} 和 Q_j^{Bader} 是原子 i 和 j 的Bader电荷。

在物理、化学与材料学层面，该特征反映了键合原子对整个键合区域平均电荷贡献。它有助于理解键的整体电子环境和键合的离子性/共价性平衡。在计算机与人工智能领域，作为一个连续数值特征，它提供了对键合区域电荷状态的量化，可用于预测材料的电学性能和化学稳定性。

4.3.9 键电荷不平衡 (bond_charge_imbalance)

bond_charge_imbalance描述符严格定义为化学键两端原子 i 和 j 的Bader电荷之差相对于其绝对值之和的归一化值。它量化了键合原子间的电荷不对称性或不平衡程度。其数学表达式为：

$$\text{Charge Imbalance}_{ij} = \frac{Q_j^{\text{Bader}} - Q_i^{\text{Bader}}}{|Q_i^{\text{Bader}}| + |Q_j^{\text{Bader}}|}$$

其中 Q_i^{Bader} 和 Q_j^{Bader} 是原子 i 和 j 的Bader电荷。分母添加一个小的容差值以避免除零错误。

在物理、化学与材料学层面，该特征反映了键中电子分布的极化程度。非零值表示电荷沿键方向存在净转移，可能与偶极子形成、局域极化和压电性相关。它有助于识别键合中的电荷转移方向和强度。在计算机与人工智能领域，作为一个连续数值特征，它提供了对键合区域电荷不平衡的归一化量化，对于预测极性材料的性质、电荷传输路径和电化学活性位点具有重要意义。

4.4 深度融合代数描述符：化学键复杂相互作用的抽象表征

本节的描述符是本章的核心创新点，它超越了传统的几何和量子描述，开创性地融合了五大核心代数思想，旨在捕捉化学键在高阶几何、拓扑、对称性和场-物质耦合中的复杂相互作用。这些特征提供了连续、可微且具有深刻物理可解释性的数学工具，对于构建能够洞察晶体材料“基因组”深层规律的可解释人工智能模型至关重要。

4.4.1 几何环境不相容性 (lie_algebra_incompatibility)

lie_algebra_incompatibility描述符严格定义为化学键两端原子 i 和 j 的局部环境结构张量 $\mathbf{T}_{\text{struct},i}$ 和 $\mathbf{T}_{\text{struct},j}$ 在旋转群的李代数 $\mathfrak{so}(3)$ 上的交换子（Commutator）的弗罗贝尼乌斯范数。这个特征通过李代数交换子 $\mathcal{C}(\mathbf{T}_{\text{struct},i}, \mathbf{T}_{\text{struct},j}) = \mathbf{T}_{\text{struct},i}\mathbf{T}_{\text{struct},j} - \mathbf{T}_{\text{struct},j}\mathbf{T}_{\text{struct},i}$ 来量化两个局部几何环境的非交换性或不可同构性。其数学表达式为：

$$\text{Incomp}_{ij} = \|\mathcal{C}(\mathbf{T}_{\text{struct},i}, \mathbf{T}_{\text{struct},j})\|_F = \sqrt{\text{Tr}((\mathcal{C}(\mathbf{T}_{\text{struct},i}, \mathbf{T}_{\text{struct},j}))^T \mathcal{C}(\mathbf{T}_{\text{struct},i}, \mathbf{T}_{\text{struct},j}))}$$

其中 $\mathbf{T}_{\text{struct},k}$ 是原子 k 的局部环境结构张量，该张量来源于第3.4.2节。

在物理上，该描述符量化了键两端原子局部环境的几何结构匹配程度和应力状态。非零的不相容性值表明两个环境无法通过简单的线性变换完全对齐，预示着晶格内部存在剪切应力、扭曲或畸变，这类似于晶体学中的应变张量或位错核。化学上，它与局部化学键合环境的失配和键合应力累积相关，高不相容性可能导致键能降低或不稳定。材料学上，lie_algebra_incompatibility是理解晶体缺陷形成、相变路径、机械性能（如脆性、韧性）以及输运性质（如离子迁移受阻）的关键；在特定晶体中，如钙钛矿，八面体畸变导致的不相容性直接影响其光电性能。在计算机与人工智能领域，该特征提供了对晶体局部结构应变和几何拓扑不匹配连续化、可微量化，是构建能够预测材料结构稳定性、缺陷形成能以及机械响应的机器学习模型的强大特征。其数学严谨性使其具有高泛化能力和可解释性。

4.4.2 键方向商代数特征：轨道大小 (quotient_algebra_orbit_size)

该特征基于商代数理论（或群作用的轨道理论），旨在量化化学键的方向在晶体对称群作用下的对称等价程度。

quotient_algebra_orbit_size：严格定义为键方向向量 \mathbf{v}_{bond} 在晶体对称群 G （具体而言，在计算中通常指晶体的空间群所包含的旋转和反射操作，即其点群操作的集合）作用下所形成的轨道（Orbit） $O(\mathbf{v}_{\text{bond}}) = \{g \cdot \mathbf{v}_{\text{bond}} \mid g \in G\}$ 的大小。

轨道（Orbit）在群论中是一个核心概念，它表示在一个群 G 对一个集合 X （这里是三维空间中的键方向向量）作用下，集合 X 中所有通过群操作可以互相转换的元素所构成的等价类。具体到化学键方向，如果两个键方向可以通过晶体的某个对称操作相互转换，那么它们就属于同一个轨道。

其数学表达式为：

$$\text{OrbitSize}(\mathbf{v}_{\text{bond}}) = |O(\mathbf{v}_{\text{bond}})| = |\{g \cdot \mathbf{v}_{\text{bond}} \mid g \in G\}|$$

其中：

\mathbf{v}_{bond} 是经过归一化的化学键方向向量； G 是晶体的对称群（通常为点群或其相关旋转操作的集合），其元素 g 代表一个对称操作（如旋转、反演）； $g \cdot \mathbf{v}_{\text{bond}}$ 表示对称操作 g 作用于键方向向量 \mathbf{v}_{bond} 后的新向量； $|\cdot|$ 表示集合中元素的数量。

轨道大小反映了该键方向在晶体中的对称性独特性：轨道越大（即包含的独特方向越多），表示该方向在对称性上越不特殊，有更多等价方向；轨道越小，则表示其对称性越高或越独特（例如，沿晶格主轴的方向通常具有较小的轨道，因为只有少数操作能将其映射到不同的方向，甚至可能只有它自身）。在计算实现中，我们遍历所有对称操作，将键方向向量进行变换，并收集所有数值上独特的变换结果，其数量即为轨道大小。

4.4.3 伪辛耦合 (pseudo_symplectic_coupling)

pseudo_symplectic_coupling描述符受辛几何理论启发，旨在量化化学键的几何方向与该键中点处电子密度梯度之间的深层耦合关系。尽管严格意义上的辛几何通常处理偶数维相空间，但此特征通过将键的方向向量视为广义坐标 \mathbf{q} ，电子密度梯度视为广义动量 \mathbf{p} ，构建了一个类比的相空间。选择电子密度梯度作为广义动量的物理依据在于，它不仅量化了电子密度的空间变化率，更深刻地反映了局域电子所受的有效力场和电荷传输的潜在驱动力。在量子力学中，力的概念与动量密切相关，电子密度梯度在空间中的分布可以被视为一种势能斜率或驱动力场，其方向指示了电子从高密度区域流向低密度区域的趋势，这与经典力学中动量驱动系统演化的物理内涵具有概念上的相似性。在此基础上定义了一个复合不变量，以捕捉两者之间的相互作用强度。此特征的值越大，表示键方向与电子密度的局域变化趋势之间存在越强的卷曲或耦合效应，预示着独特的物理化学行为。

此特征的计算核心在于定义一个伪辛不变量 $S(\mathbf{q}, \mathbf{p})$ ，它综合了哈密顿量、辛形式、角动量和相空间尺度等概念。首先，对键方向向量 bond_direction 和键中点电子密度梯度 density_gradient 进行归一化处理，得到单位向量 $\hat{\mathbf{q}}$ 和 $\hat{\mathbf{p}}$ 。接着，伪辛不变量 $S(\hat{\mathbf{q}}, \hat{\mathbf{p}})$ 通过以下复合公式计算：

$$H(\hat{\mathbf{q}}, \hat{\mathbf{p}}) = \frac{1}{2} \|\hat{\mathbf{p}}\|^2 + \frac{1}{2} \|\hat{\mathbf{q}}\|^2$$

$$\omega(\hat{\mathbf{q}}, \hat{\mathbf{p}}) = \hat{\mathbf{q}} \cdot \hat{\mathbf{p}}$$

$$L(\hat{\mathbf{q}}, \hat{\mathbf{p}}) = \|\hat{\mathbf{q}} \times \hat{\mathbf{p}}\|$$

$$\text{PS_Scale}(\hat{\mathbf{q}}, \hat{\mathbf{p}}) = \sqrt{\|\hat{\mathbf{q}}\|^2 + \|\hat{\mathbf{p}}\|^2}$$

$$S(\hat{\mathbf{q}}, \hat{\mathbf{p}}) = H(\hat{\mathbf{q}}, \hat{\mathbf{p}}) + 0.1 \cdot |\omega(\hat{\mathbf{q}}, \hat{\mathbf{p}})| + 0.05 \cdot L(\hat{\mathbf{q}}, \hat{\mathbf{p}}) + 0.01 \cdot \text{PS_Scale}(\hat{\mathbf{q}}, \hat{\mathbf{p}})$$

其中， $H(\hat{\mathbf{q}}, \hat{\mathbf{p}})$ 类比于系统总能量， $\omega(\hat{\mathbf{q}}, \hat{\mathbf{p}})$ 是键方向与密度梯度的点积（代表辛形式的直接类比）， $L(\hat{\mathbf{q}}, \hat{\mathbf{p}})$ 是两向量叉积的模（代表角动量的类比），而 $\text{PS_Scale}(\hat{\mathbf{q}}, \hat{\mathbf{p}})$ 反映了相空间的整体尺度。这些组成部分的权重是启发式的，旨在平衡各物理量的贡献。

在物理上，`pseudo_symplectic_coupling` 量化了化学键几何结构（方向）与键中电子密度变化（梯度）之间的深层相互作用。高值表明键方向与电子流或电荷极化方向高度耦合，这在理解电子传输、光电响应和非线性光学性质中至关重要。化学上，这种耦合反映了键合过程中电子分布对空间几何排布的敏感性，有助于解释特定键的稳定性或反应活性。材料学上，该特征可以作为预测材料导电性、介电常数、压电性等功能属性的指标，特别是在钙钛矿等具有显著结构-电子耦合效应的材料中，它能揭示微观结构畸变如何影响电子行为。在计算机与人工智能领域，`pseudo_symplectic_coupling` 提供了对结构-电子耦合的连续、可微且具有深刻物理内涵的量化，是构建能够预测复杂功能材料性能、优化材料设计的机器学习模型的强大特征，有助于从白箱角度理解材料的内在机制。

4.4.4 张量代数环境对齐 (`tensor_algebraic_environment_alignment`)

`tensor_algebraic_environment_alignment` 描述符严格定义为化学键两端原子 i 和 j 的局部环境结构张量 $\mathbf{T}_{\text{struct},i}$ 和 $\mathbf{T}_{\text{struct},j}$ 之间的张量对齐度。此特征旨在量化这两个局部几何环境在空间取向和形状上的相似性或一致性。高对齐度表明键两端原子所处的局部结构环境在几何上高度协调，而低对齐度则预示着结构的不匹配或应力。

此对齐度的计算综合了多种张量相似性度量，以提供一个稳健的综合指标。具体而言，它结合了归一化张量内积、主方向对齐度以及张量余弦相似度。其计算步骤如下：

1. 归一化张量内积：首先计算两个结构张量的弗罗贝尼乌斯范数 $\|\mathbf{T}\|_F = \sqrt{\text{Tr}(\mathbf{T}^T \mathbf{T})}$ 。然后，通过计算归一化张量 $\mathbf{T}_{\text{norm},k} = \mathbf{T}_k / \|\mathbf{T}_k\|_F$ 之间的迹内积 $\text{Tr}(\mathbf{T}_{\text{norm},i}^T \mathbf{T}_{\text{norm},j})$ 来衡量其整体相似性。

2. 主方向对齐度：通过计算两个结构张量各自的特征值分解，提取其具有最大特征值的主方向向量。然后计算这两个主方向向量之间夹角的余弦值（绝对值），以评估其主要各向异性方向的对齐程度。
3. 张量余弦相似度：将两个张量视为高维向量，计算它们之间的余弦相似度，即 $\frac{\sum_{m,n} (\mathbf{T}_i)_{mn} (\mathbf{T}_j)_{mn}}{\|\mathbf{T}_i\|_F \|\mathbf{T}_j\|_F}$ 。

最终的tensor_algebraic_environment_alignment值是这三个分量（归一化张量内积、主方向对齐度和张量余弦相似度）的加权平均，权重分别为0.4、0.4和0.2，并被裁剪到[0, 1]范围。键端原子的结构张量 $\mathbf{T}_{\text{struct},k}$ 来自于第3.4.2节的均方邻居距离中引入的结构张量概念。

在物理上，该特征提供了键两端局部几何环境匹配程度的连续量化。高对齐度表明局部区域的应力分布和几何形变趋势高度一致，这对于理解晶格应变传播和结构稳定性至关重要。化学上，它反映了相邻原子团簇在形状和取向上的兼容性，有助于预测键合的强度和方向性。材料学上，tensor_algebraic_environment_alignment是理解晶体结构稳定性、相变行为、机械性能（如弹性模量、硬度）以及晶界和缺陷形成的关键，在设计具有特定机械或热力学性能的材料时具有指导意义。在计算机与人工智能领域，此特征为机器学习模型提供了对复杂局部结构几何兼容性的高阶描述，有助于模型预测材料的宏观性能，尤其是在高通量筛选和逆向设计中，能够识别出具有理想局部结构匹配性的候选材料。

4.4.5 李代数键对齐 (lie_algebraic_bond_alignment_a, lie_algebraic_bond_alignment_b, lie_algebraic_bond_alignment_c)

lie_algebraic_bond_alignment_a/b/c描述符利用李代数 $\mathfrak{so}(3)$ 的伴随作用，量化了化学键方向与晶体晶格三个主轴（a, b, c）之间的代数对齐程度。该特征将键方向向量和晶格基向量视为李代数 $\mathfrak{so}(3)$ 中的元素（通过其对应的反对称矩阵表示），并计算它们之间的李代数交换子范数。小范数表示高度对齐（即键方向对晶格轴的对称性破坏较小），大范数则表示对齐度较低或存在显著的对称性破缺效应。这种方法超越了简单的几何夹角，从代数结构上评估了键方向与晶格周期性之间的协调性。

具体而言，对于单位化的键方向向量 $\hat{\mathbf{v}}_{\text{bond}}$ 以及晶格的单位基向量 $\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}$ ，我们构造它们对应的反对称矩阵 $\mathbf{M}_{\text{bond}}, \mathbf{M}_{\text{a}}, \mathbf{M}_{\text{b}}$ 和 \mathbf{M}_{c} 。李代数键对齐度通过计算李代数交换子 $\mathcal{C}(\mathbf{M}_{\text{bond}}, \mathbf{M}_{\text{axis}}) = \mathbf{M}_{\text{bond}}\mathbf{M}_{\text{axis}} - \mathbf{M}_{\text{axis}}\mathbf{M}_{\text{bond}}$ 的弗罗贝尼乌斯范数来衡量，然后通过指数函数将其归一化到[0, 1]范围，使得范数越小，对齐度值越大（接近1）。其核心数学表达式为：

$$\text{Alignment}_{\text{axis}} = \exp(-\|\mathcal{C}(\mathbf{M}_{\text{bond}}, \mathbf{M}_{\text{axis}})\|_F)$$

其中， $\|\cdot\|_F$ 表示弗罗贝尼乌斯范数。

在物理上，lie_algebraic_bond_alignment量化了化学键取向与晶格各向异性之间的深层互作用。高对齐度表明键方向与晶体学主轴高度协调，这在理解材料的各向异性物理性质（如导电性、光学响应、机械

形变) 中至关重要。化学上, 它反映了化学键在晶体环境中的优先取向性, 有助于解释键合在特定方向上的稳定性或强化效应。材料学上, 该特征是理解晶体材料宏观各向异性性能的微观根源, 对于设计具有特定晶体取向依赖性功能 (如压电、铁电、热电) 的材料具有指导意义。在计算机与人工智能领域, 此特征提供了对键方向与晶格对称性耦合的连续、可微量化, 使得机器学习模型能够更准确地预测材料的各向异性行为, 并在逆向设计中探索具有理想晶体学取向的候选结构。

4.4.6 结构-化学不相容性差异 (delta_structure_chemistry_incompatibility)

delta_structure_chemistry_incompatibility描述符严格定义为化学键两端原子 i 和 j 各自的结构-化学不相容性特征的绝对差值。此特征旨在捕捉化学键两端原子在局部结构与化学环境匹配度上的差异。高差异值表明键两端原子所处的局部环境在结构与化学信息的耦合方面存在显著不一致, 这可能预示着键合区域的应力集中或异质性。

该特征的计算直接来源于第3.5.1节中定义的structure_chemistry_incompatibility特征。其数学表达式为:

$$\Delta \text{Incomp}_{ij}^{\text{struct-chem}} = |\text{Incomp}_j^{\text{struct-chem}} - \text{Incomp}_i^{\text{struct-chem}}|$$

其中 $\text{Incomp}_k^{\text{struct-chem}}$ 是原子 k 的结构-化学不相容性。

在物理上, 该特征量化了键两端原子局部结构与电子/化学性质耦合程度的差异。高不相容性差异可能指示键合区域的局部应变、电荷分布不均匀性或几何扭曲。化学上, 它有助于理解键合的异质性以及电子如何在键合区域内重新分配, 这对于解释极性键的形成、缺陷的稳定化以及化学反应活性位点的识别至关重要。材料学上, delta_structure_chemistry_incompatibility是理解材料局部结构畸变、晶界特性、相变机理以及某些功能 (如离子传输、催化) 背后的微观驱动力的关键。在计算机与人工智能领域, 此特征通过对原子层面结构-化学耦合差异的量化, 为机器学习模型提供了更深层次的微观异质性信息, 有助于预测材料的局部应力、缺陷容忍度及其在极端条件下的行为。

4.4.7 键端各向异性失配 (bond_ends_anisotropy_mismatch)

bond_ends_anisotropy_mismatch描述符严格定义为化学键两端原子 i 和 j 各自局部环境各向异性特征的绝对差值。此特征旨在量化键两端原子周围局部几何环境在形状各向异性程度上的差异。高失配值表明键两端原子所处的环境在对称性偏离程度上存在显著不一致, 可能导致键合区域的结构应力或异质性。

该特征的计算直接来源于第3.4.3节中定义的local_environment_anisotropy特征。其数学表达式为:

$$\Delta \text{Anisotropy}_{ij}^{\text{local}} = |\text{Anisotropy}_j^{\text{local}} - \text{Anisotropy}_i^{\text{local}}|$$

其中 $\text{Anisotropy}_k^{\text{local}}$ 是原子 k 的局部环境各向异性。

在物理上，该特征量化了键两端原子局部空间填充效率或几何对称性偏离程度的差异。高失配值可能指示键合区域的局部应变集中，尤其是在不同对称性区域的界面处。化学上，它有助于理解局部键合环境的异质性以及这种异质性如何影响键的性质和稳定性。材料学上，`bond_ends_anisotropy_mismatch` 是理解晶体结构局部畸变、相变机制、机械响应（如脆性）以及某些输运性质（如热传导、声子传输）的关键，尤其是在晶格畸变显著的材料中。在计算机与人工智能领域，此特征提供了对键合区域局部几何各向异性差异的连续量化，为机器学习模型提供了关于材料内部微观应力分布和局部结构异质性的信息，有助于预测材料的机械强度、热力学稳定性以及在复杂应力下的行为。

第五章：2-单纯形描述符：三体相互作用的构型与复杂性量化

2-单纯形，即三角形，在材料科学中代表了三个原子之间的三体相互作用，是理解局部几何构型、应力分布以及电荷传输路径的关键单元。本节将详细阐述如何通过几何学、跨层级特征融合、量子化学和高级代数理论，构建一套全面量化晶体结构中2-单纯形特征的描述符。这些特征不仅捕捉了三体相互作用的局部特性，还通过多层次信息的融合，提供了更深层次的物理化学洞察。

5.1 2-单纯形识别与几何定义

本框架通过识别晶体结构中所有独特的原子三元组来构建2-单纯形。一个2-单纯形由一个中心原子及其两个近邻原子组成，并通过Pymatgen的CrystalNN工具进行精确识别，并包含容差调整的鲁棒性处理。

5.2 几何特征

几何特征直接量化了2-单纯形本身的形状和空间排布。这些特征是晶体局部结构刚性与柔性的直接体现。

5.2.1 三角形面积 (`triangle_area`)

`triangle_area` 描述符量化了2-单纯形（即由三个原子核构成的三角形）在空间中的实际面积。它通过三个原子笛卡尔坐标所形成的两个向量的叉积模长的一半来精确计算。

$$\text{Area} = \frac{1}{2} \|(\mathbf{p}_j - \mathbf{p}_i) \times (\mathbf{p}_k - \mathbf{p}_i)\|_2$$

其中 $\mathbf{p}_i, \mathbf{p}_j, \mathbf{p}_k$ 分别是构成三角形的三个原子 i, j, k 的笛卡尔坐标向量。

在物理上，三角形面积是衡量局部原子排布紧密程度和空间占据的直接指标。较大的面积可能表示原子间距相对较大，或者局部结构较为开放。化学上，它反映了三原子体系的几何构型，例如在确定键角和配位多面体形状时具有重要意义。材料学上，三角形面积的变化可以指示局部应变、缺陷区域或相变过程中的结构重构。在计算机与人工智能领域，triangle_area 提供了一个基本且不变的几何特征，用于机器学习模型理解和预测材料的局域结构特性，例如在确定声子散射机制或热膨胀系数时具有潜在应用。

5.2.2 键角方差 (bond_angle_variance)

bond_angle_variance 描述符量化了组成2-单纯形的三个键角（即 $\angle ijk, \angle jki, \angle kij$ ）的离散程度。它通过计算这三个角（以弧度表示）的统计方差得到。

$$\text{Variance}(\theta_1, \theta_2, \theta_3) = \frac{1}{3} \sum_{n=1}^3 (\theta_n - \bar{\theta})^2$$

其中 θ_n 是三角形的三个内角， $\bar{\theta}$ 是它们的平均值。

在物理上，键角方差是衡量局部几何畸变和键合灵活性的关键指标。例如，在一个理想的正三角形中，键角方差为零；而在一个高度扭曲或扁平的三角形中，方差会显著增加。化学上，它直接关联到分子的构象应变和键合的灵活性，例如在多面体扭曲研究中具有重要意义。材料学上，高键角方差可能预示着局部结构不稳定、缺陷或相变前驱态，这可能影响材料的力学性能（如硬度、延展性）和热力学稳定性。在计算机与人工智能领域，bond_angle_variance 为机器学习模型提供了对材料内部局部结构畸变程度的量化描述，有助于模型识别和预测晶格缺陷、相变行为以及材料在极端条件下的结构稳定性。

5.2.3 三角形形状因子 (triangle_shape_factor)

triangle_shape_factor 描述符量化了2-单纯形（三角形）的形状与理想正三角形的偏离程度。它是一个无量纲的归一化因子，其值在0到1之间，其中1表示完美的正三角形，0表示极度退化的三角形（例如共线）。其定义如下：

$$\text{ShapeFactor} = \frac{12\sqrt{3} \cdot \text{Area}}{\text{Perimeter}^2}$$

其中 Area 是三角形的面积，Perimeter 是三角形的周长。

在物理上，三角形形状因子反映了局部晶格的各项异性和几何理想度。接近1的值表示局部环境近似理想对称，而接近0的值则表示严重的几何畸变。化学上，它提供了对三原子配位多面体或键合网络的几何规整性的量化评估，这对于理解局部化学环境对电子结构的影响至关重要。材料学上，`triangle_shape_factor` 是评估材料结构稳定性、相变行为以及特定晶体结构中应力分布的关键特征，例如在分析钙钛矿中的八面体扭曲时，它能有效量化单个面片的畸变程度。在计算机与人工智能领域，它为机器学习模型提供了对材料内部局部几何规整度的量化描述，有助于模型预测材料的对称性破缺、晶格弛豫效应以及对外部应力的响应。

5.3 跨层级派生特征

这些特征通过对0-单纯形（原子）和1-单纯形（键）特征的统计聚合，将更低层次的物理化学信息融入到2-单纯形描述中，从而捕捉三体相互作用中原子和键的平均行为和波动性。

5.3.1 原子不兼容性统计 (`avg_atomic_incompatibility`, `variance_atomic_incompatibility`)

这两个特征分别量化了组成2-单纯形的三个原子其结构-化学不兼容性（来自第3.4.1节）的平均值和方差。它们反映了三角形内部原子层面的局部应变或几何畸变的平均水平及其波动性。

$$\text{avg_atomic_incomp} = \frac{1}{3} \sum_{i \in \text{triangle}} \text{Incomp}_{\text{S-C},i}$$

$$\text{variance_atomic_incomp} = \frac{1}{3} \sum_{i \in \text{triangle}} (\text{Incomp}_{\text{S-C},i} - \text{avg_atomic_incomp})^2$$

其中 $\text{Incomp}_{\text{S-C},i}$ 是原子 i 的结构-化学不兼容性。

在物理上，高平均不兼容性可能指示三角形区域整体处于高应力状态或存在显著的晶格畸变。高方差则表明该区域内原子间的应力分布不均匀，存在局部应力集中或软点。化学上，这些特征揭示了三原子体系中不同原子在结构-化学环境上的差异。材料学上，它们是评估材料韧性、断裂韧性以及相变稳定性的关键指标，对于理解局部应力累积和失效机制至关重要。在计算机与人工智能领域，这些特征为机器学习模型提供了对材料内部局部应力分布和结构异质性的量化描述。

5.3.2 Bader电荷统计 (`avg_bader_charge`, `variance_bader_charge`)

这两个特征分别量化了组成2-单纯形的三个原子其Bader电荷（来自第3.3.1节）的平均值和方差。它们反映了三角形区域内电荷分布的平均状态和不均匀性。

$$\text{avg_bader_charge} = \frac{1}{3} \sum_{i \in \text{triangle}} \text{Charge}_i$$

$$\text{variance_bader_charge} = \frac{1}{3} \sum_{i \in \text{triangle}} (\text{Charge}_i - \text{avg_bader_charge})^2$$

其中 Charge_i 是原子 i 的Bader电荷。

在物理上，平均Bader电荷反映了该区域的整体电子富集或贫瘠程度，与材料的电荷输运和电势分布相关。方差则衡量了电荷在三个原子间的离域或局域程度，高方差可能指示电荷极化或电荷转移效应。化学上，它们揭示了局部化学键合的离子性或共价性特征。材料学上，这些特征是设计和优化电化学材料（如电池电极、催化剂）和半导体材料的关键指标，因为它们直接关联到材料的电荷输运效率和活性位点。在计算机与人工智能领域，这些特征有助于机器学习模型预测材料的电导率、催化活性以及光电转换效率。

5.3.3 矢量不对称性统计 (avg_vectorial_asymmetry, max_vectorial_asymmetry)

这两个特征分别量化了组成2-单纯形的三个原子其矢量不对称性范数平方（来自第3.4.1节）的平均值和最大值。它们捕捉了三角形内部原子局部环境对称性破缺的平均强度和最强破缺点。

$$\text{avg_vectorial_asymmetry} = \frac{1}{3} \sum_{i \in \text{triangle}} \|\mathbf{V}_{\text{asym},i}\|^2$$

$$\text{max_vectorial_asymmetry} = \max_{i \in \text{triangle}} (\|\mathbf{V}_{\text{asym},i}\|^2)$$

其中 $\|\mathbf{V}_{\text{asym},i}\|^2$ 是原子 i 的矢量不对称性范数平方。

在物理上，高平均矢量不对称性表明该区域整体处于非对称的局部势场中，可能导致压电、铁电或非线性光学效应。最大值则指示了该三角形内最不对称的原子位点。化学上，这些特征关联到局部配位环境的扭曲和立体化学效应。材料学上，它们是筛选具有特定对称性相关材料（如压电、热释电材料）的关键指标，对于理解相变过程中对称性破缺的微观机制至关重要。在计算机与人工智能领域，这些特征有助于机器学习模型预测材料的极化行为、非线性光学响应以及对外部场的敏感性。

5.3.4 键对齐度统计 (avg_bond_alignment, variance_bond_alignment)

这两个特征分别量化了组成2-单纯形的三条键其张量代数环境对齐度（来自第4.4.4节）的平均值和方差。它们反映了三角形内部键与局部原子环境对齐的平均程度及其波动性。

$$\text{avg_bond_alignment} = \frac{1}{3} \sum_{b \in \text{triangle bonds}} \text{Alignment}_b$$
$$\text{variance_bond_alignment} = \frac{1}{3} \sum_{b \in \text{triangle bonds}} (\text{Alignment}_b - \text{avg_bond_alignment})^2$$

其中 Alignment_b 是键 b 的张量代数环境对齐度。

在物理上，高键对齐度表示键的取向与局部原子排列的内在对称性高度一致，预示着更稳定的结构。高方差则可能指示键合网络中存在取向不均匀性，这可能导致各向异性或局部应力集中。化学上，它们与化学键的刚性、取向选择性以及晶体场效应相关。材料学上，这些特征是预测材料力学各向异性、光电响应以及特定晶向生长偏好的关键指标，对于理解晶体生长和宏观性能的各向异性至关重要。在计算机与人工智能领域，这些特征有助于机器学习模型预测材料的机械强度、韧性以及在外部应力下的变形行为。

5.3.5 键密度梯度统计 (avg_bond_gradient, max_bond_gradient)

这两个特征分别量化了组成2-单纯形的三条键其键密度梯度（来自第4.3.4节）的平均值和最大值。它们反映了三角形区域内电子密度变化的平均剧烈程度和最大变化点。

$$\text{avg_bond_gradient} = \frac{1}{3} \sum_{b \in \text{triangle bonds}} |\nabla \rho|_b$$
$$\text{max_bond_gradient} = \max_{b \in \text{triangle bonds}} (|\nabla \rho|_b)$$

其中 $|\nabla \rho|_b$ 是键 b 的键密度梯度大小。

在物理上，高平均键密度梯度表示该区域内电子密度变化剧烈，可能存在强的共价键合或电荷转移通道。最大值则指示了该三角形内电荷密度变化最陡峭的键。化学上，这些特征关联到化学键的共价性强度、键断裂能以及化学反应的活性位点。材料学上，它们是设计和优化催化剂、电池电极材料以及光电转换材料的关键指标，因为它们直接关联到材料的电子传输效率和化学反应活性。在计算机与人工智能领域，这些特征有助于机器学习模型预测材料的催化活性、电子迁移率以及在电化学反应中的性能。

5.3.6 键长统计 (avg_bond_distance, variance_bond_distance)

这两个特征分别量化了组成2-单纯形的三条键其键长（来自第4.2.1节）的平均值和方差。它们反映了三角形区域内键长的平均水平和不均匀性。

$$\text{avg_bond_distance} = \frac{1}{3} \sum_{b \in \text{triangle bonds}} L_b$$
$$\text{variance_bond_distance} = \frac{1}{3} \sum_{b \in \text{triangle bonds}} (L_b - \text{avg_bond_distance})^2$$

其中 L_b 是键 b 的键长。

在物理上，平均键长反映了该区域键合的整体强度和原子间距。方差则衡量了键长的畸变程度，高方差可能指示键合网络中的不均匀应变或缺陷。化学上，它们直接关联到化学键的稳定性、键能以及局部几何构型。材料学上，这些特征是预测材料力学性能（如模量、硬度）、热膨胀系数以及相变行为的关键指标，对于理解晶体结构稳定性和力学响应至关重要。在计算机与人工智能领域，这些特征有助于机器学习模型预测材料的力学强度、热稳定性以及在不同温度和压力下的结构变化。

5.4 量子化学特征

这些特征直接从第一性原理计算的电子密度和电子局域化函数（ELF）场中提取，用于量化2-单纯形几何重心处的电子结构特性，从而揭示局部电子环境的量子本质。

5.4.1 几何重心电子密度 (geometric_centroid_density)

geometric_centroid_density 描述符量化了2-单纯形（三角形）的几何重心处的电子密度值。该值通过对整个晶胞的电子密度三维网格进行高精度三线性插值得到。

$$\rho_{\text{centroid}} = \text{Interpolate}(\rho_{\text{grid}}, \mathbf{r}_{\text{centroid}})$$

其中 ρ_{grid} 是DFT计算得到的电子密度网格数据， $\mathbf{r}_{\text{centroid}}$ 是三角形的几何重心笛卡尔坐标。

在物理上，几何重心电子密度直接反映了该空间点处电子的富集程度，这与原子间键合、电荷转移以及局部化学环境密切相关。高电子密度通常指示共价键的存在或电子的局部离域。化学上，它为理解键合性质和反应活性提供了微观信息。材料学上，geometric_centroid_density 是预测材料导电性、光电转换效率以及催化活性位点的重要指标，因为电子密度是决定这些宏观性能的根本因素。在计算机与人工智能领域，它为机器学习模型提供了对材料内部局部电子环境和键合性质的量化描述，有助于模型预测材料的电学、光学和催化性能。

5.4.2 几何重心电子局域化函数 (geometric_centroid_elf)

geometric_centroid_elf 描述符量化了2-单纯形（三角形）的几何重心处的电子局域化函数（ELF）值。该值通过对整个晶胞的ELF三维网格进行高精度三线性插值得到。

$$\text{ELF}_{\text{centroid}} = \text{Interpolate}(\text{ELF}_{\text{grid}}, \mathbf{r}_{\text{centroid}})$$

其中 ELF_{grid} 是由GPAW计算器得到的ELF网格数据， $\mathbf{r}_{\text{centroid}}$ 是三角形的几何重心笛卡尔坐标。

在物理上，ELF是一个介于0和1之间的无量纲函数，用于识别电子的局域化区域（高值，如共价键和孤对电子）和离域区域（低值，如金属键和范德华区）。几何重心处的ELF值能够精确揭示该空间点是否存在电子对的局域化，从而判断其键合特性。化学上，高ELF值表示强烈的共价键特征或孤对电子的存在，而低值则可能指示离子键或弱相互作用。材料学上，geometric_centroid_elf 是分析材料的化学键合类型、原子间相互作用强度以及判断材料稳定性、力学行为的关键特征。在计算机与人工智能领域，它为机器学习模型提供了对材料内部局部键合类型和电子局域化特性的量化描述，有助于模型预测材料的力学强度、热稳定性以及化学反应活性。

5.4.3 几何重心电子密度拉普拉斯 (geometric_centroid_laplacian_of_density)

geometric_centroid_laplacian_of_density 描述符量化了2-单纯形（三角形）的几何重心处的电子密度拉普拉斯（ $\nabla^2 \rho$ ）值。该值通过对整个晶胞的 $\nabla^2 \rho$ 三维网格进行高精度三线性插值得到。

$$\nabla^2 \rho_{\text{centroid}} = \text{Interpolate}(\nabla^2 \rho_{\text{grid}}, \mathbf{r}_{\text{centroid}})$$

其中 $\nabla^2 \rho_{\text{grid}}$ 是由电子密度网格数据计算得到的拉普拉斯场， $\mathbf{r}_{\text{centroid}}$ 是三角形的几何重心笛卡尔坐标。

在物理上，电子密度拉普拉斯的符号提供了关于电子局域化和形变的关键信息： $\nabla^2 \rho < 0$ 表示电子在此处聚集（通常是共价键的键中心和孤对电子区域），而 $\nabla^2 \rho > 0$ 表示电子在此处耗散（通常是原子核区域或离子键的原子间区域）。几何重心处的拉普拉斯值能够精确识别该点是电子的聚集盆地还是耗散区域。化学上，负值指示键合相互作用的存在，正值则表示非键合区域。材料学上，geometric_centroid_laplacian_of_density 是理解材料键合性质、电荷传输路径以及局部应力分布的重要特征，尤其在分析化学键的性质和强度时具有独特优势。在计算机与人工智能领域，它为机器学习模型提供了对材料内部局部电子形变和键合拓扑的量化描述，有助于模型预测材料的电学性能、力学响应以及化学反应活性。

5.5 代数融合特征

这些特征将结构张量、密度梯度场、点群对称性以及层次化应力概念通过代数方法融合，以捕捉2-单纯形层面的高阶相互作用、通量、对称性约简以及复杂的应力流动模式。

5.5.1 结构张量积的迹 (structural_tensor_product_trace)

structural_tensor_product_trace 描述符量化了组成2-单纯形的三个原子（来自第3.4.2节的结构张量）所对应的局部结构张量 $\mathbf{T}_i, \mathbf{T}_j, \mathbf{T}_k$ 的高阶相互作用。它被定义为这三个张量乘积的迹：

$$\text{Tr}(\mathbf{T}_i \cdot \mathbf{T}_j \cdot \mathbf{T}_k)$$

其中 $\mathbf{T}_i, \mathbf{T}_j, \mathbf{T}_k$ 分别是原子 i, j, k 的3x3结构张量。

在物理上，此特征量化了三个相邻原子局部结构环境的复杂耦合效应。非零值表示局部几何和力场的非线性相互作用，这可能是由结构扭曲、键应变或集体原子协同运动引起的。绝对值越大，这种高阶相互作用越显著。化学上，它能捕捉到三个相互作用中心之间的协同键合或排斥效应。材料学上，structural_tensor_product_trace 是评估材料力学响应、声子散射以及相变过程中多体效应的关键指标，对于理解复杂晶体结构中的耦合现象至关重要。在计算机与人工智能领域，它为机器学习模型提供了对材料内部高阶结构耦合和多体相互作用的量化描述，有助于模型预测材料的非线性力学性能、热力学稳定性以及在外部场作用下的复杂行为。

5.5.2 总密度梯度通量 (total_density_gradient_flux)

total_density_gradient_flux 描述符量化了电子密度梯度矢量场穿过2-单纯形（三角形）平面的净通量。它通过计算三角形几何重心处的平均电子密度梯度向量与三角形平面单位法向量的点积，再乘以三角形面积来近似。

$$\Phi = \left(\frac{\nabla \rho_i + \nabla \rho_j + \nabla \rho_k}{3} \right) \cdot \mathbf{n} \cdot \text{Area}$$

其中 $\nabla \rho_i, \nabla \rho_j, \nabla \rho_k$ 分别是原子 i, j, k 处的电子密度梯度向量（通过插值获得）， \mathbf{n} 是三角形平面的单位法向量，Area 是三角形面积。该特征依赖于从平面波或FD模式GPAW计算器获取的电子密度梯度场。

在物理上，总密度梯度通量反映了电子密度在通过该三角形平面时的净流动趋势。正值表示电子倾向于流出该平面，负值表示流入。这与电荷传输通道、局域电场强度以及键合极性密切相关。化学上，它揭示了三原子体系中电子转移的方向性和强度。材料学上，total_density_gradient_flux 是设计和优化电化学材料（如电池电极、电解质）、半导体以及催化剂的关键特征，因为它直接关联到材料的电荷输运效

率和活性位点。在计算机与人工智能领域，它为机器学习模型提供了对材料内部电荷输运微环境和局部电荷极化的量化描述，有助于模型预测材料的电导率、离子电导率以及催化活性。

5.5.3 点群约简因子 (point_group_reduction_factor)

point_group_reduction_factor 描述符量化了组成2-单纯形的三个原子其局部点群对称性阶数（来自第3.4.3节）相对于其平均值的约简程度。它被定义为三角形中三个原子位点对称性阶数的最小值与它们平均值的比值：

$$\text{Reduction Factor} = \frac{\min(\text{Order}_i, \text{Order}_j, \text{Order}_k)}{\text{avg}(\text{Order}_i, \text{Order}_j, \text{Order}_k)}$$

其中 Order_i 是原子 i 的位点点群阶数。

在物理上，此特征反映了局部对称性破缺的严重程度。较低的值表示三角形中存在至少一个原子其对称性远低于平均水平，这可能预示着局域结构畸变或缺陷。较高的值则表示三个原子的局部对称性较为一致。化学上，它揭示了局部配位环境的规整性，对于理解化学键合的取向性和晶体场效应至关重要。材料学上，point_group_reduction_factor 是评估材料结构稳定性、相变行为以及预测材料宏观各向异性（如光学、电学各向异性）的关键指标。在计算机与人工智能领域，它为机器学习模型提供了对材料内部局部对称性异质性和对称性破缺程度的量化描述，有助于模型识别和预测晶格畸变、相变前驱态以及特定功能性材料。

5.5.4 结构张量法向投影 (structure_tensor_normal_projection)

structure_tensor_normal_projection 描述符量化了组成2-单纯形的三个原子的总结构张量（来自第3.4.2节）在三角形平面法向上的投影。它被定义为三角形平面的单位法向量 \mathbf{n} 与总结构张量 $\mathbf{T}_{\text{total}} = \mathbf{T}_i + \mathbf{T}_j + \mathbf{T}_k$ 的双点积。其严格表达式及其与矩阵乘法的等价性推导如下：

设单位法向量 \mathbf{n} 为列向量，其分量为 $\mathbf{n} = \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix}$ 。

设总结构张量 $\mathbf{T}_{\text{total}}$ 为 3×3 矩阵，其分量为 $\mathbf{T}_{\text{total}} = \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix}$ 。

1. 双点积 $\mathbf{n} \cdot \mathbf{T}_{\text{total}} \cdot \mathbf{n}$ 的分量形式展开

首先，计算向量 \mathbf{n} 与张量 $\mathbf{T}_{\text{total}}$ 的点积 $\mathbf{v} = \mathbf{n} \cdot \mathbf{T}_{\text{total}}$ 。这个操作将产生一个新的向量 \mathbf{v} ，其第 j 个分量为：

$$v_j = \sum_i n_i T_{ij}$$

然后，将得到的向量 \mathbf{v} 与向量 \mathbf{n} 进行点积：

$$\mathbf{n} \cdot \mathbf{T}_{\text{total}} \cdot \mathbf{n} = \mathbf{v} \cdot \mathbf{n} = \sum_j v_j n_j$$

将 v_j 的表达式代入，得到双点积的最终分量形式：

$$\mathbf{n} \cdot \mathbf{T}_{\text{total}} \cdot \mathbf{n} = \sum_j \left(\sum_i n_i T_{ij} \right) n_j = \sum_i \sum_j n_i T_{ij} n_j$$

2. 矩阵乘积 $\mathbf{n}^T \mathbf{T}_{\text{total}} \mathbf{n}$ 的分量形式展开

向量 \mathbf{n} 的转置为行向量 $\mathbf{n}^T = (n_x \quad n_y \quad n_z)$ 。

首先，计算矩阵 $\mathbf{T}_{\text{total}}$ 与列向量 \mathbf{n} 的乘积 $\mathbf{u} = \mathbf{T}_{\text{total}} \mathbf{n}$ 。这个操作将产生一个新的列向量 \mathbf{u} ，其第 i 个分量为：

$$u_i = \sum_j T_{ij} n_j$$

然后，计算行向量 \mathbf{n}^T 与列向量 \mathbf{u} 的乘积：

$$\mathbf{n}^T \mathbf{T}_{\text{total}} \mathbf{n} = \mathbf{n}^T \mathbf{u} = \sum_i n_i u_i$$

将 u_i 的表达式代入，得到矩阵乘积的最终分量形式：

$$\mathbf{n}^T \mathbf{T}_{\text{total}} \mathbf{n} = \sum_i n_i \left(\sum_j T_{ij} n_j \right) = \sum_i \sum_j n_i T_{ij} n_j$$

结论：

通过比较分量形式的展开结果 () 和 ()，我们可以看到，两者是完全相同的。因此，在张量表示和矩阵表示下，以下等式成立：

$$\mathbf{n} \cdot \mathbf{T}_{\text{total}} \cdot \mathbf{n} = \mathbf{n}^T \mathbf{T}_{\text{total}} \mathbf{n}$$

这个值是一个标量，量化了总结构张量在法向量 \mathbf{n} 方向上的分量强度，反映了该三角形平面在整体结构中的各向异性或受力情况。这一标量值不仅在数学上具有明确的定义，其物理和化学意义也为我们理解材料的微观结构与宏观性能之间的联系提供了深刻的洞察。从多学科视角审视，该特征的含义可进一步阐述如下：

在物理上，此特征反映了局部结构应力或畸变场在垂直于三角形平面的方向上的分量。较大的绝对值表示三角形平面受到垂直方向的压缩或拉伸应力，或存在垂直于该平面的结构不对称性。化学上，它与键合力在三维空间中的分布以及局部电子云的变形相关。材料学上，`structure_tensor_normal_projection` 是评估材料各向异性力学响应、压电效应以及特定晶面处结构稳定性的关键指标，对于理解晶格平面内的应力传递和晶面重构至关重要。在计算机与人工智能领域，它为机器学习模型提供了对材料内部局部应力方向性和晶面响应的量化描述，有助于模型预测材料的宏观力学性能、压电性能以及在外部应力下的结构稳定性。

5.5.5 层次应力流 (hierarchical_stress_flow)

`hierarchical_stress_flow` 描述符量化了以键不兼容性（来自第4.4.1节的`lie_algebra_incompatibility`）为“大小”的“应力”矢量，沿着三角形每条边流出三角形的净通量。它通过将每条键的李代数不兼容性与其在三角形平面内垂直于该键且指向外部的单位法向量相乘，然后对所有三条边进行矢量求和，最后取其模长得到。

$$\mathbf{F}_{\text{stress}} = \sum_{b \in \text{triangle bonds}} \text{Incomp}_b \cdot \mathbf{n}_b^\perp$$
$$\text{StressFlow} = \|\mathbf{F}_{\text{stress}}\|_2$$

其中 Incomp_b 是键 b 的李代数不兼容性， \mathbf{n}_b^\perp 是垂直于键 b 且在三角形平面内指向外部的单位法向量。

在物理上，此特征捕捉了在2-单纯形内部由于键合不兼容性引起的“应力”如何在局部区域内传递和累积，形成一个伪矢量流。高值表示该三角形区域存在显著的、定向的内部应力累积或释放。化学上，它揭示了键合应力在三体相互作用中的协同或拮抗效应。材料学上，`hierarchical_stress_flow` 是理解材料塑性变形、疲劳裂纹萌生以及相变过程中原子集体运动的关键指标，对于分析材料在复杂应力条件下的失效机制至关重要。在计算机与人工智能领域，它为机器学习模型提供了对材料内部局部应力传递路径和能量耗散模式的量化描述，有助于模型预测材料的塑性、韧性以及在循环载荷下的疲劳寿命。

第六章：全局描述符：材料整体性质的宏观量化与表征

本章将聚焦于全局描述符的构建与分析，这些描述符旨在从整个晶体结构层面量化其宏观物理、化学及几何性质。不同于前几章关注原子、化学键和三体团簇的局部特征，全局描述符提供了对材料整体特性的概括性理解。它们通常作为局部特征的补充或高层次抽象，对于预测材料的宏观性能（如总能量、稳定性、晶格各向异性、堆积效率）以及在材料设计和筛选中的快速评估具有不可替代的重要性。

6.1 基础统计与几何特征：晶体宏观结构的初阶量化

本节包含的特征主要基于晶体结构的宏观几何参数和统计平均值，提供了对材料整体尺寸、形状、堆积效率和键合状态的基础性描述。这些特征直观且易于计算，是理解晶体材料全局特性的起点。

6.1.1 平均键长 (mean_bond_length)

mean_bond_length描述符严格定义为晶体结构中所有化学键键长的算术平均值。它提供了一个对材料整体原子间连接强度和空间尺度的基本量化。其计算结果直接来源于第四章1-单纯形描述符中预计算的bond_distance特征的全局平均。

在物理上，平均键长与材料的平均原子间距直接相关，影响晶体的热膨胀系数、熔点和机械刚度。化学上，它反映了材料的平均成键类型和强度，共价键通常较短，离子键和金属键则可能表现出不同的平均长度。材料学上，平均键长是材料结构稳定性和密度的重要指标，是快速评估晶体结构紧密程度的有效特征。在计算机与人工智能领域，作为最基础的结构描述符之一，它为机器学习模型提供了对材料尺度和键合特性的直接输入，常用于材料筛选和性能预测的初筛阶段。

6.1.2 晶胞体积每化学式单位 (volume_per_fu)

volume_per_fu描述符严格定义为晶体的晶胞体积除以其化学式单位的数量。这量化了在材料化学式单位（例如，CsPbI₃的“1”个单位）中每个单胞所占据的平均空间。

在物理上，该特征直接反映了材料的宏观密度和空间填充效率。较低的volume_per_fu通常意味着更高的原子堆积密度。化学上，它与原子的尺寸、离子键和共价键的相对贡献密切相关。材料学上，这个特征是评估材料稳定性和进行高压合成预测的关键参数，因为更小的单位体积通常意味着更高的能量密度或更紧密的原子堆积，对材料的热力学稳定性有重要影响。在计算机与人工智能领域，volume_per_fu是一个重要的尺寸特征，可以用于预测材料的密度、压缩性以及在不同压力条件下的结构行为。

6.1.3 晶格各向异性比 (lattice_anisotropy_ratio)

lattice_anisotropy_ratio描述符严格定义为晶体晶格三个基矢量长度（a, b, c）中的最大值与最小值之比。它量化了晶体在不同晶向上的几何不对称性，是判断材料是否具有层状、链状或高度各向异性结构的关键指标。

$$\text{Lattice Anisotropy Ratio} = \frac{\max(a, b, c)}{\min(a, b, c)}$$

在物理上，高各向异性比通常预示着材料在不同方向上具有不同的物理性质，例如力学强度、热导率、电导率或光学响应。化学上，它可能与特定方向上的弱键（如范德华力）或强共价网络有关，导致结构在某些方向上更易变形。材料学上，lattice_anisotropy_ratio是设计具有定向性能材料（如2D材料、热电材料、光伏材料）的关键特征，高各向异性往往是这些材料独特功能的来源。在计算机与人工智能领域，它为机器学习模型提供了对材料结构各向异性程度的直观量化，有助于预测材料的定向传输性质、力学稳定性以及在特定应用场景下的性能表现。

6.1.4 体相各向异性指数 (bulk_anisotropy_index)

bulk_anisotropy_index描述符旨在量化体相晶格中原子局部环境的平均“形状”各向异性，从而作为材料内在结构各向异性程度的代理。它通过计算晶体中所有不等价位点（由空间群分析识别）的局部结构张量（来自第3.4.2节中引入的结构张量概念）的特征值各向异性，然后取其平均值得到。具体而言，对于每个位点 i ，计算其结构张量 \mathbf{T}_i 的三个特征值 $\lambda_{i,1} \geq \lambda_{i,2} \geq \lambda_{i,3}$ ，然后计算局部各向异性指数 $\eta_i = (\lambda_{i,1} - \lambda_{i,3}) / (\lambda_{i,1} + \lambda_{i,2} + \lambda_{i,3})$ 。最终的bulk_anisotropy_index是所有不等价位点 η_i 的平均值。

$$\text{Bulk Anisotropy Index} = \frac{1}{N_{\text{unique}}} \sum_{i \in \text{unique sites}} \frac{\lambda_{i,1} - \lambda_{i,3}}{\sum_{j=1}^3 \lambda_{i,j}}$$

其中 N_{unique} 是不等价位点的数量， $\lambda_{i,j}$ 是第 i 个位点结构张量的特征值。

在物理上，该特征捕捉了晶体内部原子排列的局部对称性破缺和非球形特征，是微观结构应力、局域畸变以及晶格振动各向异性的直接体现。化学上，它与原子间成键方向性、配位环境畸变以及电子轨道杂化模式紧密相关。材料学上，bulk_anisotropy_index是预测材料弹性各向异性、声子传输、热导率以及甚至铁电、压电等功能特性的重要指标，因为这些特性往往源于微观尺度的结构各向异性。在计算机与人工智能领域，它为机器学习模型提供了对材料内部结构不均匀性和局部应变场的量化描述，有助于模型理解和预测材料在复杂环境下的响应，从而加速材料设计和性能优化。

6.1.5 八面体计数 (octahedral_count)

octahedral_count描述符严格计算晶体结构中符合特定化学和几何标准的八面体数量。为了确保物理意义的正确性，本特征仅对满足以下条件的原子进行计数：(1) 中心原子必须是金属（如铅Pb）；(2) 其配位数必须精确为6；(3) 所有6个近邻必须是非金属（如溴Br）。对于符合这些条件的中心原子，系统会计算其局部结构有序参数，当该有序参数大于预设物理阈值0.75时，才被认定为一个有效的八面体。

在物理上，八面体是钙钛矿等多种功能材料中的基本结构单元，其数量和连接方式直接决定了材料的晶体结构类型、稳定性以及许多宏观物理性质，例如带隙、激子绑定能和铁电性。化学上，八面体的存在和完整性反映了中心金属离子与配位非金属离子之间的键合特性和配位环境的稳定性。材料学上，octahedral_count是钙钛矿、尖晶石等含八面体结构材料的关键表征，例如在钙钛矿光伏材料中，八面体的扭曲和倾斜程度直接影响载流子传输效率和材料稳定性。在计算机与人工智能领域，它为机器学习模型提供了对材料基本结构单元数量的直接量化，有助于模型识别特定结构基元对材料性能的影响，是进行材料分类和性能预测的重要特征。

6.1.6 堆积因子 (packing_fraction)

packing_fraction描述符严格定义为晶胞中所有原子的离子体积之和与晶胞总体积之比。它量化了晶体结构的原子空间填充效率，反映了原子在晶胞中紧密堆积的程度。其计算依赖于第4.2.6节中获取的预计算离子半径数据，并确保原子顺序与CIF文件匹配。具体计算公式为：

$$\text{Packing Fraction} = \frac{\sum_{i=1}^N \frac{4}{3} \pi r_i^3}{V_{\text{cell}}}$$

其中 N 是晶胞中的原子总数， r_i 是第 i 个原子的离子半径， V_{cell} 是晶胞的体积。

在物理上，高堆积因子通常意味着更稳定的结构和更高的密度，影响材料的硬度、熔点和声子传播速度。化学上，它与原子间的排斥作用和键合类型有关，紧密堆积通常发生在具有较强各向同性键合的材料中。材料学上，packing_fraction是评估材料热力学稳定性、预测相变行为以及设计新材料结构的重要指标，例如在玻璃和非晶材料中，堆积因子与玻璃形成能力和结构弛豫行为密切相关。在计算机与人工智能领域，它为机器学习模型提供了对材料原子排列紧密程度的量化描述，有助于模型预测材料的机械性能、热力学稳定性以及在不同压力和温度条件下的结构行为。

6.2 DFT计算的基态属性：材料能量与电子结构的宏观洞察

本节的全局描述符直接来源于密度泛函理论（DFT）计算所获得的晶体基态属性。这些特征量化了材料整体的能量稳定性、电子能级特征以及平均静电环境，它们是理解和预测材料热力学稳定性、电子传输性能以及化学反应活性的关键。

6.2.1 每原子总能量 (total_energy_per_atom)

total_energy_per_atom描述符严格定义为晶体体系的总能量除以晶胞中的原子数量。它代表了材料在绝对零度下，每个原子平均分摊的能量，是衡量材料热力学稳定性的最基本和最重要的指标之一。

$$E_{\text{atom}} = \frac{E_{\text{total}}}{N_{\text{atoms}}}$$

其中 E_{total} 是体系的总能量， N_{atoms} 是晶胞中的原子总数。

在物理上，更低的每原子总能量通常意味着材料结构的热力学稳定性更高，因为它处于更深的能量阱中。化学上，这个特征反映了原子间键合的整体强度和稳定性，较低的能量表明键合更强、更优化。材料学上，total_energy_per_atom是预测材料相稳定性、判断新材料合成可行性以及评估缺陷形成能的关键参数，在晶体结构搜索和相图构建中具有核心作用。在计算机与人工智能领域，它是材料性质预测模型中最常用的输出或输入特征之一，直接关联到材料的形成能、分解能和热力学相变行为，对机器学习模型理解材料的稳定性具有决定性作用。

6.2.2 费米能级 (fermi_level)

fermi_level描述符严格定义为晶体体系的费米能级。在绝对零度下，费米能级是电子填充的最高能量状态。对于导体，它位于导带中；对于半导体和绝缘体，它位于带隙中。费米能级的位置决定了材料的导电性类型和电子跃迁的难易程度。

在物理上，费米能级是电子结构和电子输运性质的核心概念。它的精确位置对材料的电学性能（如导电性、载流子浓度）、热电性能和光学吸收边缘具有决定性影响。化学上，费米能级可以被视为材料的化学势，反映了电子的得失难易程度，高费米能级意味着材料倾向于失去电子（还原性强），低费米能级意味着倾向于获得电子（氧化性强）。材料学上，fermi_level是设计半导体、光电器件、热电材料和催化剂的关键参数，例如，在光伏材料中，费米能级与能带边缘的相对位置直接影响太阳能转换效率。在计算机与人工智能领域，它为机器学习模型提供了对材料电子态和输运性能的量化描述，是预测材料导电性、光学吸收、催化活性以及设计新型电子器件的重要特征。

6.2.3 静电势均值 (electrostatic_potential_mean)

electrostatic_potential_mean描述符严格定义为晶体体系内所有空间格点上静电势的平均值。静电势是带电粒子在空间中经历的电场能量，其平均值反映了晶体内部整体的电荷分布均匀性和平均电势环境。

$$\Phi_{\text{mean}} = \frac{1}{N_{\text{grid}}} \sum_{j=1}^{N_{\text{grid}}} \Phi(\mathbf{r}_j)$$

其中 $\Phi(\mathbf{r}_j)$ 是格点 \mathbf{r}_j 处的静电势， N_{grid} 是有效格点总数。

在物理上，静电势的均值与体系的整体电中性以及长程静电相互作用的平均强度有关。它提供了一个宏观的电势基准，有助于理解晶体对外场的响应。化学上，虽然单一平均值可能无法揭示局部的电荷转移，但它提供了对整体电荷环境的初步评估。材料学上，electrostatic_potential_mean可以作为评估材料介电性能、表面功函数以及电荷注入/提取效率的辅助指标，例如在电容器和电池材料中，平均电势环境影响电荷存储和传输。在计算机与人工智能领域，它为机器学习模型提供了对材料整体静电环境的宏观描述，有助于预测材料的电学性能、表面行为以及在电场作用下的响应。

6.2.4 静电势方差 (electrostatic_potential_variance)

electrostatic_potential_variance描述符严格定义为晶体体系内所有空间格点上静电势的方差。它量化了晶体内部静电势分布的离散程度或不均匀性。高方差意味着静电势在空间中波动剧烈，可能存在显著的电荷不均匀分布或极化效应；低方差则表示电势分布相对均匀。

$$\sigma_{\Phi}^2 = \frac{1}{N_{\text{grid}}} \sum_{j=1}^{N_{\text{grid}}} (\Phi(\mathbf{r}_j) - \Phi_{\text{mean}})^2$$

其中 $\Phi(\mathbf{r}_j)$ 是格点 \mathbf{r}_j 处的静电势， Φ_{mean} 是静电势的平均值， N_{grid} 是有效格点总数。

在物理上，静电势的方差是衡量晶体内部电荷不均匀性和局部极化强度的重要指标。它直接关联到材料的介电常数、偶极矩以及潜在的铁电或压电性质。化学上，高静电势方差通常意味着存在较强的离子键合、明显的电荷转移或局部电荷集中区域。材料学上，electrostatic_potential_variance是设计高性能介电材料、铁电体和压电体以及新型催化剂的关键特征，因为它直接反映了材料内部电荷分布的复杂性和功能性。在计算机与人工智能领域，它为机器学习模型提供了对材料内部电荷涨落和局域场效应的量化描述，有助于模型预测材料的功能特性、电荷传输路径以及在外部场下的响应，是理解和优化功能材料的关键输入。

6.3 全局高阶代数特征：超越传统描述的抽象量化

本节的全局描述符是本框架的核心创新点之一，它们引入了源自几何代数、李代数、辛几何等高阶数学概念的抽象量化。这些特征旨在捕捉晶体材料更深层次的结构对称性、力场响应以及电子与结构之间的复杂耦合，为材料的宏观功能特性提供更精细、更具物理洞察力的描述。

6.3.1 商代数与结构几何特征

本小节的特征主要利用商代数的思想来量化晶体的结构对称性，并通过几何代数的观点来描述全局的结构不对称性。它们提供了一种抽象而精确的方式来衡量晶体的序与无序程度，以及其偏离理想对称性的程度。

6.3.1.1 结构哈希值 (structure_hash)

structure_hash描述符严格定义为晶体结构的唯一图哈希值。它将晶体结构抽象为图论中的图（原子为节点，化学键为边），并使用Weisfeiler-Lehman算法，生成一个紧凑的、高度区分性的字符串。这个哈希值能够唯一标识一个晶体结构，即使其空间群相同但原子排列细节不同。

在物理上，结构哈希值本身不直接具有物理量纲，但其唯一性保证了对材料“指纹”的精确识别。它能够快速区分不同晶体结构异构体，即使它们在宏观上可能看起来相似。化学上，它确保了对同分异构体和多晶型材料的精确识别，有助于材料数据库的建立和管理。材料学上，structure_hash是材料信息学中进行结构识别、重复数据删除以及构建材料数据库索引的关键工具，对于加速高通量计算和实验筛选具有重要意义。在计算机与人工智能领域，它为机器学习模型提供了对材料结构唯一性的离散化表示，尤其适用于基于结构相似性的材料分类、推荐系统和逆向设计中对特定结构目标的匹配。

6.3.1.2 对称轨道连通性 (symmetry_orbit_connectivity)

symmetry_orbit_connectivity描述符严格定义为晶体结构中不同对称等价原子轨道之间的化学键数量占总化学键数量的比例。它量化了晶体内部对称性破缺的程度以及不同对称独立区域之间的结构连接强度。

$$\text{Symmetry Orbit Connectivity} = \frac{N_{\text{inter-orbit bonds}}}{N_{\text{total bonds}}}$$

其中 $N_{\text{inter-orbit bonds}}$ 是连接不同对称等价原子的化学键数量， $N_{\text{total bonds}}$ 是晶体中的化学键总数。

在物理上，高symmetry_orbit_connectivity值通常意味着更复杂的晶体结构，可能伴随着更低的对称性，以及不同晶体学位点之间的强相互作用。这可能导致材料在宏观上表现出极性、手性或各向异性等功能特性。化学上，它反映了不同化学环境下的原子如何通过键合连接起来，这对于理解复杂化合物的形成和稳定性至关重要。材料学上，symmetry_orbit_connectivity是设计具有特定对称性相关功能材料的关键特征，例如在铁电体中，对称性破缺导致原子位移，进而产生电极化。在计算机与人工智能领域，它为机器学习模型提供了对材料内部对称性破缺程度及其对长程结构有序性影响的量化描述，有助于模型预测材料的对称性相关功能，并指导新型功能材料的发现。

6.3.1.3 全局不对称性范数 (global_asymmetry_norm)

global_asymmetry_norm描述符严格定义为晶体结构中所有原子局部不对称性向量范数的平均值。它量化了晶体整体的“平均局部不对称性”或“平均局部扭曲程度”，而非简单地将所有局部向量求和（这可能导致向量相互抵消而无法反映实际的局部畸变）。每个原子的局部不对称性向量 V_{struct} （来自第3.4.1节）表示其配位环境偏离理想对称性的程度。最终的global_asymmetry_norm是所有原子局部不对称性向量 $\|V_{\text{struct}}\|$ 的平均值。

$$\text{Global Asymmetry Norm} = \frac{1}{N_{\text{atoms}}} \sum_{i=1}^{N_{\text{atoms}}} \|V_{\text{struct},i}\|$$

其中 N_{atoms} 是晶胞中的原子总数， $\|V_{\text{struct},i}\|$ 是第 i 个原子局部不对称性向量的欧几里得范数。

在物理上，global_asymmetry_norm提供了对晶体整体结构畸变程度的宏观量化。高值表明晶体内部存在广泛的局部结构扭曲，这可能与材料的相变行为、低频声子模式以及某些功能特性（如铁电性、压电性）有关。化学上，它反映了晶体中平均配位环境的偏离理想状态的程度，有助于理解键合环境的异质性。材料学上，这个特征是评估材料结构稳定性和功能性的重要指标，尤其适用于分析具有复杂结构畸变（如八面体倾斜、原子位移）的钙钛矿等功能材料。在计算机与人工智能领域，它为机器学习模型提供了对材料整体结构畸变和应变场的连续、可微量化，是预测材料结构相变、力学响应以及对称性相关功能（如极性）的关键特征。

6.3.1.4 李代数不对称性范数熵 (lie_asymmetry_magnitude_entropy)

lie_asymmetry_magnitude_entropy描述符严格定义为晶体中所有原子局部不对称性向量范数（ $\|V_{\text{struct}}\|$ ）分布的香农熵。它量化了晶体中原子局部结构不对称性程度分布的多样性或复杂性。高熵值表示原子局部不对称性程度分布广泛，结构可能更加无序或复杂，存在多种不同程度的局部畸变；低熵值则表示原子局部不对称性程度相似，结构更趋于规整。该熵通过对局部不对称性范数进行直方图统计，然后计算其概率分布的香农熵得到。

$$H = - \sum_{k=1}^M p_k \log_2(p_k)$$

其中 M 是直方图的箱数， p_k 是第 k 个箱中范数出现的概率。

在物理上，该熵提供了一种衡量晶体结构“信息内容”或“复杂度”的指标，反映了局部畸变模式的多样性。高熵可能与无序相、玻璃态或具有复杂构型的材料有关。化学上，它揭示了不同原子位点电荷或几何环境的异质性。材料学上，lie_asymmetry_magnitude_entropy有助于理解材料的结构相变机制、无序程度以及热力学稳定性，是评估材料在特定温度或压力下结构灵活性的重要特征。在计算机与人工智能领

域，它为机器学习模型提供了对材料内部结构变异性和复杂性的高级抽象量化，有助于模型识别和预测具有特定构型多样性（例如用于高熵合金）或结构无序性（例如用于非晶材料）的材料的功能特性，从而在复杂材料体系的设计中发挥关键作用。

6.3.2 基于力的局部扭转应力特征

本小节的特征将原子所受的力（Force）信息与局部结构不对称性向量相结合，旨在量化晶体内部由局部畸变引起的“扭转应力”。这些特征提供了对材料力学稳定性和动态行为的深层洞察，尤其适用于理解非弹性变形、声子模式和相变驱动力。

6.3.2.1 力协方差不变量1 (force_covariance_invariant_1)

force_covariance_invariant_1描述符严格定义为晶胞中所有原子所受力的协方差矩阵的迹。它量化了晶体中原子受力分布的整体离散程度，反映了材料内部的平均应力涨落或力场不均匀性。该特征的计算需要首先从GPAW计算器获取所有原子所受的力矢量 \mathbf{f}_i ，然后构建力的协方差矩阵 $\mathbf{C}_f = \frac{1}{N_{\text{atoms}}} \sum_i \mathbf{f}_i \mathbf{f}_i^T$ ，最后计算该矩阵的迹 $\text{Tr}(\mathbf{C}_f) = \sum_k \sigma_{kk}$ 。

$$\text{Force Covariance Invariant}_1 = \text{Tr} \left(\frac{1}{N_{\text{atoms}}} \sum_{i=1}^{N_{\text{atoms}}} \mathbf{f}_i \mathbf{f}_i^T \right)$$

其中 \mathbf{f}_i 是第 i 个原子所受的力矢量， N_{atoms} 是原子总数。

在物理上，force_covariance_invariant_1与材料的整体应变能密度、局部晶格软模或声子不稳定性相关。高值可能预示着材料处于高应力状态，或存在显著的局域晶格畸变。化学上，它反映了原子间相互作用力场的复杂性和不均匀性，有助于理解键合环境的异质性。材料学上，这个特征是预测材料力学性能（如韧性、脆性）、疲劳行为以及结构相变倾向的重要指标，例如在具有负热膨胀系数的材料中，其内部力场可能表现出异常的波动。在计算机与人工智能领域，它为机器学习模型提供了对材料内部力场分布和应力涨落的量化描述，有助于模型预测材料在外部载荷下的响应、缺陷形成能以及热力学稳定性，是理解材料力学行为和结构动态的关键特征。

6.3.2.2 力协方差不变量2 (force_covariance_invariant_2)

force_covariance_invariant_2描述符严格定义为晶胞中所有原子所受力的协方差矩阵的行列式。作为协方差矩阵的第二个不变量，它进一步量化了力场分布的多维度各向异性和“体积”大小。与迹（总方差）不同，行列式能捕捉到力的波动在不同正交方向上的相关性，以及整个力场分布的“散度”或“形状”。

$$\text{Force Covariance Invariant}_2 = \text{Det} \left(\frac{1}{N_{\text{atoms}}} \sum_{i=1}^{N_{\text{atoms}}} \mathbf{f}_i \mathbf{f}_i^T \right)$$

其中 \mathbf{f}_i 是第 i 个原子所受的力矢量， N_{atoms} 是原子总数。

在物理上，高force_covariance_invariant_2值可能意味着力场在空间中存在强烈的各向异性分布，这可能与材料的非弹性响应、特定的声子模式或结构相变的各向异性驱动力有关。化学上，它能更精细地揭示原子间相互作用的复杂各向异性，例如在具有强方向性键（如共价键）的材料中可能表现出高值。材料学上，这个特征是预测材料各向异性力学性能、各向异性热输运以及理解结构相变中能量景观各向异性的重要指标。在计算机与人工智能领域，它为机器学习模型提供了对材料内部力场张量性质的更高阶量化，有助于模型理解材料的结构柔性、变形机制以及特定方向上的力学响应，是材料力学设计和功能预测中的高级特征。

6.3.2.3 总扭转应力 (total_torsional_stress)

total_torsional_stress描述符严格定义为晶体中所有原子所受局部扭转应力大小的总和。单个原子的局部扭转应力被定义为其局部结构不对称性向量 $\mathbf{V}_{\text{struct},i}$ （来自第3.4.1节矢量不对称性平方范数）与该原子所受的力矢量 \mathbf{f}_i 之间的叉积范数 $\|\mathbf{V}_{\text{struct},i} \times \mathbf{f}_i\|$ 。这个叉积量化了由局部结构畸变引起的“力矩”，即力对结构不对称性的“扭转”效应。最终的total_torsional_stress是所有原子扭转应力大小的求和。

$$\text{Total Torsional Stress} = \sum_{i=1}^{N_{\text{atoms}}} \|\mathbf{V}_{\text{struct},i} \times \mathbf{f}_i\|$$

其中 $\mathbf{V}_{\text{struct},i}$ 是第 i 个原子的局部不对称性向量， \mathbf{f}_i 是第 i 个原子所受的力矢量。

在物理上，total_torsional_stress直接反映了晶体内部局部畸变与作用力场之间的耦合强度。高值表明晶体中存在显著的局部结构扭曲以及与之相伴随的内部力矩，这可能导致材料的结构不稳定、易于发生结构相变或产生非线性响应。化学上，它有助于理解特定配位环境中的键合张力以及不同原子之间应力传递的模式。材料学上，total_torsional_stress是预测材料力学性能（如塑性、蠕变）、结构相变前兆以及评估材料对外部刺激响应的关键特征，例如在铁弹材料中，该特征可以量化结构畸变与弹性应力之间的耦合。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“几何-力学”耦合状态的连续、可微量化，是预测材料结构稳定性、机械响应以及功能转换（如相变）的重要特征，在材料设计和优化中具有独特的应用价值。

6.3.3 场-密度梯度耦合特征

本小节的特征着重于从电子密度场和静电势场中提取高阶信息，通过构建场-密度梯度耦合张量来量化晶体内部的电荷分布与电场之间的复杂相互作用。这些特征提供了对材料电学响应、介电性能以及电子结构稳定性的深层洞察。

6.3.3.1 场-密度耦合不变量1 (field_density_coupling_invariant_1)

field_density_coupling_invariant_1描述符严格定义为晶体内部电场 $\mathbf{E}(\mathbf{r})$ 与电子密度梯度 $\nabla\rho(\mathbf{r})$ 构建的场-密度梯度耦合张量 \mathbf{M} 的迹。该张量 $\mathbf{M}_{ij} = \langle E_i \nabla_j \rho \rangle$ 描述了静电场与电子密度梯度的耦合，其中 $\langle \cdot \rangle$ 表示空间平均。其迹 $\text{Tr}(\mathbf{M})$ 量化了整体电场对电子密度空间变化的总效应，反映了材料内部的平均“电子应力”或电荷分布的响应性。计算过程首先从GPAW获取电子密度和静电势网格数据，然后计算电场（静电势的负梯度）和电子密度梯度，最后构建并求取张量迹。

其连续形式的严格数学表达式为：

$$\text{Field Density Coupling Invariant}_1 = \text{Tr} \left(\frac{1}{V_{\text{grid}}} \int_V \mathbf{E}(\mathbf{r}) \otimes \nabla\rho(\mathbf{r}) d^3\mathbf{r} \right)$$

其中 V_{grid} 是积分区域的总晶胞体积。

在实际DFT计算中，电子密度和电场等物理量是在离散的格点上定义的。因此，上述连续积分需要通过离散求和进行近似。设晶胞内共有 N_{grid} 个均匀分布的格点，每个格点占据的体积微元为 $d^3\mathbf{r} \approx \Delta V = V_{\text{grid}}/N_{\text{grid}}$ 。

则场-密度梯度耦合张量 \mathbf{M} 的积分近似为：

$$\begin{aligned} \mathbf{M} &\approx \frac{1}{V_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla\rho(\mathbf{r}_k) \Delta V \\ \mathbf{M} &\approx \frac{1}{V_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla\rho(\mathbf{r}_k) \frac{V_{\text{grid}}}{N_{\text{grid}}} \\ \mathbf{M} &\approx \frac{1}{N_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla\rho(\mathbf{r}_k) \end{aligned}$$

此处， $\mathbf{E}(\mathbf{r}_k) \otimes \nabla\rho(\mathbf{r}_k)$ 是在格点 \mathbf{r}_k 处的电场向量与电子密度梯度向量的外积，结果是一个二阶张量（矩阵）。

然后，我们求取这个近似张量的迹。利用张量外积的迹的性质 $\text{Tr}(\mathbf{A} \otimes \mathbf{B}) = \mathbf{A} \cdot \mathbf{B} = \sum_i A_i B_i$ (即两个向量的点积)，我们得到：

$$\begin{aligned}
\text{Field Density Coupling Invariant}_1 &= \text{Tr}(\mathbf{M}) \\
&\approx \text{Tr} \left(\frac{1}{N_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla \rho(\mathbf{r}_k) \right) \\
&= \frac{1}{N_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \text{Tr}(\mathbf{E}(\mathbf{r}_k) \otimes \nabla \rho(\mathbf{r}_k)) \\
&= \frac{1}{N_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \cdot \nabla \rho(\mathbf{r}_k) \\
&= \frac{1}{N_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \sum_i E_i(\mathbf{r}_k) \frac{\partial \rho(\mathbf{r}_k)}{\partial x_i}
\end{aligned}$$

其中 $\mathbf{E}(\mathbf{r}_k)$ 是格点 \mathbf{r}_k 处的电场， $\nabla \rho(\mathbf{r}_k)$ 是电子密度梯度， N_{grid} 是格点总数。

这与实际计算中对电场和电子密度梯度分量进行逐点乘积求和再平均的方法是完全一致的。

在明确了这一特征的严谨数学定义与计算方法之后，其深远的物理、化学和材料学意义将为我们理解材料的微观电荷分布与宏观响应机制提供关键洞察。以下将从多学科角度对其进行详细阐述：

在物理上，field_density_coupling_invariant_1与材料的介电响应、极化强度以及电子-声子耦合强度密切相关。高值可能预示着材料在电场作用下电子分布会发生显著变化，从而导致强的介电效应或非线性光学响应。化学上，它反映了电场如何影响局部键合和电荷转移过程。材料学上，这个特征是设计高性能介电材料、铁电体和非线性光学材料的关键指标，它提供了对材料微观电荷分布与宏观电场相互作用的定量描述。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“电场-电子响应”耦合的连续、可微量化，有助于模型预测材料在电场作用下的响应、电荷输运性质以及电极化行为，是理解和优化功能材料的关键高级特征。

6.3.3.2 场-密度耦合不变量2 (field_density_coupling_invariant_2)

field_density_coupling_invariant_2描述符严格定义为场-密度梯度耦合张量 \mathbf{M} 的行列式。作为耦合张量的第二个不变量，它进一步量化了电场对电子密度空间变化影响的多维度各向异性和“非正交性”。与迹（总耦合强度）不同，行列式能捕捉到电场与电子密度梯度在不同正交方向上的复杂关联，反映了电荷响应的各向异性程度以及空间分布的“体积”大小。

其连续形式的严格数学表达式为：

$$\text{Field Density Coupling Invariant}_2 = \text{Det} \left(\frac{1}{V_{\text{grid}}} \int_V \mathbf{E}(\mathbf{r}) \otimes \nabla \rho(\mathbf{r}) d^3\mathbf{r} \right)$$

其中 V_{grid} 是积分区域的总晶胞体积。

在实际DFT计算中，电子密度和电场等物理量是在离散的格点上定义的。因此，上述连续积分需要通过离散求和进行近似。设晶胞内共有 N_{grid} 个均匀分布的格点，每个格点占据的体积微元为 $d^3\mathbf{r} \approx \Delta V = V_{\text{grid}}/N_{\text{grid}}$ 。

则场-密度梯度耦合张量 \mathbf{M} 的积分近似为：

$$\begin{aligned}\mathbf{M} &= \frac{1}{V_{\text{grid}}} \int_V \mathbf{E}(\mathbf{r}) \otimes \nabla \rho(\mathbf{r}) d^3\mathbf{r} \\ &\approx \frac{1}{V_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla \rho(\mathbf{r}_k) \Delta V \\ &= \frac{1}{V_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla \rho(\mathbf{r}_k) \frac{V_{\text{grid}}}{N_{\text{grid}}} \\ &= \frac{1}{N_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla \rho(\mathbf{r}_k)\end{aligned}$$

因此，field_density_coupling_invariant_2 的近似离散形式为：

$$\text{Field Density Coupling Invariant}_2 \approx \text{Det} \left(\frac{1}{N_{\text{grid}}} \sum_{k=1}^{N_{\text{grid}}} \mathbf{E}(\mathbf{r}_k) \otimes \nabla \rho(\mathbf{r}_k) \right)$$

其中 $\mathbf{E}(\mathbf{r}_k)$ 是格点 \mathbf{r}_k 处的电场， $\nabla \rho(\mathbf{r}_k)$ 是电子密度梯度， N_{grid} 是格点总数。

这与实际计算中对电场和电子密度梯度分量进行逐点张量积求和再平均，最后计算行列式的方法是完全一致的。

在明确了这一特征的严谨数学定义与计算方法之后，其在物理、化学和材料学领域的深远意义将为我们理解材料的微观电荷分布与宏观响应机制提供关键洞察。以下将从多学科角度对其进行详细阐述：

在物理上，高field_density_coupling_invariant_2值可能意味着电场与电子密度梯度之间存在强烈的各向异性耦合，这可能导致材料表现出显著的各向异性介电行为、非线性光学效应或特定的电荷输运路径。化学上，它能更精细地揭示电场如何影响特定方向上的电荷极化和键合重构。材料学上，这个特征是设计具有方向性电学性能材料、光电器件和新型催化剂的关键指标，例如在低对称性晶体中，该特征可以量化其电学响应的各向异性。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“电场-电子响应”张量性质的更高阶量化，有助于模型理解材料的电学各向异性、载流子传输机制以及在特定电场方向上的功能转换，从而在高级功能材料的设计中发挥关键作用。

6.3.3.3 总梯度范数 (total_gradient_norm)

total_gradient_norm描述符严格定义为晶体内部所有空间格点上电子密度梯度范数的平方的平均值。它量化了晶体中电子密度在空间中变化的总剧烈程度或“崎岖性”。高值表明电子密度在空间中变化迅速，可能存在明显的化学键、电荷转移区域或原子核附近；低值则表示电子密度分布相对平缓。该特征通过GPAW计算器获取电子密度网格数据，计算其梯度，然后计算梯度范数的平方，最后对所有有效格点求平均。

$$\text{Total Gradient Norm} = \frac{1}{N_{\text{grid}}} \sum_{j=1}^{N_{\text{grid}}} \|\nabla \rho(\mathbf{r}_j)\|^2$$

其中 $\|\nabla \rho(\mathbf{r}_j)\|$ 是格点 \mathbf{r}_j 处电子密度梯度的欧几里得范数。

在物理上，total_gradient_norm与材料的电子局域化程度、化学键合强度以及电子动能密度密切相关。高值通常意味着存在较强的共价键或局域电子态。化学上，它直接反映了电荷密度的空间分布特征，特别是在键区域和原子核周围的电荷集中和耗散。材料学上，total_gradient_norm是评估材料化学键合类型、电子传输路径以及结构稳定性的重要指标，例如在半导体和绝缘体中，它与带隙大小和电子有效质量相关。在计算机与人工智能领域，它为机器学习模型提供了对材料内部电子密度空间变化模式的连续、可微量化，有助于模型预测材料的电子传输性能、介电性质以及化学反应活性，是理解材料电子结构与功能之间联系的关键特征。

6.3.4 基于静态相空间的全局伪辛几何特征

本小节的特征借鉴了辛几何 (Symplectic Geometry) 的深刻思想，构建了一个“伪辛相空间”来量化晶体中化学键的几何自由度与键中点处的电子密度梯度之间的耦合涨落。尽管其并非严格的辛流形构造，但这种类比提供了一个强大的框架来描述材料中结构与电子行为的深层耦合。该特征旨在捕捉材料中结构-电子相互作用的复杂性和内在的“体积”大小，反映了静态晶格中结构与电子响应的协同涨落。最终的特征值返回该耦合度量的对数，以避免数值下溢并更好地反映不同量级上的变化。

6.3.4.1 伪辛涨落体积的对数 (log_pseudo_symplectic_fluctuation_volume)

log_pseudo_symplectic_fluctuation_volume描述符严格定义为晶体中所有化学键所构成的“伪辛相空间”中状态矢量分布协方差矩阵的行列式的自然对数。每个化学键 k 在此伪辛相空间中被表示为一个六维状态矢量 $\mathbf{s}_k = (\mathbf{q}_k, \mathbf{p}_k)$ ，其中：

\mathbf{q}_k 是该化学键的键方向向量（其笛卡尔坐标表示），代表结构的广义坐标（来自第4.1.1节所涉及的键几何）。

\mathbf{p}_k 是该化学键中点处的电子密度梯度（其笛卡尔坐标表示），代表电子行为的广义动量。该梯度通过GPAW计算获得的电子密度梯度场进行高精度三线性插值得到。

所有化学键的状态矢量集合 $\{\mathbf{s}_k\}$ 构成了伪辛相空间中的一个点云。我们计算这个点云的协方差矩阵 \mathbf{C}_s 。协方差矩阵的行列式 $\text{Det}(\mathbf{C}_s)$ 量化了这个点云在六维空间中占据的“体积”，反映了键的几何结构与键中点电子密度梯度之间的耦合涨落程度。为避免数值下溢并更好地处理量级差异，最终特征取其自然对数 $\ln(\text{Det}(\mathbf{C}_s))$ 。在计算中，会引入标准化操作以确保不同量纲特征的可比性，并对协方差矩阵对角线添加小扰动以保证数值稳定性。

$$\text{Log Pseudo Symplectic Fluctuation Volume} = \ln (\text{Det} (\text{Cov}(\{\mathbf{s}_k\})))$$

其中 $\mathbf{s}_k = (\mathbf{q}_k, \mathbf{p}_k)$ 是第 k 个化学键的状态矢量，Cov 表示协方差操作。

在物理上，高log_pseudo_symplectic_fluctuation_volume值意味着晶体内部的结构与电子行为之间存在复杂且广泛的耦合涨落，这可能与材料的非谐性振动、热力学不稳定性以及某些功能特性（如超导、铁电性）的起源机制密切相关。它提供了一种量化“结构-电子耦合强度”的新颖视角。化学上，它揭示了化学键的几何构型如何与其局域电子环境的动态变化相互影响，有助于理解键的柔韧性和反应活性。材料学上，这个特征是设计和筛选具有强结构-电子耦合效应材料的关键指标，尤其适用于探索新型光电材料、热电材料以及超导材料。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“结构-电子耦合涨落空间”的连续、可微量化，是预测材料复杂功能特性、理解材料响应机制以及在数据驱动的材料设计中发现新颖结构-性能关联的关键高级特征，其抽象性使其具有强大的泛化能力。

6.4 图相关路径特征：晶体内部长程关联的图论洞察

本节的全局描述符采用图论路径分析的创新方法，旨在捕捉晶体结构中原子之间更复杂、更长程的相互作用和关联。这些特征通过在晶体图（原子为节点，化学键为边）上采样多种路径，然后聚合路径上原子或键的局部特征，从而量化晶体内部的能量传输、电荷流动、应力传播以及结构异质性等动态关联。与传统的全局平均值不同，路径特征能够揭示空间关联的模式和方向性。为应对复杂晶体图的计算挑战，本方法采用了统一重要性采样策略，以高效地从大量可能路径中选择最具代表性的子集。

6.4.1 路径协变扭转应力 (path_cov_torsional_stress)

path_cov_torsional_stress描述符量化了沿晶体结构中采样路径的“结构-化学不相容性”与“局部扭转应力”之间的协方差。它反映了路径上局部结构畸变程度与其所受力矩之间的线性关联强度。该特征的计算首先在晶体图上采样路径（通过随机行走或带早期停止的完全枚举），然后获取路径上每个原子的结构-化学不相容性（来自第4.4.1节）序列和局部扭转应力（来自第6.3.2.3节中定义的第每个原子的扭转应力大小）序列，最后计算这两个序列的协方差。最终的特征值包括这些路径协方差的均值 (path_cov_torsional_stress_mean)、标准差 (path_cov_torsional_stress_std) 和最大值 (path_cov_torsional_stress_max)，以全面描述这种关联。

$$\text{Cov}(\text{Incomp}_{S-C}, \text{Torsion}) = E[(\text{Incomp}_{S-C} - E[\text{Incomp}_{S-C}])(\text{Torsion} - E[\text{Torsion}])]$$

其中 $\text{Incomp}_{\text{S-C}}$ 代表路径上的结构-化学不相容性， Torsion 代表路径上的局部扭转应力。

在物理上，高协方差表示结构畸变越大，其所承受的力矩也越大，这可能预示着材料内部存在协同的结构不稳定性或应力集中区域，影响材料的力学响应和相变行为。化学上，它揭示了结构中化学键合的局部不匹配如何通过力场传播和累积，形成长程的应力路径。材料学上， $\text{path_cov_torsional_stress}$ 是评估材料韧性、断裂韧性、疲劳强度以及裂纹扩展路径的关键指标，高协方差路径可能指示材料的薄弱环节。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“应力传播路径”和“结构-力学耦合”的量化描述，有助于模型预测材料在极限条件下的行为和失效模式，从而指导材料的结构优化和损伤容限设计。

6.4.2 路径熵 (path_entropy)

path_entropy 描述符量化了沿晶体结构中采样路径的“结构-化学不相容性”分布的香农熵。它衡量了路径上局部结构畸变模式的多样性或随机性。高熵值表示路径上原子或键的结构-化学不相容性（来自第 4.4.1 节）分布广泛，存在多种程度的畸变；低熵值则表示路径上的畸变程度相似，结构更为均一。最终的特征值包括这些路径熵的均值 (path_entropy_mean)、标准差 (path_entropy_std) 和最大值 (path_entropy_max)。

$$H_{\text{path}} = - \sum_{k=1}^M p_k \log_2(p_k)$$

其中 M 是直方图的箱数， p_k 是路径上结构-化学不相容性落在第 k 个箱中的概率。

在物理上， path_entropy 反映了材料内部结构缺陷、局部无序或异质性的空间分布。高熵路径可能预示着能量传输或电荷传输的无序性增加，影响材料的输运性能。化学上，它揭示了沿特定路径的键合环境或配位结构的异质性，有助于理解化学反应的活性位点分布。材料学上， path_entropy 是评估材料缺陷容忍度、无序相形成以及热力学稳定性的重要指标，尤其适用于分析高熵合金、非晶材料和复杂固溶体。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“路径结构异质性”和“无序度”的量化描述，有助于模型预测材料的相变行为、热力学稳定性以及对随机扰动的响应，从而在材料的相设计和性能调控中发挥作用。

6.4.3 路径化学势差 (path_chempot_diff)

path_chempot_diff 描述符量化了沿晶体结构中采样路径的电负性（来自第 4.2.2 节）的最大值与最小值之差。它反映了路径上原子电荷分布或电子吸引能力的宏观波动，预示着潜在的电荷传输通道或电势梯度。高值表示路径上存在显著的电负性（化学势）差异，可能形成电子流动的驱动力；低值则表示路径上的电荷环境相对均匀。最终的特征值包括这些路径化学势差的均值 ($\text{path_chempot_diff_mean}$)、标准差 ($\text{path_chempot_diff_std}$) 和最大值 ($\text{path_chempot_diff_max}$)。

$$\Delta\chi_{\text{path}} = \max_{i \in \text{path}} (\chi_i) - \min_{i \in \text{path}} (\chi_i)$$

其中 χ_i 是路径上原子 i 的电负性。

在物理上，path_chempot_diff与材料的电荷传输、载流子迁移率以及局部电场强度密切相关。高化学势差路径可能指示优先的电子或离子传输通道，例如在电池电极材料或催化剂中。化学上，它揭示了沿特定路径的原子间电荷极化和电子吸引力的不均匀分布。材料学上，path_chempot_diff是设计和优化电化学材料（如电池、燃料电池）、半导体以及催化剂的关键特征，因为它直接关联到材料的电荷输运效率和活性。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“电荷传输路径”和“电荷环境梯度”的量化描述，有助于模型预测材料的电学性能、离子电导率以及催化活性，从而在能源材料和催化剂设计中发挥关键作用。

6.4.4 路径最大扭矩 (path_max_torque)

path_max_torque描述符量化了沿晶体结构中采样路径上所有原子局部扭转应力（来自第6.3.2.3节中定义的每个原子的扭转应力大小）的最大值。它识别了路径上承受最大“力矩”或最显著局部畸变-力耦合的原子位点。高值表示路径上存在一个或多个原子遭受强烈的扭转效应，这可能预示着局部结构不稳定或存在高应变区域。最终的特征值包括这些路径最大扭矩的均值 (path_max_torque_mean)、标准差 (path_max_torque_std) 和最大值 (path_max_torque_max)。

$$\text{MaxTorque}_{\text{path}} = \max_{i \in \text{path}} (\|\mathbf{V}_{\text{struct},i} \times \mathbf{f}_i\|)$$

其中 $\|\mathbf{V}_{\text{struct},i} \times \mathbf{f}_i\|$ 是第 i 个原子的局部扭转应力大小。

在物理上，path_max_torque与材料的局域软模、相变前驱效应以及缺陷诱导的应力集中相关。它能识别晶体结构中对外部刺激最敏感的“弱点”。化学上，它揭示了路径上特定原子或键合环境在力场作用下的最大张力。材料学上，path_max_torque是预测材料早期失效、疲劳裂纹萌生以及设计高应力环境下稳定材料的关键特征，对于理解材料在服役条件下的可靠性至关重要。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“关键应力点”和“最脆弱路径”的量化描述，有助于模型预测材料的极限强度、疲劳寿命以及在极端条件下的结构稳定性，从而在结构材料和功能材料的可靠性设计中发挥关键作用。

6.4.5 路径曲率 (path_curvature)

path_curvature描述符量化了沿着晶体结构中采样路径的键向量序列的平均弯曲程度。它反映了路径在空间中的几何偏离直线性的程度。对于图中的每一条简单路径 $P = (v_1, v_2, \dots, v_L)$ ，我们考虑其连续的键向量 $\mathbf{b}_k = \mathbf{r}_{v_{k+1}} - \mathbf{r}_{v_k}$ 。路径曲率通过计算连续键向量之间的夹角之和来近似，并除以转折点数量进行归一化。这提供了对路径几何复杂性的连续化量化。

其数学定义如下：

$$\text{Curvature}_P = \frac{1}{N_{\text{angles}}} \sum_{k=1}^{L-1} \theta(\mathbf{b}_k, \mathbf{b}_{k+1})$$

其中 N_{angles} 是路径上的转折点数量（即 $L - 2$ ）， $\theta(\mathbf{u}, \mathbf{v}) = \arccos\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}\right)$ 是两个向量之间的夹角。最终的全局特征是对所有采样路径的这些曲率值进行统计分析，包括其均值 (path_curvature_mean)、标准差 (path_curvature_std) 和最大值 (path_curvature_max)，以全面捕捉晶体内部路径弯曲模式的特征。

在物理上，path_curvature反映了沿路径的结构“弯曲”或“折叠”程度。高曲率路径可能意味着该区域存在显著的晶格畸变，或其原子排列更复杂，这可能影响材料的声子传输和热传导。化学上，这个特征与键角应变、构象自由度以及分子链的柔韧性密切相关，对于理解聚合物和生物大分子中的柔性结构尤为重要。材料学上，path_curvature是理解材料柔韧性、塑性变形、晶体结构重构以及声子传播路径的关键指标，例如，在特定方向上具有高曲率的路径可能指示材料在机械载荷下的薄弱环节。在计算机与人工智能领域，它提供了对晶体内部路径几何复杂性的量化，使得机器学习模型能够更准确地预测材料的机械柔韧性、相变行为和能量传输特性，从而在材料的结构设计和性能调控中发挥作用。

6.4.6 路径缠绕范数 (path_wrapping_norm)

path_wrapping_norm描述符量化了沿着晶体结构中采样路径的“周期性边界条件”缠绕程度。它反映了路径在穿越晶胞边界时所积累的“整体位移向量”的欧几里得范数。对于图中的每一条简单路径 $P = (v_1, v_2, \dots, v_L)$ ，我们考虑其连续的键向量在分数坐标下的周期性映像。这些映像向量 $\mathbf{f}_k^{\text{image}}$ （表示键 $v_k \rightarrow v_{k+1}$ 穿越的晶胞整数位移）沿路径累积求和，得到一个总的缠绕向量 $\mathbf{W}_P^{\text{frac}} = \sum \mathbf{f}_k^{\text{image}}$ 。为了获得物理意义明确的笛卡尔坐标下的范数，这个分数坐标下的缠绕向量被转换到笛卡尔坐标系下，通过与晶格矩阵 \mathbf{L} 相乘得到 $\mathbf{W}_P^{\text{cart}} = \mathbf{L} \mathbf{W}_P^{\text{frac}}$ 。最终的路径缠绕范数即为这个笛卡尔缠绕向量的范数 $\|\mathbf{W}_P^{\text{cart}}\|_2$ ，并按路径长度归一化。

其数学定义如下：

$$\begin{aligned} \mathbf{W}_P^{\text{frac}} &= \sum_{k=1}^{L-1} \mathbf{f}_k^{\text{image}} \\ \mathbf{W}_P^{\text{cart}} &= \mathbf{L} \mathbf{W}_P^{\text{frac}} \\ \text{WrappingNorm}_P &= \frac{\|\mathbf{W}_P^{\text{cart}}\|_2}{L} \end{aligned}$$

其中 $\mathbf{f}_k^{\text{image}}$ 是键 k 的周期性映像向量， \mathbf{L} 是晶格矩阵， L 是路径的长度（原子数）。最终的全局特征是

对所有采样路径的这些缠绕范数值进行统计分析，包括其均值 (path_wrapping_norm_mean)、标准差 (path_wrapping_norm_std) 和最大值 (path_wrapping_norm_max)。

在物理上，path_wrapping_norm反映了晶体结构中沿着特定路径的“长程拓扑连接性”或“周期性形变累积”。高缠绕范数路径可能意味着该区域的原子排列受到周期性边界条件的强烈影响，或者存在导致宏观应变累积的微观结构路径。化学上，它与原子在晶格中的迁移路径、扩散机制以及晶体缺陷（如位错）的传播有关，因为这些过程往往伴随着晶胞的穿越。材料学上，path_wrapping_norm是理解材料长程输运性质（如离子扩散、热膨胀）、塑性变形机制以及晶格匹配/失配的重要指标，尤其适用于分析多晶材料的晶界行为和扩散通道。在计算机与人工智能领域，它提供了对晶体内部路径穿越周期性边界复杂性的量化，有助于机器学习模型预测材料的扩散系数、相变机制以及在周期性应力下的结构响应，从而在能源材料和晶格工程设计中发挥关键作用。

6.4.7 路径力场梯度 (path_force_gradient)

path_force_gradient描述符量化了沿着晶体结构中采样路径的原子所受力矢量的变化率或不均匀性。它反映了路径上力场的空间梯度累积，预示着潜在的应力集中、晶格软模或能量传输通道。对于图中的每一条简单路径 $P = (v_1, v_2, \dots, v_L)$ ，我们获取路径上每个原子 v_k 所受的力矢量 \mathbf{f}_k 。路径力场梯度通过计算连续原子所受力矢量之差的范数平方和来近似，并按路径长度归一化。

其数学定义如下：

$$\text{ForceGradient}_P = \frac{1}{L-1} \sum_{k=1}^{L-1} \|\mathbf{f}_{v_{k+1}} - \mathbf{f}_{v_k}\|^2$$

其中 \mathbf{f}_v 是原子 v 所受的力矢量， L 是路径的长度（原子数）。这些力矢量直接来源于DFT计算结果。最终的全局特征是对所有采样路径的这些力场梯度值进行统计分析，包括其均值

(path_force_gradient_mean)、标准差 (path_force_gradient_std) 和最大值

(path_force_gradient_max)。

在物理上，path_force_gradient反映了晶体内部力场的“崎岖性”和局部应力分布的剧烈程度。高梯度路径可能意味着该区域存在显著的晶格畸变、原子位移或相变前驱效应，这些都与材料的力学稳定性和声子行为密切相关。化学上，它与原子间相互作用势能面的陡峭程度和键合刚度变化有关。材料学上，path_force_gradient是预测材料力学性能（如硬度、断裂韧性）、热输运（通过声子散射）以及理解结构相变机制的关键指标，例如在超弹性材料中，力场的非线性梯度对其柔韧性至关重要。在计算机与人工智能领域，它提供了对晶体内部“力场路径”和“应力波传播”的量化描述，有助于机器学习模型预测材料的机械响应、热导率以及在不同应力条件下的结构稳定性，从而在结构材料和功能材料的设计中发挥关键作用。

6.4.8 路径结构自相关 (path_structure_autocorr)

path_structure_autocorr描述符量化了沿着晶体结构中采样路径的“结构-化学不相容性”序列（来自第4.4.1节）的一阶自相关系数。它反映了路径上局部结构畸变程度的“记忆性”或“连续性”，即一个原子的不相容性与其下一个近邻原子的不相容性之间的线性相关强度。高正值表示相邻原子具有相似的结构畸变程度，低值或负值则表示畸变模式在路径上变化较大。最终的全局特征是对所有采样路径的这些自相关系数进行统计分析，包括其均值 (path_structure_autocorr_mean)、标准差 (path_structure_autocorr_std) 和最大值 (path_structure_autocorr_max)。

$$\text{Autocorr}_P(\text{Incomp}_{\text{S-C}}) = \frac{\text{Cov}(\text{Incomp}_{\text{S-C},k}, \text{Incomp}_{\text{S-C},k+1})}{\sigma(\text{Incomp}_{\text{S-C},k})\sigma(\text{Incomp}_{\text{S-C},k+1})}$$

其中 $\text{Incomp}_{\text{S-C},k}$ 是路径上第 k 个原子的结构-化学不相容性。

在物理上，path_structure_autocorr反映了晶体内部结构畸变或应变场的空间连续性。高自相关性路径可能指示晶体中存在长程有序的畴、共价网络或特定晶向上的刚性链，这影响材料的晶格稳定性、声子传输和宏观各向异性。化学上，它揭示了化学键合环境的相似性或差异性如何沿特定路径传播。材料学上，path_structure_autocorr是理解材料微观结构同质性、相界行为、缺陷分布以及晶体生长模式的关键指标，对于控制材料微观结构以获得期望的宏观性能至关重要。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“结构序列模式”和“空间相关性”的量化描述，有助于模型识别晶体中的特定结构 motif、预测材料的组织结构演变以及在不同合成条件下的形态学特征，从而在材料的结构设计和工艺优化中发挥作用。

6.4.9 路径电荷-电势协方差 (path_charge_potential_cov)

path_charge_potential_cov描述符量化了沿着晶体结构中采样路径的原子Bader电荷（来自第3.3.1节）与原子位点处静电势（来自第6.2.3节的静电势均值，后经插值得到）之间的协方差。它反映了路径上电荷分布与电势环境之间的线性关联强度，预示着潜在的电荷传输效率或电化学反应活性。高正协方差表示电荷富集区域伴随着高电势，或电荷耗散区域伴随着低电势；负协方差则可能指示电荷转移的方向性。最终的全局特征是对所有采样路径的这些协方差值进行统计分析，包括其均值 (path_charge_potential_cov_mean)、标准差 (path_charge_potential_cov_std) 和最大值 (path_charge_potential_cov_max)。

$$\text{Cov}(\text{Charge}, \text{Potential}) = E[(\text{Charge} - E[\text{Charge}])(\text{Potential} - E[\text{Potential}])]$$

其中 Charge 代表路径上的Bader电荷，Potential 代表路径上的静电势。

在物理上，path_charge_potential_cov与材料的电荷输运机制、电子散射过程以及局部电场强度密切相关。高协方差路径可能指示优先的电子或离子传导通道，或具有极高化率的区域。化学上，它揭示了沿

特定路径的电荷转移模式和化学势梯度，有助于理解化学反应的电荷传递步骤。材料学上，`path_charge_potential_cov`是设计和优化电化学材料（如电池、超级电容器）、半导体以及催化剂的关键特征，因为它直接关联到材料的电荷输运效率、能量存储能力和表面反应活性。在计算机与人工智能领域，它为机器学习模型提供了对材料内部“电荷-电势路径”和“电荷输运微环境”的量化描述，有助于模型预测材料的电导率、电容性能以及催化活性，从而在能源材料和电子器件设计中发挥关键作用。

第七章：结果与讨论：多尺度拓扑-量子-图论融合描述符的验证与发现

本章旨在系统地展示并深入讨论本研究中提出的多尺度拓扑-量子-图论融合描述符在表征周期性材料结构与性能方面的有效性、敏感性及发现潜力。我们以三例典型的卤素钙钛矿材料——立方相 CsPbBr_3 (Pm-3m)、正交相 CsPbI_3 (Pnma) 和初始被标记为立方Pm-3m的 CsPbCl_3 为案例。通过细致比较这些材料在0-单纯形、1-单纯形、2-单纯形直至全局层面的关键特征统计量，我们严谨地验证了描述符捕捉结构对称性破缺与畸变的能力，并揭示了一项至关重要的结构真相，此事件充分彰显了本框架作为“结构探测器”的独特价值。

7.1 对称性基准： CsPbBr_3 (Pm-3m) 的特征分析

CsPbBr_3 在室温下稳定于立方Pm-3m空间群，是一种典型的高对称性、无显著结构畸变的钙钛矿。因此，本节分析旨在验证我们的描述符在理想对称体系中的正确行为和零值敏感性。那些专门设计用于量化结构不对称性、畸变和应力的描述符，其平均值应趋近于零，且方差也应极小。

7.1.1 0-单纯形特征：原子局部环境的均匀性

针对 CsPbBr_3 ，其0-单纯形描述符的分析结果清晰地展现了高对称性特征。例如，`vectorial_asymmetry_norm_sq`的平均值与方差均为0.000000，精确验证了该特征能够准确识别立方相中原子局部配位环境的高度对称性。同理，`symmetry_breaking_quotient`的平均值为1.000000，方差为0.000000，表明所有原子位点均完美保持了其理想的局部对称性。

7.1.2 1-单纯形特征：化学键环境的规则性

在1-单纯形层面，`lie_algebra_incompatibility`的平均值与方差均为0.000000。这一关键结果明确指出，在立方相中，所有Pb-Br化学键的局部环境是完全兼容的，不存在扭转或剪切应力，与理想八面体网络的高度协同性完全吻合。

7.1.3 2-单纯形特征：三体团簇的理想构型

2-单纯形描述符同样准确反映了CsPbBr₃的结构对称性。avg_vectorial_asymmetry 和 max_vectorial_asymmetry 的平均值和方差均为0.000000。hierarchical_stress_flow 的平均值和方差也均为0.000000，这表明在任意三体团簇内部，由键不兼容性产生的应力流是完全平衡的。

7.1.4 全局特征：整体宏观性质的无畸变体现

在全球层面，lattice_anisotropy_ratio 的平均值为1.000000，与立方晶胞特性完美吻合。global_asymmetry_norm 和 total_torsional_stress 的平均值均为0.000000，有力验证了整个晶体中没有累积的局部不对称性或扭转应力。octahedral_count 为8.000000，准确表明结构中存在8个完整的、未畸变的PbBr₆八面体。这些结果共同且无可辩驳地证明了我们的描述符框架能够精确、零误差地识别和量化理想对称结构，为后续分析奠定了可靠基线。

7.2 量化已知畸变：CsPbI₃ (Pnma) 的特征分析

CsPbI₃稳定于正交Pnma空间群，具有显著的八面体倾斜和结构畸变。本节将展示描述符如何定量地捕捉这些已知的对称性破缺。与CsPbBr₃的零值基线形成鲜明对比，CsPbI₃的许多不对称性和应力相关特征呈现出显著的非零平均值和更大的方差。

7.2.1 0-单纯形特征：原子局部环境的畸变

vectorial_asymmetry_norm_sq 的平均值为0.557791，方差为0.199638。这一显著的非零平均值明确指示了原子局部配位环境偏离理想立方对称性的程度，其非零方差则表明不同原子位点的畸变程度存在空间差异。

7.2.2 1-单纯形特征：化学键环境的扭曲

lie_algebra_incompatibility 的平均值高达15.998363，方差为183.027194。与CsPbBr₃的零值相比，这一巨大的非零平均值和显著的方差明确指出，在正交相中，Pb-I化学键的局部环境发生了剧烈的扭转和剪切，精确反映了复杂的八面体倾斜模式所导致的局部应变异质性。

7.2.3 2-单纯形特征：三体团簇的复杂畸变

avg_vectorial_asymmetry 的平均值为0.622157，hierarchical_stress_flow 的平均值为27.546488。这些显著的非零值表明，构成三角形的三体团簇环境是高度非对称的，且存在显著的、不平衡的应力流。

7.2.4 全局特征：整体宏观性质的各向异性

lattice_anisotropy_ratio 的平均值为1.859983，远大于CsPbBr₃的1.0，直接量化了正交晶胞的几何特征。global_asymmetry_norm (0.644180) 和 total_torsional_stress (0.092686) 的显著非零值再次验证了整个晶体结构中普遍存在的局部不对称性及累积的扭转应力。

7.3 敏感性案例研究：CsPbCl₃ 的结构揭示

7.3.1 描述符与Pm-3m标签的不一致性

在分析报告中，CsPbCl₃最初被标记为立方Pm-3m空间群。然而，本框架的描述符分析结果却呈现出与理想立方相预期行为系统性的、显著的偏差。

7.3.1.1 0-单纯形特征：强烈的局部畸变信号

尽管标称空间群为Pm-3m，vectorial_asymmetry_norm_sq 的平均值却高达3.437875，远超Pnma相CsPbI₃的0.557791，且其极高方差（18.007008）表明局部畸变分布极不均匀。这构成了第一个重要的不一致信号。

7.3.1.2 1-单纯形与2-单纯形特征：系统性的应力与非对称性

在1-单纯形和2-单纯形层面，lie_algebra_incompatibility（平均值0.570241）、avg_vectorial_asymmetry（平均值2.462962）和 hierarchical_stress_flow（平均值0.975182）等特征均呈现显著的非零值，与理想立方晶格的预期相悖，系统性地指向了结构中存在不可忽视的应力与非对称性。

7.3.2 描述符作为“结构探测器”的验证

面对上述所有描述符呈现的系统性偏差，我们采取了“反向推理”的逻辑：即这些偏差并非源于描述符的计算错误，而是输入数据本身物理真实的反映。这一逻辑促使我们对输入的CsPbCl₃结构文件本身进行了复核。

最终的调查确认：原始用于计算的CsPbCl₃结构文件实际上对应于正交Amm2空间群，而非立方Pm-3m。这一发现源自一个数据处理流程中的标注错误。

本案例研究的结论，不仅成功纠正了一个重要的结构错误，更强有力地验证了我们多尺度融合描述符框架的卓越敏感性和作为“结构探测器”的独特能力：

- 超越显式标签的洞察力：**描述符能够穿透标签的表象，直达结构最真实的几何和物理内涵，提供超越传统晶体学分类的深层洞察。

2. **白盒可解释性**：当描述符给出非预期值时，我们能够清晰地追溯其物理源头（如几何畸变、键扭曲应力等），这种固有的可解释性是黑盒模型所不具备的。
3. **自验证机制**：本次发现是一次描述符的自验证。描述符在理想对称体系上表现完美，在已知畸变体系上精确量化畸变，而在“错误标签”体系上发出“警报”，最终成功揭示了真实结构，证明了本框架的稳健性与科学严谨性。

7.4 综合讨论：多尺度特征的协同效应与物理洞察

7.4.1 多尺度描述的互补性与层级信息流

本框架通过多尺度分层设计，实现了从微观到宏观信息的无缝流动与深度整合。0-单纯形描述符捕捉原子尺度的局部畸变；1-单纯形描述符量化化学键层面的扭曲和应力；2-单纯形描述符聚焦于三体团簇的高阶效应；全局描述符则将这些局部信息进行高阶整合和抽象。这种层层递进的体系，使得我们不仅能够判断“结构是否畸变”，更能精确地定位“在哪里畸变”、“以何种方式畸变”以及“这种畸变如何影响整体性能”。

7.4.2 方差数据的价值：揭示结构异质性与涨落

除了平均值，特征的方差在本次分析中也发挥了至关重要的诊断作用。对于CsPbI₃和实际为Amm2的CsPbCl₃，许多关键特征不仅平均值显著非零，方差也普遍较大。这些非零方差揭示了结构内部的异质性，即不同原子、化学键或三体团簇的畸变程度是不均匀的。它们量化了结构微观涨落和内在复杂程度，为预测材料的动态行为、缺陷容忍度及非平衡响应提供了关键信息。

7.4.3 描述符的普适性与白盒优势

本研究结果充分证明了我们构建的多尺度融合描述符框架的普适性与白盒优势。无论晶体对称性高低、是否存在已知畸变或数据标注错误，本框架都能稳定、准确地提供具有深刻物理意义的量化特征。其固有的可解释性使得科学家能够理解材料性质的物理根源，从而极大加速科学知识的发现和基于原理的理性材料设计。

7.5 本章总结

本章通过对三种钙钛矿的详尽对比分析，全面验证了多尺度拓扑-量子-图论融合描述符的卓越性能。我们清晰地展示了该框架不仅能够精确识别和量化理想与畸变结构，更通过成功揭示CsPbCl₃的真实结构，强有力地证明了其作为结构探测器的强大敏感性和发现能力。本章深入强调了多尺度描述符在不同层级上捕捉结构信息的协同效应，以及方差数据在揭示结构内部异质性与动态涨落方面的重要价值。这些分析结果共同奠定了本框架作为一种高度可解释、高敏感度、普适性强的材料基因组表征工具的坚实基础。

第八章：综合分析与未来展望：多尺度融合描述符的价值、影响与未来图景

本章对本研究提出的多尺度融合描述符框架进行全面的综合分析，阐明其在材料科学领域的核心价值、深远影响与未来图景。我们将总结该框架如何通过其创新性设计，超越传统材料表征的局限，并详细阐述其高通量虚拟筛选和AI驱动的逆向设计两大核心应用场景中的变革潜力。

8.1 多尺度融合描述符的核心价值与突破

本研究的核心在于构建了一套能够从0-单纯形（原子）到全局层面（晶体整体）层层递进、逐级抽象且信息高度互联的描述符体系。这些描述符深度融合了晶体几何、量子化学信息，并开创性地引入了高阶现代数学概念，实现了对材料微观结构、电子行为及其复杂耦合的精细化、定量化、高度可解释性表征。

8.1.1 超越传统描述符的局限性

本框架通过以下关键方面显著超越了传统描述符的局限性与“黑箱”困境：

- 多尺度连续性与信息流：** 框架实现了从原子到整体的无缝信息衔接与层级抽象，使得从微观物理现象到宏观材料性能的溯源与解释成为可能。
- 固有物理可解释性：** 每一个描述符都拥有明确的数学定义和深刻的物理意义。这种透明性使得模型输出不再是难以理解的数字，而是可以直接与材料微观机制关联的科学洞察，极大地促进了“知识发现”。
- 连续可微性与优化潜力：** 核心特征的数学构造确保了其连续可微性，使得我们的特征空间成为一个可被梯度优化算法高效探索的“设计空间”，这是传统离散描述符难以企及的。
- 高敏感性与发现能力：** 正如第八章案例所展示的，本框架能够以惊人的敏感度捕捉到数据中隐藏的、细微的结构异常，展现了其作为“结构探测器”和“科学发现工具”的独特价值。

8.1.2 晶体材料“基因组”的深度编码

本描述符框架旨在为周期性材料提供一套“深度基因组”编码，实现材料的精确“指纹识别”，揭示结构-性能的内在关联，并为理性设计与智能调控提供坚实基础。

8.2 核心应用场景与变革潜力

8.2.1 高通量虚拟筛选：加速材料发现的“白盒”引擎

本框架提供了一种全新的“白盒”高通量虚拟筛选范式。通过构建一个庞大而丰富的材料特征数据库，并以此训练高效的机器学习代理模型，我们能够对任何新的候选材料进行瞬时性能评估。这种“即插即用”的能力将显著减少传统的实验试错成本，直接为高通量合成实验提供精准指导，实现从“海量筛选”到“精准筛选”的飞跃。

8.2.2 AI驱动的反向设计：从性能到结构的知识驱动路径

材料的反向设计，即从期望性能反向推导其原子结构，是材料科学的终极目标之一。本框架为此提供了前所未有的“知识驱动”路径：

- 构建可搜索的“特征空间”：** 我们构建了一个高维度、连续可微、可解释且物理真实的特征空间。一个理想的性能目标可以被明确地转化为一个具体的目标特征向量，即一个可操作的“材料设计蓝图”。
- 智能生成算法的“搜索空间”优化：** 本特征空间是各种先进生成式AI算法（如GA, RL, VAE）的理想“搜索空间”。这些算法可以在此空间中进行高效的优化搜索，而非在离散的原子结构空间中盲目探索。
- 特征向量到结构的“高保真解码”：** 一旦找到最优的“特征蓝图”，下一步就是将其“翻译”回一个物理可实现的晶体结构。我们的高度可解释性特征为这个解码过程提供了关键的物理约束和指导，使得生成的结构更可能具有物理合理性和化学稳定性。
- “知识驱动”的闭环设计：** 本研究促成了一个良性循环的“知识驱动”智能设计范式：从微观原理中提取知识（构建特征），利用知识训练AI（预测与设计），再通过对生成蓝图的物理解释来深化科学理解，从而形成持续迭代、自我优化的科学发现闭环。

这种“精确设计、智能发现”的范式有望彻底改变材料科学的研究模式，将其从传统的“试错式”探索提升到一个全新的、更高效、更具洞察力的新时代。

8.3 未来展望：理论深化与应用拓展

8.3.1 理论深化与模型扩展

- 更高阶拓扑结构融入：** 探索将更高阶的单纯形（如3-单纯形）和更深层的代数拓扑概念（如持久同调）融入描述符体系，以捕捉孔隙、通道等更高级的几何拓扑特征。
- 时间维度与动态过程的量化：** 将框架扩展到非静态体系，如分子动力学轨迹，以量化材料的动态行为、相变路径和输运过程。

3. **非晶与准晶材料的普适性表征：** 进一步普适化框架，使其能够有效描述非晶态、准晶态等缺乏长程有序性的复杂体系。
4. **不确定性量化与鲁棒性评估：** 开发描述符的概率版本或引入贝叶斯方法，量化特征值中的不确定性，并评估其对计算参数的鲁棒性。

8.3.2 应用拓展与跨领域融合

1. **力学性能的精准预测：** 结合力场和应力相关特征，开发更精细的机器学习模型，精准预测材料的弹性模量、硬度、断裂韧性等，为设计高性能结构材料提供“基因型”指导。
2. **功能材料的定制设计：** 将描述符应用于特定功能材料（如催化剂、热电材料、电池材料）的定制设计，通过在特征空间中优化，实现“功能先行”的设计理念。
3. **实验验证与数据反馈闭环：** 建立从描述符预测到实验合成、表征、再到数据反馈的紧密闭环研究流程，实现理论-计算-实验的无缝衔接。
4. **与深度生成模型的集成：** 探索与先进的深度生成模型（如Graph Neural Networks, Diffusion Models）的无缝集成，实现从特征空间到原子坐标的端到端、高保真度逆向解码。
5. **跨学科知识的深度融合：** 持续加强与凝聚态物理、量子化学、工程力学等领域的交叉融合，将更前沿的理论和计算方法引入描述符设计，例如，整合电子带结构、声子谱等更详细的量子信息。

第九章：结论

本研究开创性地提出并构建了一套周期性材料的多尺度拓扑-量子-图论融合描述符框架，旨在克服传统材料表征方法在精确捕捉复杂微观结构、电子行为及其深层耦合方面的固有局限。该框架通过将晶体结构抽象为多尺度单纯复形，并创新性地融合了几何学、量子化学信息以及源自代数拓扑、李代数、辛几何和图论中的高阶数学概念，成功开发出一系列连续可微且具有深刻物理可解释性的材料特征。这些描述符实现了从0-单纯形（原子）、1-单纯形（化学键）、2-单纯形（三体团簇）直至全局（晶体整体）的全面、精细化基因组编码。

通过对立方相 CsPbBr_3 、正交相 CsPbI_3 以及具有错误标签的 CsPbCl_3 （实际为正交 $\text{Amm}2$ 相）的详尽对比分析，本研究严谨地验证了该描述符框架的卓越性能。结果清晰表明，该框架不仅能够精确量化理想与畸变结构，更以其高度敏感性成功揭示了数据中隐藏的结构异常，强有力地证明了其作为“结构探测器”的独特价值。

本描述符框架的核心价值在于为材料科学领域提供了一个兼具效率与洞察力的定量分析工具。与依赖高昂计算成本的纯第一性原理筛选或缺乏深层物理内涵的传统经验描述符相比，本框架通过提供一套高效且富有洞察力的材料基因组编码，显著提升了高通量实验的可行性与成功率。它不仅为高通量虚拟筛选

提供了物理驱动的“白盒”引擎，能够极大加速新材料的发现，更重要的是，它为人工智能驱动的逆向设计构建了一个信息丰富、物理真实的特征空间，使得从期望性能到实际结构的精准反演成为可能。

综上，本研究的成果是推动材料科学向更智能化、知识驱动型研究范式转变的重要里程碑，为未来在材料设计、性能预测和机理阐明方面开启了广阔的前景。