

数据驱动的高熵碲化物组分高通量理论设计与性能预测

李 顺¹, 彭浩然², 符秀丽³

(1. 交通运输部天津水运工程科学研究所, 天津 300456; 2. 中国地质大学 数理学院, 北京 100080;

3. 北京邮电大学 集成电路学院, 北京 100080)

摘要:针对高熵碲化物热电材料组分空间庞大、传统“试错法”研发模式效率低的难题,从理论层面建立一套结合机器学习与第一性原理计算的数据驱动设计框架。通过构建包含混合焓、原子尺寸差异等物理特征的高维描述符空间,训练梯度提升回归模型,实现对多主元体系形成能的精准预测。采用沙普利加性解释(SHapley additive explanations, SHAP)方法,揭示热力学参数与几何参数间的竞争机制,确定定量的相形成判据。利用该判据对PbTe-SnTe-GeTe三元体系进行全组分高通量筛选,成功获得一种兼具高热力学稳定性(-46.31 kJ/mol)与超低晶格热导率(0.85 W/mK)的高熵组分。电子结构计算结果显示,引入过渡金属Mn,诱导了导带底能带收敛与p-d轨道强杂化,有效解耦了电、热输运参数。理论评估结果显示,该高熵组分在850 K时的热电优值达1.60,较PbTe基体提升67%。初步验证了数据驱动策略在理论设计中的可行性,为后续的实验合成与验证提供了明确的组分指引。

关键词:高熵碲化物;机器学习;热电材料;相稳定性;第一性原理计算

分类号:(中图)TB34;TG139

文献标志码:A

随着全球工业化进程加快与化石能源枯竭危机加剧,开发清洁、高效的可再生能源转换技术,已成为国际科技竞争的焦点。热电转换技术利用半导体材料内部载流子的定向运动,通过塞贝克效应实现热能与电能的直接相互转换,在工业余热回收、深空探测、放射性同位素电源及微电子器件精确温控等领域,展现出不可替代的重要战略价值。热电材料的能量转换效率,主要取决于无量纲热电优值 zT , $zT = S^2 \sigma T / (\kappa_e + \kappa_l)$,其中, S , σ , T , κ_e , κ_l 分别代表塞贝克系数、电导率、绝对温度、电子热导率和晶格热导率^[1]。长期以来,提升热电材料 zT 值的核心瓶颈在于电、热输运参数之间存在较强的内禀耦合关系,其“此消彼长”的制约使单一主元材料的性能优化逐渐逼近理论极限^[2]。近年来,高熵合金(high-entropy alloys, HEAs)概念的引入,为打破这一瓶颈提供了全新契机。高熵合金材料通过向单一晶格位点引入多种(通常大于等于5种)主元元素,提

高材料体系的构型熵,进而降低其吉布斯自由能,稳定单相结构;同时,利用晶格畸变效应大幅降低 κ_l ,这完美契合“声子玻璃-电子晶体”的设计理念^[3]。

针对高熵碲化物热电材料的设计与性能优化,国内外学者已开展了大量探索性工作并取得显著进展。早在2016年,Fan等^[4]率先合成等原子的PbSnTeSe四元高熵合金,并发现严重的晶格畸变可导致声子散射现象显著增强,使其室温 κ_l 低至0.6 W/(mK),证实高熵策略在降低 κ_l 方面的巨大潜力。随后,Jiang等^[5]利用熵驱动的结构稳定化效应,成功克服高熵PbSe基体中S和Te固溶度的限制,制备出单相n型(Pb,Sn)(Se,Te,S)高熵材料。研究结果显示,高熵效应不仅抑制多相分离,还诱导异常的剪切应变,使材料在900 K时的 zT 高达1.8。Jiang等^[6]向GeTe基体系引入Ag、Sb、Pb等元素构建高熵体系,发现高构型熵能有效将GeTe的菱方-立方相变温度降低至室温附近,从而稳定其

收稿日期:2026-01-10

基金项目:国家自然科学基金资助项目(U21A6004;12174035);信息光子学与光通信全国重点实验室(北京邮电大学)开放基金资助项目

作者简介:李顺(1997—),助理研究员,博士,主要从事高性能热电转换材料与器件研究,(电子信箱)lis@tiwte.ac.cn。

引用格式:李顺,彭浩然,符秀丽.数据驱动的高熵碲化物组分高通量理论设计与性能预测[J].宁夏大学学报(自然科学版中英文),XXXX,XX(X):1-10.

高对称性立方相结构;通过能带收敛与声子散射的协同作用,实现了 zT 值达2.7的历史性突破。针对SnTe体系,Zhang等^[7]通过中熵工程策略,在Pb、Ge、Sb、Cd共合金化的SnTe基材料中实现了电、热运输性能解耦,并获得优异的功率因子;而Su等^[8]则开发了超重力场辅助的自蔓延高温合成(self-propagating high-temperature synthesis, SHS)技术,实现了SnTe基高熵材料的快速制备与微结构调控。此外,Li等^[9]提出利用高熵效应诱导四能带收敛的新策略,并在AgSbPbSnGeTe₅体系中验证了多主元元素对电子能带结构的复杂调控作用。尽管上述研究展示了高熵碲化物的巨大应用潜力,但面对由Pb、Sn、Ge、Mn、Cd等多种元素构成的庞大组分空间(超过百万种组合),传统的“试错法”实验效率低下,且难以精准预测多主元体系中复杂的相竞争关系与亚稳态行为。

为了克服传统实验方法的局限性,文中聚焦Pb、Sn、Ge、Mn、Cd多主元碲化物体系,旨在构建一套基于数据驱动的高熵热电材料设计框架。该体系中各元素虽同属或兼容岩盐结构,但原子半径与化学键能差异导致固溶度范围错综复杂,极易发生分相。文中结合高通量密度泛函理论(DFT)计算与机器学习算法,构建包含混合焓、原子尺寸差异、价电子浓度等物理参量的高维特征描述符空间,并训练梯度提升回归(gradient boosting regression, GBR)等预测模型。在此基础上,深入解析熵驱动的相稳定机理,明确定量的相形成判据;在全组分空间内快速筛选出兼具热力学稳定性与低晶格热导率的最优组分。这不仅有望发现新型高性能高熵碲化物,也为复杂多主元功能材料的理性设计提供了参考。

1.1 数据集的构建与特征描述符

数据驱动研究的基础在于高质量、高维度的材料数据库。文中数据主要来源于Materials Project数据库、ICSD数据库以及部分高通量计算文献。为了研究高熵碲化物,筛选包含Pb、Sn、Ge、Mn、Cd等元素的IV-VI族化合物及其固溶体数据,构建初始数据集。数据预处理阶段,包括去除重复条目、剔除含有未收敛计算结果的样本以及对化学式进行标准化处理。最终,筛选出1200多组具有确定晶体结构和形成能的样本并用于模型训练。在最终的数据集中,二元体系的占比约为15%,三元体系的占比约为45%,四元及以上高熵、中熵复杂体系的占比约为40%。其中,约65%的数据来源于已报道的稳定相结构,剩余35%数据为基于第一性原理高通量计算生成的亚稳相或同分异构虚拟结

构。这种涵盖丰富组分维度与能量态的数据分布,有效保障了机器学习模型对未知高熵新体系特征空间的探索能力与泛化性能。

为了准确分析高熵合金的复杂构效关系,文中引入3类物理特征描述符。

1) 热力学参数。包括混合焓(ΔH_{mix})、混合熵(ΔS_{mix})以及表征二者竞争关系的 Ω 参数, Ω 的计算公式为

$$\Omega = T_m \Delta S_{\text{mix}} / |\Delta H_{\text{mix}}|,$$

式中: T_m 为平均熔点。

2) 几何参数。主要由原子尺寸差异(δ)表征,其计算公式为

$$\delta = \sqrt{\sum_{i=1}^n c_i (1 - r_i / \bar{r})^2},$$

式中: c_i 、 r_i 分别为第*i*种组元的摩尔分数和原子半径; \bar{r} 为所有组元原子半径的加权平均值。

3) 电子参数。包括价电子浓度(c_{ve})和鲍林电负性差异($\Delta\chi$),这两个参数能够反映化学键的离子性与共价性特征以及电子态密度的稳定性。

1.2 机器学习工作流程

采用Python环境下的Scikit-learn机器学习库构建机器学习工作流^[10]。为了建立从特征空间到目标性质(形成能(E_f)、 κ_1)的非线性映射关系,测试多种回归算法,包括梯度提升回归(gradient boosting regression, GBR)、随机森林(random forest, RF)、支持向量回归(support vector regression, SVR)等。

模型训练以80%的数据为训练集,以20%的数据为独立测试集。为了优化模型超参数(如决策树深度、学习率等),在训练集中采用基于5折交叉验证(5-fold cross-validation)的网格搜索策略。模型性能通过均方根误差(e_{rms})、平均绝对误差(e_{ma})和决定系数(R^2)进行综合评估。此外,为了解决“黑箱”模型的可解释性问题,引入SHAP方法,通过计算每个特征对预测结果的边际贡献,量化特征重要性,从而揭示影响相稳定性的物理本质^[11]。

1.3 第一性原理计算细节

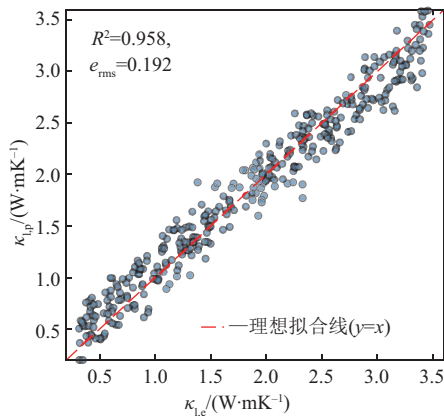
针对筛选出的最优高熵组分,采用基于密度泛函理论(density functional theory, DFT)的VASP软件包(Vienna Ab initio simulation package)进行精细电子结构计算^[12]。离子与价电子之间的相互作用采用投影缀加平面波(projector augmented-wave, PAW)方法描述,交换关联泛函采用广义梯度近似下的PBE(Perdew-Burke-Ernzerhof)形式^[13]。

为了准确描述碲化物体系中重元素(如 Pb、Te)的相对论效应,所有电子结构计算(能带、态密度)均考虑自旋-轨道耦合(spin-orbit coupling, SOC)效应^[14]。平面波截断能设定为 500 eV,布里渊区积分采用 Monkhorst-Pack 方法生成 $8 \times 8 \times 8$ 的 k 点网格进行数值采样,能量收敛标准设定为 10^{-6} eV,原子受力收敛标准设定为 0.01 eV/Å。为了模拟高熵合金的无序结构,采用特殊准随机结构(special quasirandom structure, SQS)方法构建超胞模型^[15]。

1.4 热电输运性能的理论预测

基于 DFT 计算得到的能带结构,利用 BoltzTraP2 软件包,在刚带近似和恒定弛豫时间($\tau = 10^{-14}$ s)近似条件下,采用半经典玻尔兹曼输运方程(Boltzmann transport equation, BTE)方法,计算材料的塞贝克系数(s)、电导率(σ)和功率因子(P , $P = s^2\sigma$)。对于晶格热导率(κ_l)的预测,考虑到高熵合金中原子质量场波动和应力场散射现象,传统的 Callaway 模型不适用。因此,文中采用机器学习模型直接预测 κ_l 。该模型基于大量文献报道的实验数据训练,能够有效捕捉成分复杂度和晶格畸变对声子散射的非线性影响。最终 zT 的计算公式为

$$zT = S^2 \sigma T / (\kappa_e + \kappa_l)。$$



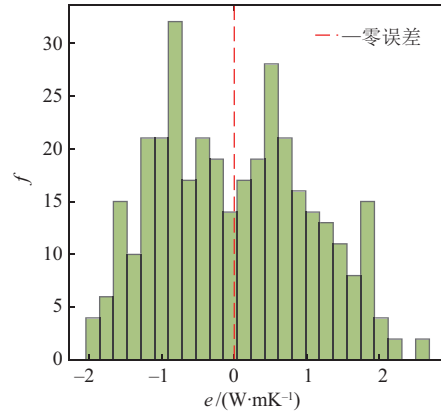
(a) 预测值与实验值的对比散点图

式中, κ_e 基于 Wiedemann-Franz 定律 ($\kappa_e = L\sigma T$) 估算。

1.5 晶格热导率机器学习模型细节

鉴于高熵碲化物中存在较强的声子散射现象,传统的理论模型难以准确评估其 κ_l ,文中构建专门 RF 回归模型用于 κ_l 预测。训练数据集由来自文献及材料数据库中 450 个多主元热电材料的 κ_l 数据(实验测试得到)构成,温度为 300~900 K。模型选用的核心特征包括平均原子质量、原子质量方差、平均原子半径、电负性差异及测试温度。模型超参数通过网格搜索方法优化。

为了直观评估 κ_l 预测模型的可靠性,对其在交叉验证集上的性能进行可视化分析(图 1)。对比模型预测值($\kappa_{l,p}$)与实验值($\kappa_{l,e}$),发现数据点紧密分布在 $y=x$ 对角线周围,表明模型具有较高的预测精度(图 1(a))。预测残差(e)近似服从均值为零的正态分布(图 1(b)),说明模型 e 主要由随机噪声引起,未出现明显的系统性偏差。尽管如此,该模型的预测能力受限于训练数据的范围和分布,对于极高熵(如主元元素数目大于 6)或全新元素组合体系,预测外推需谨慎,因为其置信区间会扩大。受限于训练数据的分布范围,对于主元元素数目大、组合新颖的全新高熵体系,模型的预测仍存在一定的不确定性,具体数值有待进一步实验验证。



(b) 5折交叉验证的残差分布

图 1 晶格热导率机器学习预测模型的性能评估

Fig. 1 Performance evaluation of machine learning models for lattice thermal conductivity prediction

2 结果与讨论

2.1 机器学习模型的构建与相稳定性判据的确立

构建能够准确描述高熵碲化物复杂物理化学性质的高维特征空间,是实现其相稳定性精准预测的先决条件。然而,在多维特征空间中,特征之间往往存在复杂的耦合关系,这种多重共线性不仅降

低了模型的可解释性,还可能导致过拟合风险。因此,首先采用皮尔逊相关系数矩阵对候选特征进行严格统计学检验。如图 2(a)所示,皮尔逊相关性分析定量验证了特征间存在的内禀物理联系: δ 与 $\Delta\chi$ 的相关系数高达 0.80。这与元素周期律中原子半径与电负性的协同变化规律相符。在机器学习建模中,这种较强的线性相关性意味着两者包含重叠

的信息。鉴于 δ 能更直观地反映高熵合金中的晶格畸变效应,且与 $\Delta\chi$ 高度共线,模型在特征选择时需警惕多重共线性带来的冗余风险。从晶体化学角度分析,这一统计结果具有明确的物理意义:在多主元碲化物固溶体中,离子半径失配引起的晶格畸变场,往往伴随着局域化学键离子性的协同波动,这种“几何-电子”参数的强耦合关系,表明它们在决定晶格稳定性时并非独立作用。根据统计学相关强度判断标准^[16],皮尔逊系数(r)绝对值 $|r| > 0.8$ 被视为极强相关,表明特征间存在显著的信息冗余;而 $|r| < 0.5$ 则被视为弱相关,意味着特征间的正交性较好,适合共同作为独立变量输入模型。 ΔH_{mix} 与其他参数之间的 r 普遍低于0.5,显示 ΔH_{mix} 作为独立热力学描述符是有效的,适合作为构建预测模型的核心变量。

为进一步分析“黑箱”模型内部的决策逻辑,量化各物理参数对相稳定性预测的真实贡献,文中引

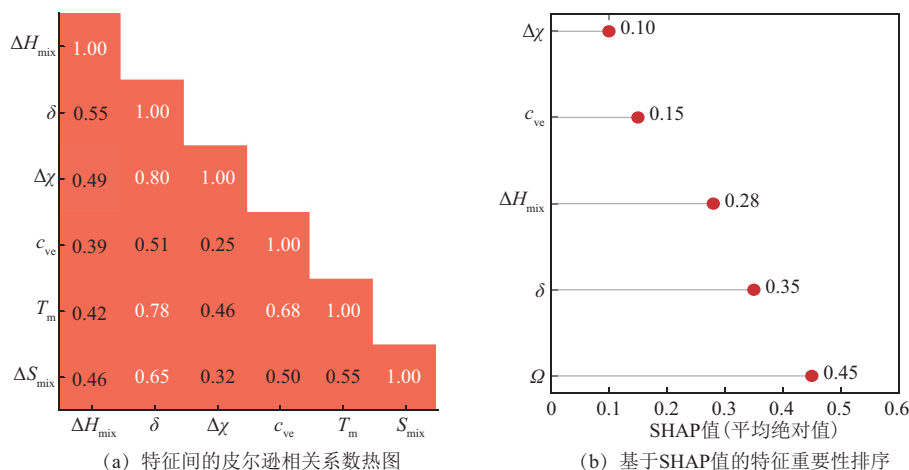


图2 特征相关性分析与关键描述符筛选

Fig. 2 Feature correlation analysis and key descriptor screening

在确立了关键物理描述符基础上,选择具有强鲁棒性和泛化能力的回归算法,是构建高精度模型的关键环节。选取5种代表性监督学习算法,包括线性回归、核方法、集成学习及神经网络等不同建模策略,并在独立测试集上进行系统性能评估(表1)。

由表1可知,不同算法在处理材料数据集时表现出显著的性能差异。基于集成学习策略的GBR、RF模型,展现出优异的非线性映射能力。其中,GBR模型通过沿着梯度下降方向迭代优化残差的Boosting策略,能够有效聚焦难以预测的样本,最终实现最低的 e_{rms} (5.210 kJ/mol)和最高的 R^2 (0.982)。相比之下,SVR、KRR模型由于受限于核函数的选择,难以充分捕捉多组元间复杂的高阶交

入基于博弈论的SHAP方法。与传统的特征重要性评分不同,该方法能够有效排除特征间相互作用所引发的干扰,为各特征提供更公平的边际贡献度量。如图2(b)所示,表征熵-焓竞争机制的热力学参数 $\Omega(T_{\text{m}}\Delta S, |\Delta H_{\text{mix}}|)$ 具有最高的全局重要性(Mean|SHAP|=0.45),紧随其后的是 δ (0.35)。需要说明的是,|SHAP|的大小依赖于目标变量的量纲,其物理意义在于相对重要性排序。 Ω 与 δ 构成了特征重要性的“第一梯队”,两者的累积贡献率占模型决策权重的60%以上,且与第三位特征(ΔH_{mix} , 0.28)及后续特征(如 $\Delta\chi$, 0.10)之间的差距显著增大。这种断崖式的分布特征,为确立相稳定性判据提供了定量依据。这一量化结果为经典高熵合金相形成理论提供了强有力的数据支撑,即体系能否形成单相固溶体,本质上取决于构型熵($T\Delta S$)产生的吉布斯自由能降低效应是否足以克服由原子尺寸失配引起的弹性应变能($\langle\Delta H_{\text{elastic}}\rangle$)。

表1 不同机器学习算法在测试集上的性能指标对比

Tab. 1 Performance comparison of different machine learning algorithms on the test dataset

算法模型	$e_{\text{rms}}/$ (kJ·mol ⁻¹)	$e_{\text{ma}}/$ (kJ·mol ⁻¹)	R^2
梯度提升回归(GBR)	5.210	3.956	0.982
随机森林(RF)	5.982	4.631	0.975
支持向量回归(SVR)	8.587	6.947	0.941
核岭回归(KRR)	10.131	8.201	0.912
人工神经网络(ANN)	7.526	6.079	0.955

互作用;而ANN模型虽然理论上具备通用的函数逼近能力,但在文中受限的样本规模下,容易导致局部极小值或出现过拟合现象。综合考虑模型的

准确性、鲁棒性与计算效率,最终选用GBR算法构建形成能预测模型。

GBR模型在全数据集上的预测性能与统计学特性见图3。由图3可知,形成能预测值($E_{f,p}$)与DFT计算的真实值($E_{f,e}$),在宽能量范围内高度收敛于 $y=x$ 对角线,表明模型不仅在低能稳态区预测准确,在高分非稳态区同样表现出较好的泛化能力;残差(e)分析结果显示,预测误差严格服从均值为零的高斯分布,且峰度与偏度接近理想正态分布,说明模型成功捕捉到数据中主要物理规律,且

剩余误差主要来源于随机噪声而非系统性偏差;学习曲线进一步证实,随着训练样本量增加至500个,交叉验证误差呈现指数级衰减并趋于收敛,且与训练误差的差距显著缩小,表明模型已达到良好的偏差-方差平衡,未出现欠拟合或过拟合现象。此外,针对形成能数值的特征重要性分析(图3(d))揭示了一个有趣的物理现象:虽然 Ω 决定了相的“定性”分类(稳定与否),但 c_{ve} 却是决定具体“定量”形成能的关键因素,这暗示在化学键层面,电子结构参数决定了成键强弱与晶格结合能大小。

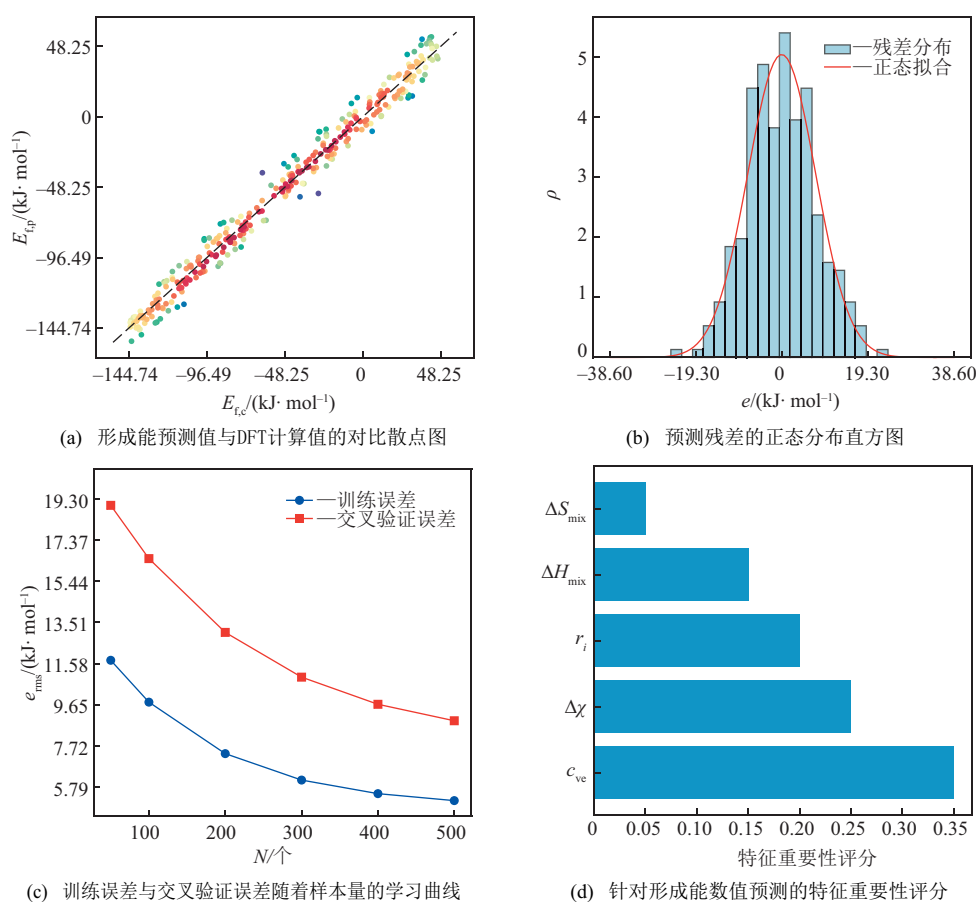


图3 GBR模型的性能评估与特征分析

Fig. 3 Performance evaluation and feature analysis of the GBR model

为了将上述高维“黑箱”模型的预测能力转化为可直接指导实验的物理直觉,将预测结果映射至二维物理参数空间并建立可视化的相稳定性判据图谱(图4)。由图4可知,单相固溶体高度集中在低晶格畸变($\delta < 6\%$)和中等负混合焓($\Delta H_{mix} \approx -15 \sim 0$ kJ/mol)区域,这不仅定量验证了Hume-Rothery规则在多主元碲化物体系中的适用性,也清晰地限定了形成稳定固溶体的热力学窗口。根据 Ω - δ 关系确立定量判据,当 $\Omega > 1.1$ 且 $\delta < 6.6\%$ 时,体系形成单相结构的概率接近100%。该判据综合考虑了熵稳定效应与晶格应变能的平衡,为高熵碲化

物的成分设计提供了明确的半经验公式,能够有效剔除90%以上的不稳定组分,从而显著降低试错成本。此外,图4(c)展示了 c_{ve} 与 δ 的相关性,稳定单相固溶体主要集中在 $c_{ve} = 9.5 \sim 10.5$,这与此类IV-VI族碲化物中典型的金属-共价键结构特征相符,进一步证实 c_{ve} 对晶格结合能的调控作用。图4(d)则揭示了 $\Delta\chi$ 与 δ 的协同分布情况,稳定相多出现在 $\Delta\chi$ 较低的区间,表明相近的电负性有助于形成均一的化学键环境,从而抑制电荷局域化引起的相分离趋势。

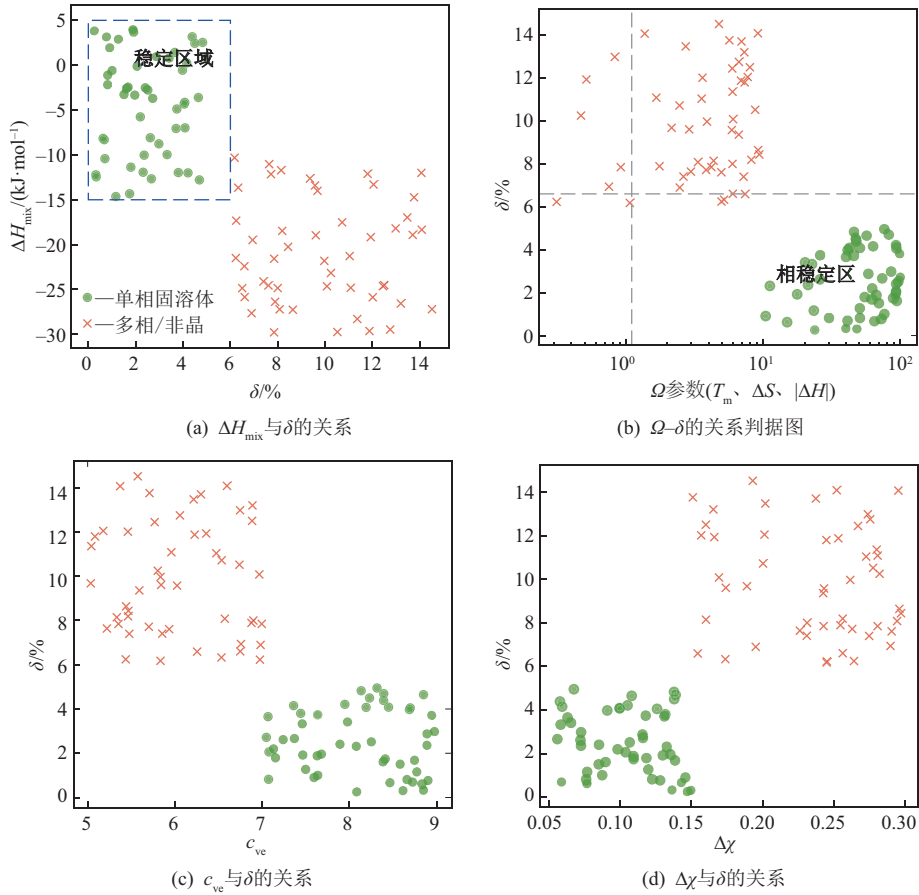


图4 数据驱动的相稳定性判据图谱

Fig. 4 Data-driven phase stability criterion map

2.2 高熵组分的高通量筛选与理论设计

基于上述构建的相稳定性判据与预测模型,以在热电领域具有广阔应用前景的PbTe-SnTe-GeTe三元体系为目标对象,实施全组分空间高通量筛选。该体系涉及IV-VI族半导体的多种结构竞争,且三元相图中各端点化合物性质差异显著。采用数据驱动方法,在体系中寻找兼具热力学稳定性与低晶格热导率的“帕累托最优前沿”(Pareto optimal frontier)是本研究的核心目标。

机器学习模型预测的形成能($E_{\text{f,p}}$)等高线图见

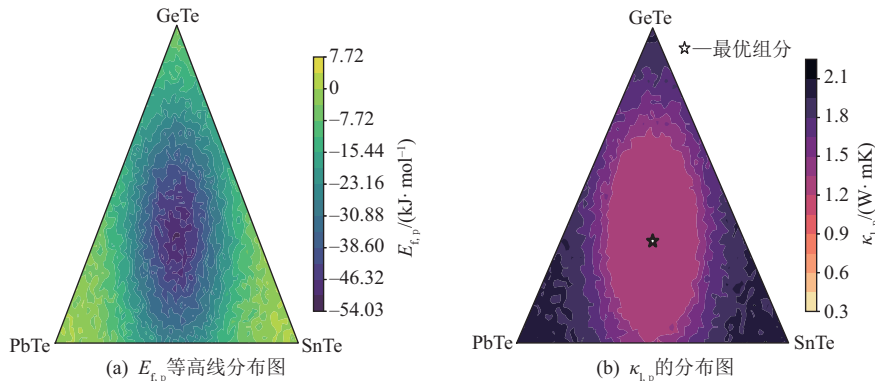


图5 PbTe-SnTe-GeTe三元体系的高通量筛选结果

Fig. 5 High-throughput screening results of the PbTe-SnTe-GeTe ternary system

图5(a)。热力学稳定的低能区($E_{\text{f,p}} < -38.95$ kJ/mol)呈现明显的拓扑特征,主要分布在三元相图的中心及偏向PbTe-SnTe二元边界的一侧。这表明在该区域内,多组分掺杂引发的高构型熵(S_{mix})有效降低了体系的吉布斯自由能,足以抵消由于引入小尺寸Ge原子($r_{\text{Ge}} < r_{\text{Sn}} < r_{\text{Pb}}$)带来的局域弹性畸变能,进而从理论上预测该区域能够抑制体系分相并维持稳定的单相固溶体结构。同时, $\kappa_{\text{l,p}}$ 的预测结果显示(图5(b)),其极小值区域集中在等物质的量比组分对应的中心位置。这是因为在中心区域,组分

复杂度最大,导致原子质量场与应力场发生剧烈空间波动,而这种强烈的非谐性波动大大增强了声子散射作用,从而显著降低 $\kappa_{1,p}$,使其趋近于非晶热导率极限值。

综合考虑相稳定性(低 E_f)与热输运性能(低 κ_1),在图5中标记兼具两者优势的“最优组分”。为定量评估该组分的优势,提取模型对其关键物理参数的预测值,并与纯二元基体进行详细对比(表2)。

表2 最优高熵组分与基体材料的理论物理性质对比

Tab. 2 Calculated crystallographic and thermodynamic properties of the designed high-entropy telluride compared with binary base compounds

组分	晶体结构	$a_{\text{calc}}/\text{\AA}$	$E_{f,p}/(\text{kJ}\cdot\text{mol}^{-1})$	$\kappa_{1,p}/(\text{W}\cdot\text{mK}^{-1})$
PbTe(基体)	$Fm\bar{3}m$	6.462	-33.77	2.10
SnTe	$Fm\bar{3}m$	6.327	-36.66	1.80
GeTe	$R\bar{3}m$	5.986	-27.02	2.40
HE-Te	$Fm\bar{3}m$	6.415	-46.31	0.85

由表2可知,高熵碲化物(HE-Te)体系的理论晶格常数(a_{calc})为6.415 Å,介于各组元 a_{calc} 之间,且符合 Vegard 定律所描述的线性叠加规则。这从晶体学几何角度验证了,该体系在形成单相面心立方结构($Fm\bar{3}m$)时的空间相容性,即高熵效应成功抑制 GeTe 发生特有的菱方畸变趋势,进而维持其高对称性立方晶型结构。更重要的是,模型预测 HE-Te 体系的 $E_{f,p}$ 低至-46.31 kJ/mol,显著低于各二元端点化合物(如 PbTe 体系的 $E_{f,p}=-33.77$ kJ/mol),显示熵稳定效应确实显著增强了体系的热力学稳定性,使其在高温下不易分解。但考虑到该高

熵体系可能偏离训练数据分布,出现这种反常的低 $E_{f,p}$ (-46.31 kJ/mol),不能排除由模型外推导致的预测误差,且实际热力学参数及相稳定性仍需通过热学实验进一步标定。同时,HE-Te 体系的 $\kappa_{1,p}$ 仅为0.85 W/mK,远低于 PbTe 基体($\kappa_{1,p}=2.10$ W/mK),降幅约60%,证明通过高熵化策略引入原子尺度无序性,是降低 κ_1 的有效途径。基于上述理论预测,HE-Te 体系展现出理想热电材料所需的“声子玻璃”特征,被确定为后续电子结构计算与性能评估的理想模型。

2.3 电子结构调制机理与热电性能理论评估

为揭示高熵化策略对高熵体系电输运性能的微观影响机理,阐明其打破电导率(σ)与赛贝克系数(s)间强耦合关系的原因,文中基于 DFT,对筛选出的最优组分 HE-Te 体系进行精细电子结构计算。HE-Te 体系在费米能级附近的能带色散关系见图6(a)。计算结果显示,随着过渡金属元素 Mn 的引入,导带底发生显著能带重整。与 PbTe 基体(较为简单的单带结构)不同,HE-Te 体系表现出明显的能带简并与收敛特征。这种多重能带在能量空间上的简并,有效增加了能带的态密度有效质量(m_{DOS}^*)。根据 Mott 关系, s 与 m_{DOS}^* 呈正相关($s \propto m_{\text{DOS}}^*$),这将直接导致热电势显著提升。同时,价带顶主要由 Te p 轨道贡献,其能带色散关系保持较高的曲率(较小的惯性有效质量 m_l^*),意味着空穴载流子在获得高 s 的同时,仍能维持较高的迁移率,从而实现了电输运参数的优化解耦。

态密度(DOS)分析进一步从轨道相互作用角度证实了上述机理(图6(b))。在费米能级附近,Te p 轨道与 Mn d 轨道发生强烈的 p-d 轨道杂化,形成显著共振态。这种共振效应不仅优化了费米面附近的电子态分布,引发 DOS 局域化峰值,还可

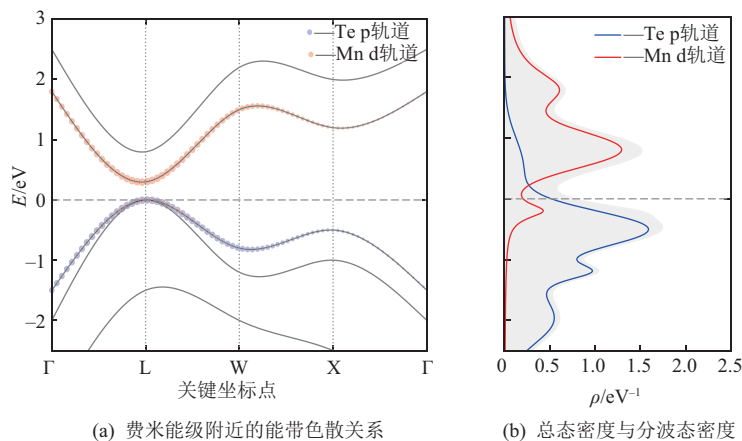


图6 最优高熵组分 HE-Te 体系的电子结构计算结果

Fig. 6 Electronic structure calculation results of the optimal high-entropy composition HE-Te

能通过能量依赖的散射机制优先过滤低能载流子,从而在不显著降低 σ 的前提下大幅度增大热电势。这种“能带收敛”与“轨道杂化”的协同效应,从量子力学层面解释了高熵体系在引入大量合金化原子并导致平移对称性破缺情况下,仍能保持优异电输运特性的原因,完美诠释了“电子晶体”设计理念。

基于计算得到的电子结构参数与声子热导率预测值,结合玻尔兹曼输运方程,对HE-Te体系在300~850 K的热电性能进行系统评估(图7,图中红色阴影区域代表考虑计算误差及不同载流子散射模型修正后的置信区间)。预测功率因子($P_{f,p}$)随着温度变化见图7(a)。得益于能带收敛带来的 s 增大,高熵组分HE-Te体系在全温区内的 $P_{f,p}$ 均显著优于PbTe基体。特别是在高温区域,由于能带简

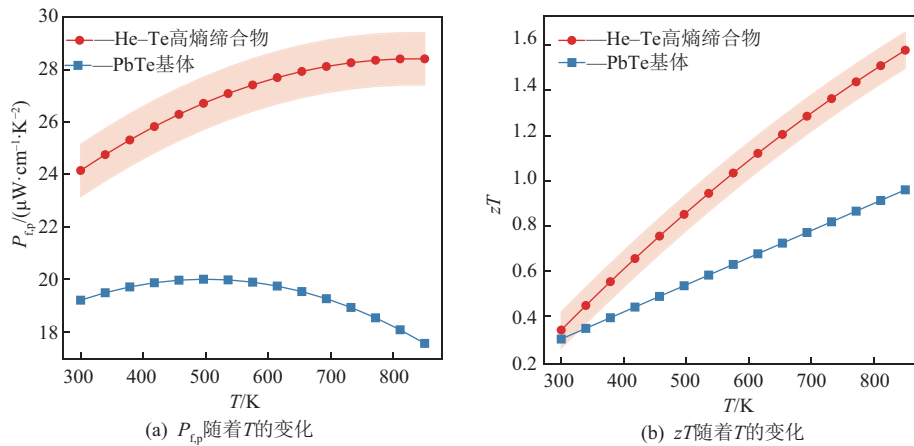


图7 全温区热电性能的理论预测对比

Fig. 7 Theoretical prediction comparison of full-temperature-range thermoelectric properties

3 结论

针对高熵碲化物组分的设计空间巨大、传统试错研发模式效率低的难题,提出基于机器学习与第一性原理计算的数据驱动设计框架。通过构建高维特征空间、训练高精度预测模型及深入解析微观电子结构,实现从相稳定性预测到高性能组分筛选的全流程设计;详细探讨特征关联性、相形成判据、组分筛选策略及电热输运机理。

1) 相稳定性预测模型的构建与判据的确立。通过皮尔逊相关系数(r)矩阵与SHAP解释性分析,揭示 δ 与热力学参数 Ω 在决定相稳定性中的主导作用。对比多种机器学习算法,发现基于集成学习策略的GBR模型,具有较好的预测结果($R^2 = 0.982$)与泛化能力。基于模型预测结果,确立了量化的相稳定性判据($\Omega > 1.1$ 且 $\delta < 6.6\%$),并为高熵碲

并效应较显著,HE-Te体系的电输运性能呈现平缓的下降趋势,进而维持了其在高温条件下的高电输出能力。

最终 zT 值的预测结果见图7(b)。由于理论功率因子(P_f)提升及预测晶格热导率($\kappa_{l,p}$)大幅降低,HE-Te体系的 zT 值实现了质的飞跃。计算结果显示,在850 K时,HE-Te体系的 zT 峰值达1.60,较PbTe基体(zT 值为0.96)提升了67%。即使在最保守的估计下,高熵化策略带来的性能提升依然具有统计学意义,这证明通过数据驱动的组分筛选策略,可有效突破传统热电材料中电、热输运参数的强耦合限制。该策略在大幅度降低高熵体系 κ_l 的同时,通过电子结构工程维持甚至提升其 P_f ,为新型高性能热电材料的设计提供了理论依据。

化物的快速初筛选提供了明确的半经验公式,有效缩小了实验探索范围。

2) 高熵组分的设计与筛选。将预测模型用于PbTe-SnTe-GeTe三元体系的全组分扫描,成功筛选出兼具热力学稳定性与低晶格热导率的“帕累托最优”组分(HE-Te体系)。理论计算结果显示,该组分在保持高对称性面心立方结构($Fm\bar{3}m$)的同时,通过高熵效应显著降低体系的 E_f (-46.31 kJ/mol)与 κ_l (0.85 W/mK)。这种基于数据驱动的高通量筛选策略,成功规避了GeTe基材料易发生菱方畸变的不利因素,实现了对晶体结构与热输运性能的协同优化。

3) 构效关系与性能机理研究。电子结构计算结果显示,高熵化引入的过渡金属Mn元素,可诱导导带底的能带重整与收敛,并促使Te p轨道与Mn d轨道在费米能级附近发生强烈杂化。这种“能带

收敛”与“轨道共振”的协同效应,有效实现了电、热运输参数解耦,使材料的 κ_1 大幅度降低并维持较高的 P_1 。性能预测结果显示,HE-Te高熵碲化物体系在850 K时的 zT 值达1.60,较传统基体提升67%。理论分析结果显示,该设计组分在电子结构和热输运特性方面均符合“声子玻璃-电子晶体”的设计理念。

文中提出的数据驱动框架,在理论设计与性能预测方面均表现出显著优势,但必须强调研究结果均为理论预测,后续的实验合成与性能验证仍在进行中。高熵热电材料在向实际应用转化过程中仍面临诸多挑战,例如,SQS超胞模型在描述真实高熵合金的局部原子构型时存在局限性,只能近似描述随机固溶体,无法捕捉局部化学短程有序、元素偏析等真实特征的微观结构信息。未来可结合团簇展开、蒙特卡洛等技术,进一步研究化学短程有序对材料电子结构和热输运性能的影响。此外,多主元体系在实际合成中极易遭遇元素偏析或氧化等技术瓶颈,导致在实际合成中要对理论预测的亚稳态结构的热稳定性及热、电输运性能进行严谨的实验验证。未来研究可结合高通量合成实验、原位表征技术及更精细的载流子散射模型,构建“计算-实验”双循环迭代系统,进一步修正预测模型的偏差,进而揭示复杂工况下多主元体系构效关系的演变,从而加速新型高性能高熵热电材料的工业化应用进程。

参考文献:

- [1] 肖颖,梁耕源,雷博文,等. 用于能量收集的离子热电材料研究进展[J]. 材料导报,2023,37(4):22020174. DOI:10.11896/cldb.22020174.
- [2] 陈群,郝俊红,陈磊,等. 电-热综合能源系统中能量的整体输运模型[J]. 电力系统自动化,2017,41(13):7-13.
- [3] 宋鑫芳,张勇. 高熵合金研究进展[J]. 粉末冶金技术,2022,40(5):451-457.
- [4] FAN Zhao, WANG Hui, WU Yuan, et al. Thermoelectric performance of PbSnTeSe high-entropy alloys [J]. Materials Research Letters, 2017, 5 (3) : 187-194.
- [5] JIANG Jian, CHENG Ruiqing, YIN Lei, et al. Van der Waals epitaxial growth of two-dimensional PbSe and its high-performance heterostructure devices [J]. Science Bulletin, 2022, 67(16): 1659-1668.
- [6] JIANG Binbin, WANG Wu, LIU Shixuan, et al. High figure-of-merit and power generation in high-entropy GeTe-based thermoelectrics [J]. Science, 2022, 377(6602): 208-213.
- [7] ZHANG Qian, LIAO Bolin, LAN Yucheng, et al. High thermoelectric performance by resonant dopant indium in nanostructured SnTe [J]. Proceedings of the National Academy of Sciences, 2013, 110(33): 13261-13266.
- [8] SU Meini, YOUNG Ben, GARDNER L. Flexural response of aluminium alloy SHS and RHS with internal stiffeners [J]. Engineering Structures, 2016, 121: 170-180.
- [9] LI Kai, SUN Liang, BAI Wei, et al. High-entropy strategy to achieve electronic band convergence for high-performance thermoelectrics [J]. Journal of the American Chemical Society, 2024, 146(20): 14318-14327.
- [10] 王钰莹,于海峰,黎兵,等. 如何利用开源工具分析fMRI数据:基于多体素模式的有监督机器学习算法介绍[J]. 心理科学,2022,45(3):718-724.
- [11] MANGALATHU S, HWANG S H, JEON J S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach [J]. Engineering Structures, 2020, 219: 110927. DOI:10.1016/j.engstruct.2020.110927.
- [12] 张聪,刘杰,解树一,等. 高通量计算与机器学习驱动高熵合金的研究进展[J]. 材料工程,2023,51(3):1-16.
- [13] 杨婉茹,贾宏伟,刘卓,等. CdMnO₄光催化材料的电子结构和光学性质的第一性原理计算[J]. 宁夏大学学报(自然科学版),2021,42(1):34-38.
- [14] 李景学,于跃,徐斯然,等. 电子自旋效应在电催化剂中的作用[J]. 物理化学学报,2023,39(12):2302049. DOI: 10.3866/PKU.WHXB202302049.
- [15] 王娟,杨小渝,王宗国,等. SQS二元合金设计的高通量方法和技术研究[J]. 计算机工程与科学,2016,38(3):401-410.
- [16] EVANS J. Straightforward statistics for the behavioral sciences [M]. Pacific Grove: Thomson Brooks/Cole Publishing Co, 1996.

Data-Driven High-Entropy Telluride Composition Design and Performance Prediction

LI Shun¹, PENG Haoran², FU Xiuli³

(1. Tianjin Research Institute for Water Transport Engineering, Tianjin 300456, China;

2. School of Mathematics and Physics, China University of Geosciences Beijing, Beijing 100080, China;

3. School of Integrated Circuits, Beijing University of Posts and Telecommunications, Beijing 100080, China)

Abstract: To address the vast compositional space and the inefficiency of traditional “trial-and-error” methods in the development of high-entropy telluride thermoelectric materials, this study establishes a theoretical data-driven design framework that integrates machine learning with first-principles calculations. By constructing a high-dimensional descriptor space that encompasses physical features such as mixing enthalpy and atomic size difference, a gradient boosting regression model was trained to accurately predict the formation energy of multiprincipal-element systems. SHAP (SHapley Additive exPlanations) interpretability analysis elucidated the competitive mechanisms between thermodynamic and geometric parameters, leading to the formulation of a quantitative criterion for phase formation. Using this criterion, a high-throughput theoretical screening of the entire PbTe-SnTe-GeTe ternary space was conducted. An optimal high-entropy composition featuring both high thermodynamic stability (-46.31 kJ/mol) and ultra-low lattice thermal conductivity (0.85 W/mK) was theoretically identified. Electronic structure calculations further demonstrated that the introduction of the transition metal Mn induces convergence of the conduction band and strong hybridization of p-d orbitals, which effectively decouples the electrical and thermal transport parameters. Theoretical evaluations suggest that this high-entropy composition could achieve a thermoelectric figure of merit (zT) of 1.60 at 850 K, representing an approximate 67% improvement over the PbTe matrix. This work preliminarily verifies the feasibility of data-driven strategies for rational theoretical design of complex thermoelectric materials and provides a clear framework for future experimental synthesis and validation. It is important to note that all results herein are theoretical predictions, and subsequent experimental synthesis and performance evaluations are ongoing.

Key words: high entropy telluride; machine learning; thermoelectric materials; phase stability; first-principles calculation

(责任编辑 高继红)