

数据论文

面向软晶格筛选的立方钙钛矿体模量
可解释性描述符研究*姜锦铭 孙庆德[†] 张卫兵[‡]

(长沙理工大学物理与电子科学学院, 柔性电子材料基因工程湖南省重点实验室, 长沙 410114)

(2025年5月18日收到; 2025年7月3日收到修改稿)

近年来, 软晶格被认为是钙钛矿材料实现缺陷容忍性的主要物理来源, 体模量则作为晶格“软硬度”的关键衡量指标. 本文针对立方钙钛矿体系, 基于 SISO (sure independence screening and sparsifying operator) 与 VS-SISO (集成迭代变量选择机制的 SISO 框架) 方法, 构建了两类低维、物理可解释性强的体模量预测模型. 首先, 基于共价半径、熔点和体积等结构和热力学特征构建的热-结构耦合描述符模型, 在测试集上实现了均方根误差 $RMSE = 7.41$ GPa, 决定系数 $R^2 = 97.8\%$ 的良好预测性能; 进一步引入电负性、原子价态与未配对电子数等电子层级特征后, 构建了电子-热-结构三重耦合描述符模型, 预测精度显著提升, 在测试集上 $RMSE$ 降至 5.34 GPa, R^2 提升至 98.35% . 基于该模型, 本文对超过 10000 个卤族和硫族立方钙钛矿进行高通量预测, 筛选出约 170 种体模量位于 $10\text{--}20$ GPa 区间、与 Pb-I 钙钛矿相近的候选体系. 研究结果为软晶格机制在无铅体系中的适用性提供了初步支持, 并为高通量筛选具缺陷容忍潜力的稳定无铅钙钛矿材料提供了理论依据与数据支撑. 本文数据集可在 <https://doi.org/10.57760/sciencedb.j00213.00161> 中访问获取.

关键词: 体模量, 缺陷容忍性, 软晶格, 钙钛矿, 可解释机器学习

PACS: 02.60.Pn, 71.15.Mb, 88.40.H-, 07.05.Mh

DOI: 10.7498/aps.74.20250652

CSTR: 32037.14.aps.74.20250652

1 引言

近年来, 以 $\text{CH}_3\text{NH}_3\text{PbI}_3$ 为代表的有机-无机杂化钙钛矿太阳能电池, 其光电转换效率已提升至接近 27% ^[1], 可与晶体硅太阳能电池相媲美. 这一突破主要得益于钙钛矿材料优异的光电性能, 包括近似最优的带隙、高光吸收系数, 以及电子与空穴的双极性迁移能力^[2]. 其中, 材料本征的缺陷容忍性被认为是实现高效率的关键因素之一, 能够显著延长载流子寿命和扩散长度, 从而提高电荷提取与收集效率, 促进器件性能的提升^[3,4].

为理解钙钛矿的缺陷容忍机制, 研究者从不同物理视角提出了多种理论模型. 一种典型观点认为, Pb 原子的 6s 孤对电子与卤素 p 轨道之间形成强反键合耦合, 抬升了价带顶 (VBM) 并增强带边弥散性, 有助于形成浅受主缺陷^[5]; 此外, Pb 的自旋-轨道耦合 (spin-orbit coupling, SOC) 效应被认为能够下移导带底 (CBM), 抑制缺陷诱导的局域晶格畸变, 从而降低施主缺陷能级^[6]. 在电子结构机制之外, 研究还发现该体系中易形成大极化子, 其局域晶格极化可屏蔽缺陷散射并抑制非辐射复合, 有效提高了缺陷容忍性^[7-9]; 这一行为本质上源于钙钛矿所具有的软晶格特性: 软晶格结构更易发

* 国家自然科学基金 (批准号: 12474219) 和湖南省自然科学基金 (批准号: 2023JJ40041) 资助的课题.

[†] 通信作者. E-mail: qingdesun@csust.edu.cn

[‡] 通信作者. E-mail: zhangwb@csust.edu.cn

生局域畸变, 促进极化子形成; 同时, 低频声子与载流子耦合较弱, 有助于延长载流子寿命^[10-13]. 其中, 较低的体模量可作为软晶格的关键量化指标.

尽管上述理论从不同角度有效解释了铅卤钙钛矿的缺陷容忍性, 但其在非铅钙钛矿中的适用性仍缺乏系统验证. 鉴于铅卤钙钛矿存在稳定性不足及 Pb 的毒性问题, 研究者正积极探索高性能、环境友好、结构稳定的无机非铅钙钛矿作为其替代方案^[14-18]. 基于此, 本文以体模量作为衡量晶格“软硬度”的关键结构参数, 筛选出与 Pb-I 钙钛矿具有相近体模量的无铅钙钛矿候选材料, 识别具备软晶格特征的体系, 并进一步探索其缺陷容忍潜力, 从而验证软晶格机制在非铅体系中的适用性, 为新型稳定、环保钙钛矿的设计提供理论支持.

近年来, 机器学习技术被广泛应用于材料性质预测, 包括体模量等关键物性的建模. 目前已有多种模型被开发用于体模量预测, 主要可分为 3 类: 第 1 类是传统特征工程结合机器学习的方法, 如 Dummy^[19], RF^[20], XGBoost^[21] 和 CatBoost^[22], 依赖元素属性及人工构建的描述符进行建模; 第 2 类为非图结构的深度学习模型, 如 MODNet^[23] 和 CrabNet^[24], 通过端到端方式从元素组合中自动学习有效特征; 第 3 类则是基于图神经网络 (GNN) 的模型, 通过晶体结构构建图并进行结构感知学习, 代表性模型包括 CGCNN^[25], SchNet^[26], MegNet^[27], ALIGNN^[28], DimeNet++^[29], Matformer^[30] 和 coNGN^[31].

在统一数据集 MatBench(包含来自 Materials Project 的 10987 个晶体结构)^[19,32] 上的基准评估表明, coNGN 模型在预测对数体模量 $\log_{10}K_{\text{VRH}}$ (单位 GPa) 方面表现最优, 其平均绝对误差 (mean absolute error, MAE) 低至约 0.05. GNN 模型在结构信息充分的前提下具备较高的预测精度, 但其物理可解释性相对较弱; 相比之下, 传统机器学习模型虽预测精度略低, 却更适于进行特征归因与机制分析.

值得注意的是, 立方相钙钛矿具有相同的晶体结构框架, 仅化学组成发生变化, 因此即使忽略结构信息, 仅依赖元素特征也能获得良好的预测效果. Akinpelu 等^[33] 曾采用传统机器学习方法 (CatBoost, XGBoost, RF) 结合 SHAP(shapley additive explanations) 分析, 建立了 199 种钙钛矿材料的体模量预测模型, 其中 RF 模型表现最佳 (MAE =

7.79 GPa, $R^2 = 0.96$). 分析表明, 弹性常数是影响体模量的关键特征之一, 但其作为一种非易得参数, 通常需依赖第一性原理计算获得, 难以在高通量材料筛选中广泛适用.

因此, 本文的核心问题是: 能否构建一种仅依赖易获取特征的体模量简易描述符, 兼具高预测精度与物理可解释性. SISSO(sure independence screening and sparsifying operator) 方法作为一种结合符号回归与压缩感知的机器学习框架, 能够从海量特征组合中筛选出形式简单、物理意义明确的解析表达式, 在保证良好预测性能的同时具备出色的可解释性. 本文将采用 SISSO 方法^[34] 及其扩展形式 VS-SISSO(集成迭代变量选择机制的 SISSO 框架)^[35] 对候选特征进行逐步筛选与建模, 构建立方钙钛矿体模量的高精度预测公式. 该方法不仅能够实现对体模量的高效预测, 为发现具备 Pb-I 钙钛矿类软晶格特征的无铅候选材料提供有效工具, 同时揭示了影响立方钙钛矿体模量的关键元素特征, 为缺陷容忍性理论在新材料体系中的拓展应用奠定基础.

2 数据与方法

2.1 数据集

以立方相钙钛矿结构 (空间群 $Pm\bar{3}m$) 的 CsPbI_3 和 CaTiO_3 为原型, 通过元素替换构建出 24 种 $\text{A}^+\text{B}^{2+}\text{X}_3$ (其中 A 位为一价金属, B 位为二价金属, $\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{I}$) 的卤族单钙钛矿, 以及 27 种 $\text{A}^{2+}\text{B}^{4+}\text{Y}_3$ (其中 A 位为二价金属, B 位为四价金属, $\text{Y} = \text{O}, \text{S}, \text{Se}$) 的硫族单钙钛矿. 在满足八电子守恒原则的前提下, $\text{A}^+\text{B}^{2+}\text{X}_3$ 型钙钛矿可通过“ $2\text{B}^{2+} \rightarrow \text{B}^+ + \text{B}^{3+}$ ”的化学演化机制, 生成 108 种 $\text{A}^+\text{B}^+\text{B}^{3+}\text{X}_6$ 型卤族双钙钛矿 (B^+ 为一价金属, B^{3+} 为三价金属); 同理, $\text{A}^{2+}\text{B}^{4+}\text{Y}_3$ 型钙钛矿可通过“ $2\text{B}^{4+} \rightarrow \text{B}^{3+} + \text{B}^{5+}$ ”演化生成 54 种 $\text{A}^{2+}\text{B}^{3+}\text{B}^{5+}\text{Y}_6$ 型硫族双钙钛矿 (B^{3+} 为三价金属, B^{5+} 为五价金属). 图 1 所示为上述 4 类钙钛矿材料的元素分布情况.

体模量通过第一性原理计算得到的能量-体积 ($E-V$) 关系曲线进行拟合而获得. 具体而言, 首先在晶格缩放系数范围为 0.95—1.03 的区间内均匀插值生成 9 个不同的体积点, 计算不同体积下的总能, 获得 $E-V$ 数据. 随后对该数据拟合三阶 Birch-Murnaghan 状态方程 (EOS)^[36,37], 并对拟合

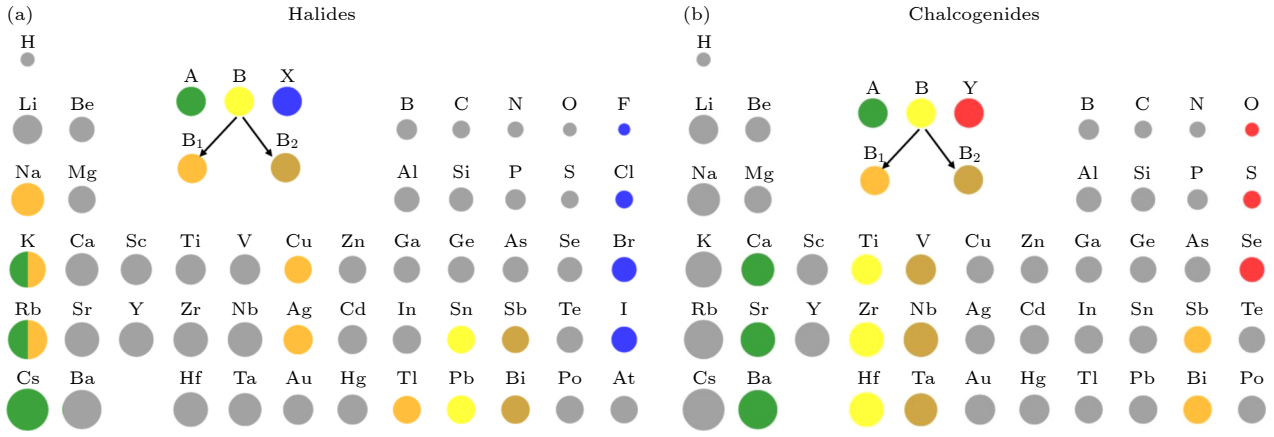


图 1 (a) 卤族立方钙钛矿包括典型的单钙钛矿 ABX_3 ($B = M^{2+}$) 和双钙钛矿 $A_2B_1B_2X_6$ ($B_1 = M^+$, $B_2 = M^{3+}$); (b) 硫族立方钙钛矿包括 ABY_3 ($B = M^{4+}$) 型单钙钛矿和 $A_2B_1B_2Y_6$ ($B_1 = M^{3+}$, $B_2 = M^{5+}$) 型双钙钛矿; 图中展示了数据集中 213 个钙钛矿的元素组成分布, 其中双色半圆表示该元素既可作为 A 位也可作为 B 位

Fig. 1. (a) Halide cubic perovskites include typical ABX_3 -type single perovskites ($B = M^{2+}$) and $A_2B_1B_2X_6$ -type double perovskites ($B_1 = M^+$, $B_2 = M^{3+}$); (b) chalcogenide cubic perovskites include ABY_3 -type single perovskites ($B = M^{4+}$) and $A_2B_1B_2Y_6$ -type double perovskites ($B_1 = M^{3+}$, $B_2 = M^{5+}$); the figure summarizes the elemental composition of 213 cubic perovskites in the dataset. Dual-colored semicircles indicate elements that can occupy both A and B sites.

后的 E - V 方程关于体积求导, 得到压力-体积 (P - V) 关系曲线, 进一步提取体模量. B-M 状态方程如下所示:

$$E(V) = E_0 + \frac{9V_0B_0}{16} \left\{ \left[\left(\frac{V_0}{V} \right)^{(2/3)} - 1 \right] B'_0 + \left[\left(\frac{V_0}{V} \right)^{(2/3)} - 1 \right]^2 \left[6 - 4 \left(\frac{V_0}{V} \right)^{(2/3)} \right] \right\}, \quad (1)$$

$$P(V) = \frac{3B_0}{2} \left[\left(\frac{V_0}{V} \right)^{(7/3)} - \left(\frac{V_0}{V} \right)^{(5/3)} \right] \times \left\{ 1 + \frac{3}{4} (B'_0 - 4) \left[\left(\frac{V_0}{V} \right)^{(2/3)} - 1 \right] \right\}, \quad (2)$$

$$B_0 = -V \left(\frac{\partial P}{\partial V} \right)_{P=0}, \quad (3)$$

$$B'_0 = \left(\frac{\partial B}{\partial P} \right)_{P=0}, \quad (4)$$

其中, $E(V)$ 和 $P(V)$ 分别表示体积变化下的总能量与压力; 拟合参数 E_0 , V_0 , B_0 , B'_0 分别为平衡状态下的总能、体积、体模量以及体模量关于压力的一阶导数. 所有第一性原理计算均在 VASP 软件中进行^[38], 采用的计算参数包括: 平面波截断能为 400 eV, 电子自洽迭代的能量收敛精度设为 10^{-4} eV, 离子弛豫的力收敛准则为 0.02 eV/Å, K 点网格密度设为 $2\pi \times 0.025 \text{ \AA}^{-1}$.

以立方相钙钛矿材料的体模量作为预测目标, 本文采用 Matminer^[39] 工具包中提供的 22 项元素

特征 (如原子序数、原子半径、原子质量等), 对材料化学式中涉及的各元素属性进行统计处理, 提取其最小值 (min)、最大值 (max)、极差 (range)、平均值 (mean)、平均偏差 (average deviation) 和众数 (mode) 6 种统计量, 从而为每个材料构建共计 132 个统计学描述符. 进一步引入元素价态 (n_A , n_B , n_X) 与原子半径 (r_A , r_B , r_X) 等 6 项个体特征, 最终形成包含 138 维的材料描述符集合, 用于机器学习模型训练. 需特别指出的是, 对于双钙钛矿结构中 B 位由两种不同元素共同占据的情况, 其相关特征取两者的平均值. 最终建立了一个包含 213 个立方钙钛矿体模量数据的训练数据集.

2.2 模型参数优化与评估

本文所采用的机器学习模型基于 Scikit-learn (sklearn)^[40] 平台, 包括随机森林 (random forest, RF)、支持向量回归 (support vector regression, SVR)、核岭回归 (kernel ridge regression, KRR) 以及极端随机树 (extremely randomized trees, EXT). 各模型参数均经调优设定, 具体如表 1 所列. 其中, RF 和 EXT 模型的 $n_estimators$ 分别设置为 200 和 110, $max_features$ 设置为 “auto”, 并在 RF 中启用 $bootstrap = True$; SVR 模型采用 “rbf” 核函数, 惩罚系数 $C = 497$, 核函数参数 $gamma = 0.01$; KRR 模型使用 “poly” 核函数, 正则化参数 $alpha = 0.001$.

表 1 各类机器学习模型参数设置
Table 1. Machine learning model parameters.

模型	参数设置
RF	n_estimators=200, max_features="auto", bootstrap=True
SVR	kernel="rbf", C=497, gamma=0.01
KRR	kernel="poly", alpha=0.01
EXT	n_estimators=110, max_features="auto"

通过上述模型训练得到的最优模型中的重要特征被提取为描述符, 并作为 SISO 模型的输入特征. 为确保 SISO 与 VS-SISO 两种模型之间的可比性, 统一将描述符维度设为 1, 并在 213 个样本基础上采用回归方式预测立方钙钛矿的体模量. 特征构造中引入了如+, -, ×, ÷, exp, sqrt, log 等 11 种常见数学算符, 以生成候选组合. 通过设置 nsf = 12 和 fcomplexity = 6, 在 12 个初始特征中筛选出 6 个最优组合, subs_sis = 10000 用于控制候选特征子集的搜索规模. 在稀疏回归阶段, 引入 ℓ_0 范数最小化方法 (method = 'L0') 作为稀疏化策略, 以在确保模型预测精度的同时, 构建形式简洁、物理意义明确的解析表达式. 在 VS-SISO 模型中进一步引入随机子集筛选与迭代变量选择机制, 其中超参数 Sa = 4, 最大迭代次数设为 200, 以提升特征选择效率与最终模型性能. 具体参数设定见表 2.

表 2 SISO 与 VS-SISO 参数设置
Table 2. Parameters for SISO and VS-SISO.

参数名称	设置值	说明
desc_dim	1	描述符维度
ptype	1	回归任务
opset	11	运算符: (+)(-)(*)(/)(exp)(exp) (^(-1))(^2)(sqrt)(log)()
ntask	1	任务数(无任务划分)
nsample	213	样本数
nsf	12	特征总数上限
fcomplexity	6	最大选中特征数
subs_sis	10000	SIS阶段筛选特征子集
method	"L0"	稀疏回归方法
metric	"RMSE"	评价指标
nm_output	100	最终模型输出数量
Sa	4	随机子集维数 (VS-SISO)
max_iter	200	最大迭代次数 (VS-SISO)

为全面评估所采用的 6 种模型 (包括 4 种机器学习模型: RF, SVR, KRR, EXT, 以及符号回归

模型: SISO 和 VS-SISO) 的预测性能, 本文选取平均绝对误差 (MAE)、均方根误差 (RMSE) 和决定系数 (R^2) 三项指标作为评价标准. 每个模型均采用十折交叉验证 (10-fold cross-validation) 进行评估, 并对各项指标取平均值, 分别记作 CV-MAE, CV-RMSE 和 CV- R^2 , 以衡量模型在不同数据划分下的泛化能力和稳定性.

3 结果与讨论

对比 RF, SVR, KRR 和 EXT 四种模型的预测性能, 如图 2 所示, 所有模型在训练集与测试集上均表现出良好的拟合趋势, 数据点大致分布在理想预测线 $y = x$ 的两侧. 4 种模型的决定系数 (R^2) 均超过 95%, 其中 SVR 表现最为优异, 在测试集上取得了最高的 R^2 (97.86%), 并在 MAE (3.54 GPa) 和 RMSE (7.35 GPa) 方面也优于其他模型.

在十折交叉验证 (CV) 中, SVR 同样展现出较强的泛化能力, 其 CV-MAE 和 CV-RMSE 分别为 3.08 GPa 和 8.37 GPa, CV- R^2 达到 96.86%, 优于其他 3 种模型. 综上, SVR 是本研究中预测性能最优的回归模型. 4 种模型在原始预测与交叉验证中的 MAE, RMSE 和 R^2 指标汇总于表 3 中.

表 3 各模型性能评估结果
Table 3. Evaluation results of different models.

	RF	SVR	KRR	EXT
MAE/GPa	3.79	3.54	4.89	4.27
RMSE/GPa	8.42	7.35	10.88	11.00
R^2 /%	97.20	97.86	95.31	95.22
CV-MAE/GPa	3.09	3.08	3.96	2.70
CV-RMSE/GPa	9.04	8.37	11.36	13.47
CV- R^2 /%	95.36	96.86	95.40	95.39

为了获得 SISO 模型所需的输入描述符, 对最佳 SVR 模型训练过程中所使用的特征进行了重要性排序分析. 图 3(a) 展示了特征重要性占比超过 1% 的 20 个主要描述符. 其中, 熔点的平均值 (meanMT) 具有最高的权重, 占整体特征重要性的 7.2%; 其次是熔点的众数值 (modeMT), 其权重占比为 4.6%. 这表明熔点在预测立方钙钛矿的体模量中起着关键作用.

除熔点外, 原子质量、高斯体积、共价半径以及电负性等特征也具有较高的重要性. 这些原子属

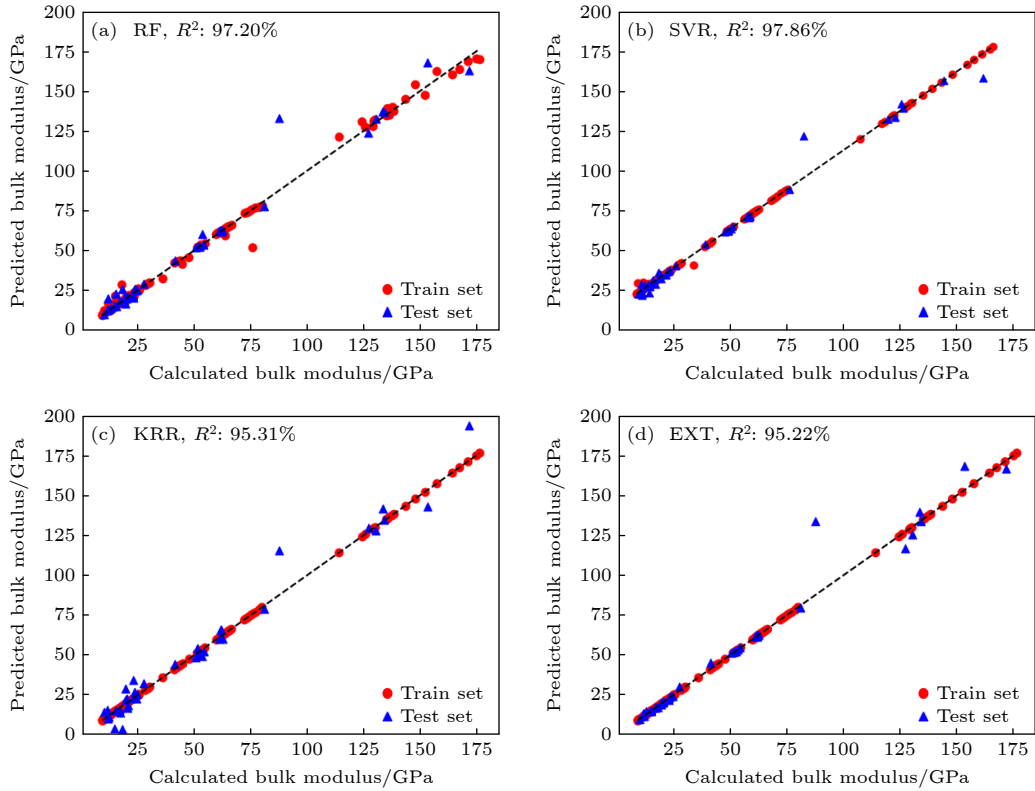


图 2 各机器学习模型预测值与真实值的对比图 (a) 随机森林; (b) 支持向量回归; (c) 核岭回归; (d) 极端随机树; 红色点和蓝色三角形分别表示训练集和测试集数据, 虚线表示理想预测线 $y = x$

Fig. 2. Comparison between predicted and calculated bulk modulus for different machine learning models: (a) Random forest (RF); (b) support vector regression (SVR); (c) kernel ridge regression (KRR); (d) extremely random trees (EXT); red dots and blue triangles correspond to the training and testing datasets, respectively, while the dashed line represents the ideal prediction ($y = x$).

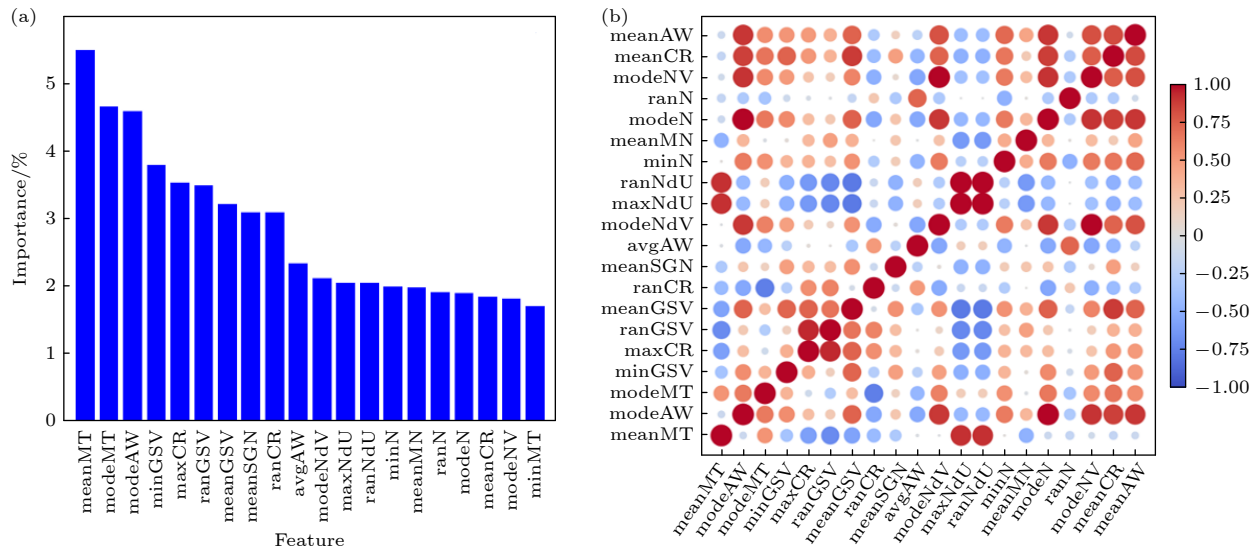


图 3 (a) 基于支持向量回归模型的特征重要性排序; (b) 特征之间的皮尔逊相关系数热图

Fig. 3. (a) Feature importance ranking based on the support vector regression (SVR) model; (b) Pearson correlation coefficients between features.

性与材料的力学性能密切相关: 熔点反映了原子间相互作用的强弱, 原子质量影响振动行为和声子传播, 原子半径和体积决定了原子的填充密度与堆积

方式, 而电负性则与键的极性及其键合强度有关. 因此, 这些特征共同决定了晶体结构在外力作用下的抵抗能力, 成为影响材料体模量的主要因素.

在完成特征重要性排序后,为进一步提升 SISSO 模型的稳健性与泛化能力,对初步筛选出的 20 个高重要性特征进行相关性分析.考虑到高度相关的特征可能导致信息冗余并引发多重共线性问题,采用皮尔逊相关系数^[41]对所有特征进行了两两线性相关性评估.如图 3(b)所示,颜色越深、球体越大表示特征间的相关程度越高.基于此,设定 85% 的相关性阈值,筛除高度冗余特征,最终筛选出 12 个具有代表性且相互独立性较高的主要特征.这些特征在保留关键信息的同时,有助于构建更加简洁且物理可解释性更强的 SISSO 模型,其具体名称与物理含义如表 4 所示.

表 4 SISSO 模型的输入特征及其含义

Table 4. Input features and their meanings in the SISO model.

特征名称	含义	特征名称	含义
meanMT	平均熔点	meanGSV	平均基态高斯体积
modeMT	众数熔点	meanSGN	平均空间群数
modeAW	众数原子质量	avgAW	原子质量的平均偏差
minGSV	最小基态晶胞体积	minN	最小原子序数
maxCR	最大共价半径	ranN	原子序数极差
ranCR	共价半径极差	meanMN	平均门捷列夫序数

为系统比较 SISSO 模型与 VS-SISSO 模型在体模量预测方面的性能,本文采用 RMSE 与 R^2 两项评估指标,对所构建的立方钛矿体模量预测公式进行对比分析.在 SISSO 模型中,前期通过特征选择筛选出与立方钛矿体模量最相关的一组原子属性变量,并据此构建特征空间.在此基础上,模型进一步提取出一维最优特征组合,得到如下预测表达式,以下简称热-结构耦合描述符:

$$B(\text{GPa}) = 1.09 \times \frac{(\max\text{CR} + \text{ranCR}) \times \text{meanMT}}{\text{meanGSV} \times (\max\text{CR} - \text{ranCR} + \text{minN})} - 2.54, \quad (5)$$

式中包含的特征包括最大共价半径 ($\max\text{CR}$)、共价半径极差 (ranCR)、平均熔点 (meanMT)、平均高斯体积 (meanGSV) 和最小原子序数 (minN).该描述符体现了原子半径、晶体热稳定性与体积致密度等因素对体模量的协同调控机制.其中,熔点越高表征键合能越强,有助于增强材料的抗压性能;共价半径则反映了化学键的空间尺度;而平均高斯体积越小,结构越致密,通常对应更高的体模

量.该趋势与 Cohen 等^[42]提出的经验公式 $B \propto d_{\text{B-X}}^{-n}$ ($n = 3-4$) 相一致,即晶体越紧凑、结合能越强,体模量越大,具有良好的物理可解释性.与 Guo 等^[43]提出的 $d_{\text{B-X}}^{-4}$ 描述符模型相比 ($\text{RMSE} = 13.62 \text{ GPa}$, $R^2 = 73.53\%$),本文构建的热-结构耦合描述符在预测精度方面表现更优, RMSE 降至 7.41 GPa , R^2 提升至 97.8% ,如图 4(a)–(c) 所示.此外,与 Verma 等^[44]提出的依赖晶格常数的半经验体模量公式 $\frac{(Z_{\text{A}}Z_{\text{B}}Z_{\text{X}})^{0.35}}{a^{3.5}}$ 不同,本文模型仅基于元素原子属性特征,无需结构优化参数,便于高通量材料筛选.

在 VS-SISSO 模型中,以初始的 138 项原子层级特征作为输入,考虑到模型迭代过程中采用随机子集抽样策略,为避免误差路径的波动,对模型进行 3 轮独立训练.图 4(a) 所示为 3 轮训练中获得的最小 RMSE 分别为 5.34 , 5.66 和 5.98 GPa ,均优于 SISO 所构建的热-结构耦合描述符.

最终获得的最佳描述公式如下,本文将其命名为电子-热-结构三重耦合描述符:

$$B(\text{GPa}) = 46.85 \times \frac{(\text{meanE})^2 \times |n_{\text{X}} + n_{\text{B}}|}{\min\text{CR} \times \sqrt{\frac{\max\text{NPU}}{\text{meanMT}}}} - 3.89, \quad (6)$$

式中, meanE 表示平均电负性, n_{X} 为阴离子 X 的原子价态, n_{B} 为阳离子 B 的原子价态;若为双钙钛矿结构,则 n_{B} 表示 B_1 与 B_2 原子的平均价态. $\min\text{CR}$ 表示原子间的最小共价半径, $\max\text{NPU}$ 为原子 p 轨道中未被填充的最大电子数,用于表征价层未配对电子情况, meanMT 表示原子平均熔点.对于离子型化合物而言,原子价态越高通常对应更大的马德隆能,从而增强晶体的静电约束力与结构刚性,进而提高体模量.如图 4(d) 所示,高价态的硫族钙钛矿材料体模量普遍高于价态较低的卤族钙钛矿.此外,电负性可作为衡量离子键强度的有效指标.在 ABX_3 钙钛矿系列中, A, B 元素相同的情况下,阴离子 X 的电负性越大,形成的 B–X 键越强,整体晶体的体模量也越高.例如, F 基体钙钛矿的体模量普遍高于 Cl 基、Br 基与 I 基钙钛矿,呈现出 $\text{F} > \text{Cl} > \text{Br} > \text{I}$ 的递减趋势.通过引入价电子、电负性以及原子价态等电荷相关特征, VS-SISSO 模型在体模量预测性能上实现了进一步提升.电子-热-结构三重耦合描述符在测试集上实现了 $\text{RMSE} = 5.34 \text{ GPa}$, $R^2 = 98.35\%$, 优于热-结构耦合描述符,如图 4(d) 所示.

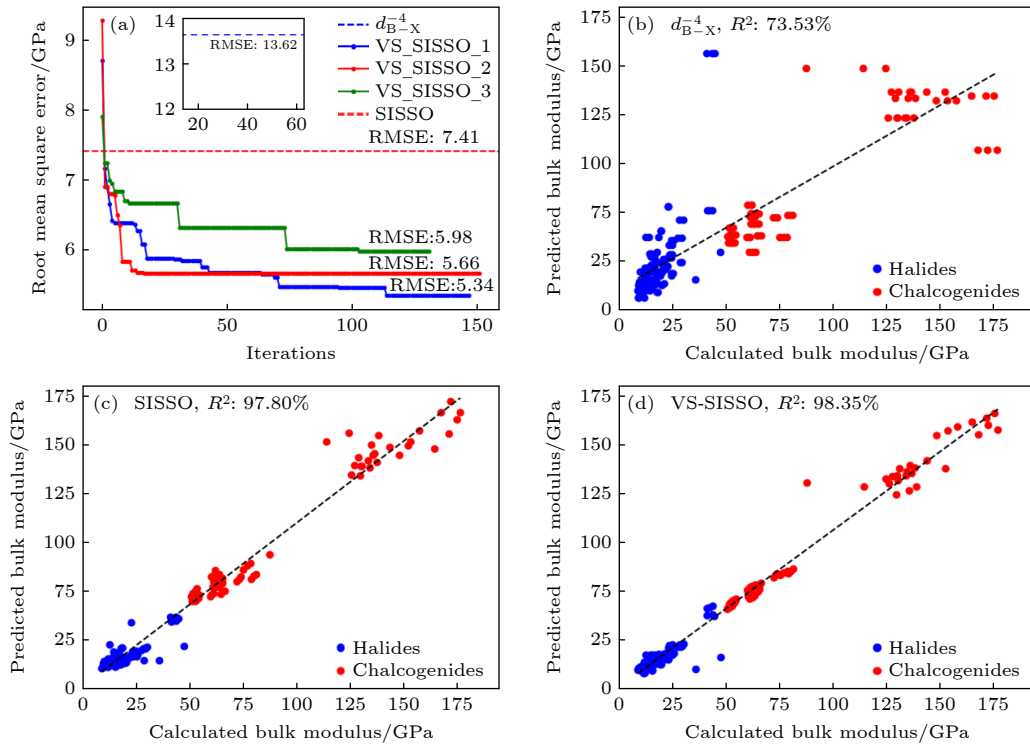


图4 (a) VS-SISSO (三轮独立训练) 与 d_{B-X}^{-4} , SISSO 所构建体模量描述符的预测 RMSE 对比; (b)–(d) 3 种不同模型的预测结果对比图, 其中 (b) 为使用 d_{B-X}^{-4} 描述符的拟合结果, (c) 为使用 SISSO 构建的热-结构耦合描述符的拟合结果, (d) 为使用 VS-SISSO 构建的电子-热-结构三重耦合描述符的拟合结果; 横轴表示 DFT 计算所得体模量, 纵轴为模型预测值; 蓝色点表示卤族钙钛矿, 红色点表示硫族钙钛矿

Fig. 4. (a) Comparison of prediction RMSE among the VS-SISSO (three independent trainings), the d_{B-X}^{-4} descriptor, and the SISSO-derived descriptor; (b)–(d) comparison of predicted vs. DFT-calculated bulk modulus using three different models, (b) the d_{B-X}^{-4} descriptor, (c) the SISSO-derived thermo-structural descriptor, (d) the VS-SISSO-derived electronic-thermo-structural descriptor; the x -axis represents the bulk modulus obtained from DFT, and the y -axis indicates model predictions; blue dots denote halide perovskites, while red dots represent chalcogenide perovskites.

为筛选出体模量与铅碘钙钛矿相近的候选材料, 本文基于构建的电子-热-结构三重耦合描述符, 对超过 10000 个立方钙钛矿的体模量进行了预测. 结果表明, 体模量分布在 10—20 GPa 区间的无机非铅钙钛矿约有 170 种, 具体列表见补充材料表 S4 (online) (本文所有数据集均以 Word 表格形式在补充材料给出). 这些材料在晶格柔性上与典型铅卤钙钛矿相近, 有望在缺陷容忍性方面展现相似潜力. 上述筛选结果为后续探索具备软晶格特征的稳定无铅钙钛矿及其缺陷物理特性提供了理论参考与数据支持.

4 结论

本文基于 SISSO 与 VS-SISSO 方法, 分别构建了热-结构耦合描述符模型和电子-热-结构三重耦合描述符模型, 用于立方钙钛矿体模量的可解释

性预测. 结果表明, 热-结构耦合描述符揭示了原子半径、熔点和体积等结构与热力学因素对体模量的调控作用; 进一步引入电负性、原子价态和未配对电子数等电子结构特征后, 电子-热-结构三重耦合模型在测试集上达到了 $RMSE = 5.34$ GPa, $R^2 = 98.35\%$ 预测性能, 显著优于传统模型. 基于该模型对超过 10000 个立方钙钛矿的体模量进行了高通量预测, 筛选出约 170 种体模量分布在 10—20 GPa 区间、具有类 Pb-I 软晶格特征的无铅钙钛矿体系, 为其潜在缺陷容忍性提供数据支持, 也为新型稳定、环保钙钛矿材料的快速发现与机制理解提供理论依据.

数据可用性声明

支撑本研究的数据集可在科学数据银行 <https://doi.org/10.57760/sciencedb.j00213.00161> 中访问获取.

参考文献

- [1] NREL. Best research-cell efficiency chart. <https://www.nrel.gov/pv/cell-efficiency.html> [2025-05-18]
- [2] Yin W, Shi T, Yan Y 2014 *Adv. Mater.* **26** 4653
- [3] Huang J, Yuan Y, Shao Y, Yan Y 2017 *Nat Rev Mater* **2** 17042
- [4] Xian Y M, Wang X M, Yan Y F 2024 *Chin. Phys. B* **33** 096803
- [5] Yin W J, Shi T, Yan Y 2014 *Appl. Phys. Lett.* **104** 063903
- [6] Ming C, Wang H, West D, Zhang S, Sun Y Y 2022 *J. Mater. Chem. A* **10** 3018
- [7] Miyata K, Meggiolaro D, Trinh M T, Joshi P P, Mosconi E, Jones S C, De Angelis F, Zhu X Y 2017 *Sci. Adv.* **3** e1701217
- [8] Zhu X Y, Podzorov V 2015 *J. Phys. Chem. Lett.* **6** 4758
- [9] Bonn M, Miyata K, Hendry E, Zhu X Y 2017 *ACS Energy Lett.* **2** 2555
- [10] Yang J F, Wen X M, Xia H Z, Sheng R, Ma Q S, Kim J, Tapping P, Harada T, Kee T W, Huang F Z, Cheng Y B, Green M, Ho-Baillie A, Huang S J, Shrestha S, Patterson R, Conibeer G 2017 *Nat. Commun.* **8** 14120
- [11] Chu W B, Zheng Q J, Prezhdo O V, Zhao J, Saidi W A 2020 *Sci. Adv.* **6** eaaw7453
- [12] Chu W B, Saidi W A, Zhao J, Prezhdo O V 2020 *Angew Chem Int Ed* **59** 6435
- [13] Wu X W, Ming C, Shi J, Wang H, West D, Zhang S B, Sun Y Y 2022 *Chin. Phys. Lett.* **39** 046101
- [14] Zhao X G, Yang J H, Fu Y, Yang D, Xu Q, Yu L, Wei S H, Zhang L 2017 *J. Am. Chem. Soc.* **139** 2630
- [15] Sun Q D, Wang J, Yin W J, Yan Y F 2018 *Adv. Mater.* **30** 1705901
- [16] Sun Q, Yin W J, Wei S H 2020 *J. Mater. Chem. C* **8** 12012
- [17] Ghorpade U V, Suryawanshi M P, Green M A, Wu T, Hao X, Ryan K M 2023 *Chem. Rev.* **123** 327
- [18] Yu Y, An Z D, Cai X Y, Guo M L, Jing C B, Li Y Q 2021 *Acta Phys. Sin.* **70** 048503 (in Chinese) [余毅, 安治东, 蔡晓艺, 郭明磊, 敬承斌, 李艳青 2021 物理学报 **70** 048503]
- [19] Dunn A, Wang Q, Ganose A, Dopp D, Jain A 2020 *npj Comput. Mater.* **6** 138
- [20] Geurts P, Ernst D, Wehenkel L 2006 *Mach Learn* **63** 3
- [21] Chen T, Guestrin C 2016 *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining San Francisco California USA* pp785–794
- [22] Hancock J T, Khoshgoftaar T M 2020 *J. Big Data* **7** 94
- [23] De Breuck P P, Hautier G, Rignanese G M 2021 *npj Comput. Mater.* **7** 83
- [24] Wang A Y T, Kauwe S K, Murdock R J, Sparks T D 2021 *npj Comput. Mater.* **7** 77
- [25] Xie T, Grossman J C 2018 *Phys. Rev. Lett.* **120** 145301
- [26] Schütt K T, Saucedo H E, Kindermans P J, Tkatchenko A, Müller K R 2018 *J. Chem. Phys.* **148** 241722
- [27] Chen C, Ye W, Zuo Y, Zheng C, Ong S P 2019 *Chem. Mater.* **31** 3564
- [28] Choudhary K, DeCost B 2021 *npj Comput. Mater.* **7** 185
- [29] Gasteiger J, Giri S, Margraf J T, Günnemann S 2020 *arXiv: 2206.13578 [cs. LG]*
- [30] Guerrero P, Hašan M, Sunkavalli K, Měch R, Boubekeur T, Mitra N J 2022 *ACM Trans. Graph.* **41** 1
- [31] Ruff R, Reiser P, Stühmer J, Friederich P 2024 *Digit. Discov.* **3** 594
- [32] Ong S P, Cholia S, Jain A, Brafman M, Gunter D, Ceder G, Persson K A 2015 *Comput. Mater. Sci.* **97** 209
- [33] Akinpelu S B, Abolade S A, Okafor E, Obada D O, Ukpong A M, Kumar R. S, Healy J, Akande A 2024 *Res. Phys.* **65** 107978
- [34] Ouyang R, Curtarolo S, Ahmetcik E, Scheffler M, Ghiringhelli L M 2018 *Phys. Rev. Mater.* **2** 083802
- [35] Guo Z, Hu S, Han Z K, Ouyang R 2022 *J. Chem. Theory Comput.* **18** 4945
- [36] Roy P B, Roy S B 2005 *J. Phys.: Condens. Matter* **17** 6193
- [37] Heyd J, Peralta J E, Scuseria G E, Martin R L 2005 *J. Chem. Phys.* **123** 174101
- [38] Kresse G, Furthmüller J 1996 *Phys. Rev. B* **54** 11169
- [39] Ward L, Dunn A, Faghaninia A, Zimmermann N E R, Bajaj S, Wang Q, Montoya J, Chen J, Bystrom K, Dylla M, Chard K, Asta M, Persson K A, Snyder G J, Foster I, Jain A 2018 *Comput. Mater. Sci.* **152** 60
- [40] Fabian P, Gaël V, Alexandre G, Vincent M, Bertrand T, Olivier G, Mathieu B, Peter P, Ron W, Vincent D, Jake V, Alexandre P, David C, Matthieu B, Matthieu P, Duchesnay É 2011 *J. Mach. Learn. Res.* **12** 2825
- [41] Benesty J, Chen J, Huang Y, Cohen I 2009 *Pearson Correlation Coefficient* (Berlin: Springer) pp1–4
- [42] Cohen M L 1993 *Science* **261** 307
- [43] Guo Z, Wang J, Yin W J 2022 *Energy Environ. Sci.* **15** 660
- [44] Verma A S, Kumar A 2012 *J. Alloy. Comp.* **541** 210

DATA PAPER

Descriptors for the interpretability of cubic perovskite bulk modulus oriented towards soft lattice screening*

 JIANG Jinming SUN Qingde[†] ZHANG Weibing[‡]

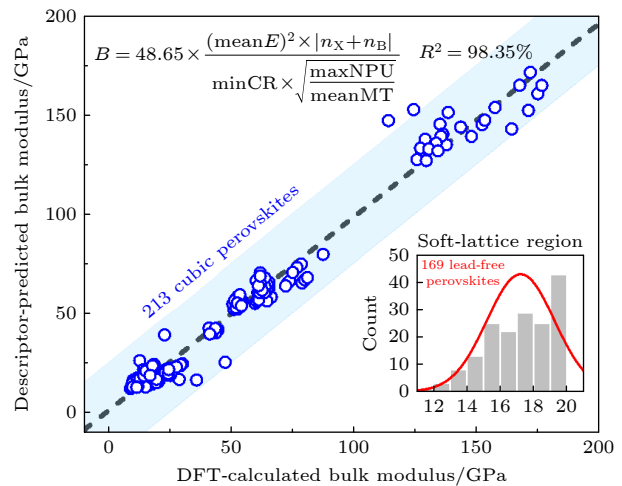
(Hunan Provincial Key Laboratory of Flexible Electronic Materials Genome Engineering, School of Physics & Electronic Science, Changsha University of Science and Technology, Changsha 410114, China)

(Received 18 May 2025; revised manuscript received 3 July 2025)

Abstract

In recent years, soft lattices have been considered a primary physical origin of defect tolerance in lead-halide perovskite materials, with bulk modulus serving as a key indicator of lattice “softness”. This work focuses on cubic perovskites and constructing a dataset of bulk moduli for 213 compounds based on density functional theory (DFT) calculations. A total of 138 features are compiled, including 132 statistical features extracted using the Matminer toolkit and 6 manually selected elemental descriptors. Four conventional machine learning regression models (RF, SVR, KRR, and EXR) are employed for prediction. Of them, the SVR model shows the best performance, achieving a test-set Root Mean Square Error (RMSE) of 7.35 GPa and Coefficient of Determination (R^2) of 97.86%. Feature importance analysis reveals that thermodynamic-structural features such as melting point, covalent radius, and atomic volume play dominant roles in determining bulk modulus. Based on the 12 most important features, a thermodynamic-structural coupling descriptor is constructed using the SISSO method, yielding a test-set RMSE of 7.41 GPa and R^2 of 97.80%. The resulting descriptor indicates that the bulk modulus is proportional to melting point and inversely proportional to atomic volume. Furthermore, the VS-SISSO method combined with a random subset selection and iterative variable screening strategy is used, enabling the selection of electronic-level features such as electronegativity, valence state, and number of unpaired electrons. The resulting electronic-thermodynamic-structural coupling descriptor further improves the prediction accuracy, reaching an RMSE of 5.34 GPa and R^2 of 98.35% on the test set. Notably, due to the difference in valence states, this model effectively distinguishes between the bulk moduli of chalcogen-based (divalent) and halogen-based (monovalent) perovskites.

Based on this model, high-throughput screening is performed on over 10000 cubic chalcogenides and halide perovskites, and approximately 170 lead-free candidates with bulk moduli in the range of 10–20 GPa are identified, which are comparable to Pb-I perovskites. These results provide preliminary evidence for supporting the applicability of the soft-lattice mechanism in lead-free systems and offer theoretical guidance and data support for the high-throughput discovery of stable, defect-tolerant, lead-free perovskite materials. All the data presented in this paper are openly available at <https://doi.org/10.57760/sciedb.j00213.00161>.



Keywords: bulk modulus, defect tolerance, soft lattice, perovskite, interpretable machine learning

PACS: 02.60.Pn, 71.15.Mb, 88.40.H-, 07.05.Mh

DOI: 10.7498/aps.74.20250652

CSTR: 32037.14.aps.74.20250652

* Project supported by the National Natural Science Foundation of China (Grant No. 12474219) and the Natural Science Foundation of Hunan Province, China (Grant No. 2023JJ40041).

[†] Corresponding author. E-mail: qingdesun@csust.edu.cn

[‡] Corresponding author. E-mail: zhangwb@csust.edu.cn

面向软晶格筛选的立方钙钛矿体模量可解释性描述符研究

姜锦铭 孙庆德 张卫兵

Descriptors for the interpretability of cubic perovskite bulk modulus oriented towards soft lattice screening

JIANG Jinming SUN Qingde ZHANG Weibing

引用信息 Citation: *Acta Physica Sinica*, 74, 170201 (2025) DOI: 10.7498/aps.74.20250652

CSTR: 32037.14.aps.74.20250652

在线阅读 View online: <https://doi.org/10.7498/aps.74.20250652>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

溶液法原位大面积制备钙钛矿光电薄膜成膜的同步辐射可视化结晶过程研究

Solution slot-die coating perovskite film crystalline growth observed by *in situ* GIWAXS/GISAXS

物理学报. 2024, 73(6): 063201 <https://doi.org/10.7498/aps.73.20231847>

基于SnO₂:DPEPO混合电子传输层的钙钛矿太阳能电池性能研究

Performance of perovskite solar cells based on SnO₂:DPEPO hybrid electron transport layer

物理学报. 2024, 73(19): 198401 <https://doi.org/10.7498/aps.73.20240827>

基于钙钛矿的二维铁磁体Sr₂RuO₄单层

Perovskite-based two-dimensional ferromagnet Sr₂RuO₄ monolayer

物理学报. 2024, 73(22): 226101 <https://doi.org/10.7498/aps.73.20241042>

光谱稳定的混合卤素蓝光钙钛矿LED设计

Compositional design of spectrally stable blue mixed-halide perovskite LEDs

物理学报. 2025, 74(14): 148502 <https://doi.org/10.7498/aps.74.20250297>

单个CsPbBr₃钙钛矿量子点的荧光闪烁特性

Photoluminescence blinking properties of single CsPbBr₃ perovskite quantum dots

物理学报. 2021, 70(20): 207802 <https://doi.org/10.7498/aps.70.20210908>

铅卤钙钛矿法布里-珀罗谐振腔激光器

Lead halide perovskites Fabry-Pérot resonant cavity laser

物理学报. 2021, 70(5): 054205 <https://doi.org/10.7498/aps.70.20201302>