

智能模型高通量筛选无机钠固态电解质

刘悒泓, 毕文柱, Mohamed Ait Tamerd, 杨孟昊

(同济大学材料科学与工程学院车用新能源研究院, 上海市金属材料开发应用重点实验室, 上海 201804)

摘 要: 钠固态电解质因其能量密度高、安全性好和成本低等优势被视作极具潜能的下一代电解质材料, 然而较低的离子电导率成为限制钠固态电解质发展的主要瓶颈之一。本工作提出了一种机器学习与第一性原理计算结合的智能模型实现高通量筛选, 将离子电导率作为主要的筛选指标, 先后使用极端梯度提升(XGBoost)模型和从头算分子动力学(AIMD)预测并验证材料的离子电导率, 最终筛选得到 4 种潜在的无机钠固态电解质材料即 $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$, $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPCO}_7$ 以及 $\text{Na}_2\text{LuPWO}_8$, 并通过进一步分析探究钠离子的迁移机理。

关键词: 钠固态电解质; 离子电导率; 钠离子电池; 机器学习; 第一性原理计算

中图分类号: TM912; TP181 文献标志码: A 文章编号: 0454-5648(2025)07-1801-08

网络出版时间: 2025-05-26



随着人类对于能源短缺与环境污染的问题愈发重视, 新能源领域迎来了飞速发展。目前大规模商业化的二次电池主要为以有机溶剂和锂盐为电解液的锂离子电池, 该种电池虽然有更高的能量密度, 更快的充放电速度和更长的循环寿命, 但其存在两个明显的缺陷: 一是全球锂资源不足同时分布不均, 对锂日益增长的需求使得其价格水涨船高, 况且我国的锂资源又长期依赖进口, 寻找锂离子电池的替代品迫在眉睫; 二是由于有机溶剂易燃, 以有机溶剂作为电解液存在安全隐患, 尤其是在目前的技术下, 锂的枝晶生长无法得到完全的抑制, 容易导致电路短路引起电池起火乃至爆炸。这些缺陷严重地阻碍了锂离子电池的进一步发展。

钠资源充足且成本低, 被视为替代锂的绝佳材料^[1]。同时, 固态电解质(SSE)相较于传统的液态的电解液, 具有能量密度高, 热稳定性好、机械刚性强、安全性好、可以抑制枝晶生长的优势^[2-3], 因此钠固态电解质成了未来电池电解质的理想材料。为了提高钠固态电解质的性能, 现有研究主要聚焦在寻找离子电导率高的 SSE 新材料, 针对性设计 SSE 结构, 以及优化 SSE 与电极之间的界面等方向上^[4]。

针对钠固态电解质的研究可以追溯到 20 世纪 60 年代, $\beta\text{-Al}_2\text{O}_3$ 型化合物被发现具有较高的钠离子电导率。1976 年, Goodenough 等发现了钠超离子导体(sodium super ionic conductor, NASICON)^[5], 其在离子导电性、热稳定性和电化学稳定性上均有优异的表现, 被视作最有潜力的钠固态电解质材料, 但钠固态电解质的发展一度陷入停滞。近年来, 随着寻找锂离子电池替代品的需求不断增加, 伴随着人工智能的兴起, 钠固态电解质的发展来到了新的转折点。时至今日, 钠固态电解质的发展始终难以突破离子电导率低的瓶颈。离子电导率是 SSE 材料最重要的评价指标之一, 高离子电导率确保了钠离子能够在 SSE 中快速直接地传输, 从而提升电池的功率密度和充放电速率。

在钠固态电解质中, 无机固态电解质(ISE)相对于其他 2 类钠固态电解质(固态聚合物电解质, 复合固态电解质)通常具有更高的离子电导率和电化学稳定性^[6], 因此更有希望突破离子电导率低的限制。本工作旨在针对 ISE, 使用机器学习与第一性原理计算结合的智能模型高通量筛选方法搜索潜在的钠固态电解质材料。

收稿日期: 2024-12-11。 修订日期: 2024-12-16。

基金项目: 国家自然科学基金(52302302)。

第一作者: 刘悒泓(2001—), 男, 硕士研究生;

毕文柱(2005—), 男, 本科生。

通信作者: 杨孟昊(1990—), 男, 博士。

Received date: 2024-12-11. Revised date: 2024-12-16.

First author: LIU Yihong (2001—), male, Master candidate.

BI Wenzhu (2005—), male, Undergraduate.

E-mail: yihong_liu@tongji.edu.cn; down@tongji.edu.cn

Correspondent author: YANG Menghao (1990—), male, Ph.D.

E-mail: menghaoyoung@tongji.edu.cn

目前关于高通量筛选钠固态电解质的研究中, 大多是基于某种已知的 SSE 类型, 通过掺杂或取代生成新的材料并用机器学习算法进行筛选, 从而预测出新的 SSE 材料, 例如 NASICON 类型^[7–9]、argyrodite 类型^[10]。因此通过上述途径筛选得到的新材料的晶体结构与原先已知的材料相似。而本工作的筛选范围为所有已知的无机含钠化合物, 希望从中筛选出多种不同类型的新的钠固态电解质。

高通量筛选的流程主要分为 3 步: 首先建立钠固态电解质的数据集; 然后训练一套机器学习模型用以预测含钠无机化合物的离子电导率; 最后用从头算分子动力学(AIMD)模拟对预测得到的离子电导率进行验证, 完成筛选。在整个筛选过程中, 本工作不只关注离子电导率, 也考虑了电子绝缘性、热稳定性和化学稳定性等影响 SSE 性能的重要性质^[11]。

为了能更深入地研究这些筛选得到的 SSE 材料, 本工作进一步对 AIMD 的模拟结果进行分析, 观测单个离子的运动轨迹, 并且通过计算得到激活能, 钠离子概率密度分布和 van Hove 相关函数等信息^[12]。在此基础上, 使用爬升图像轻推弹性带(CI-NEB)方法研究计算特定扩散路径上的扩散能垒^[13]。通过第一性原理计算, 试图去理解材料结构、钠离子的传导机理与宏观离子电导率之间的联系。

1 方法

1.1 数据集

本数据集中所有数据均来源于 Materials Project 数据库^[14]。在 Materials Project 数据库所有的 153 235 种无机材料中, 含钠的共有 12 823 种。本数据集收集了这 12 823 种材料的晶体信息、带隙以及高于凸包的能量(E_{hull}), 其中后两者的作用是对数据库进一步进行筛选。

固态电解质除了需要具有较高的离子电导率外, 还需要确保材料的电子传导能力弱, 即电子电导率低, 从而尽可能避免电解质自放电引起的电池性能降低、寿命减短、失控风险提高等问题。通常低于 2 eV 的带隙不足以确保材料具有很好的绝缘能力^[13]。另外, 材料的热力学稳定性也是重要的评价固态电解质的指标, 用 E_{hull} 衡量热力学稳定性, E_{hull} 的值越小代表材料越稳定。综上所述, 针对固态电解质的电子传导能力和热力

学稳定性, 对原先的数据集进行筛选, 带隙小于 1.5 eV 或者 E_{hull} 大于 0.03 eV/atom 的材料被排除^[15], 最终经筛选的数据集共有 3631 种含钠无机化合物。

1.2 机器学习

基于上述数据集, 建立了机器学习模型来预测数据集中所有材料的离子电导率。

在建立机器学习回归模型之前, 需要将数据库中的材料组分信息转化为可以被机器学习模型理解和处理的数据。使用 Magpie(materials agnostic platform for informatics and exploration)生成材料描述符^[16]。Magpie 会编码生成 145 种属性, 这些属性描述了材料的化学计量信息、元素性质统计信息、电子结构信息以及离子化合物信息。

为了能使回归模型达到最佳性能, 总共尝试了 4 种机器学习算法: 随机森林(RF)^[17], 极端梯度提升(XGBoost)^[18], 类别型特征提升(CatBoost)^[19], K 近邻(KNN)^[20]。这些算法基于 scikit-learn 库实现^[21]。

为了训练机器学习模型, 需要另外收集材料离子电导率数据。数据是训练机器学习模型的基础, 其质量直接决定了模型性能^[22]。材料领域中数据普遍存在来源广、样本少、维度高等问题, 因此需要重视材料数据品质的治理^[23]。因此选择了来源可靠的动态固态电解质数据库(DDSE)^[24], 收集材料的离子电导率作为训练机器学习模型的数据, 并且对这些数据进行了二次校验来确保机器学习模型的准确性和可靠性。

将所有钠固态电解质数据按照 7:3 的比例随机划分为训练集和测试集, 在训练集上使用 5 折交叉验证筛选最佳超参数组合并测试模型预测能力, 在测试集上测试模型的泛化能力。采用决定系数(R^2)评估模型预测值与真实值之间的误差。

除了预测离子电导率外, 用同样的方法训练了机器学习模型来分别预测氧化势和还原势, 以此来考察材料的电化学稳定性。当预测的离子电导率数值相近时, 将氧化势和还原势作为进一步筛选的依据。

1.3 AIMD

为了进一步验证所筛选得到的钠固态电解质的性能(验证机器学习预测的可靠性), 在对晶体结构进行优化后, 使用 AIMD 模拟来计算其离子电导率。

使用 VASP(vienna ab initio simulation package) 进行 AIMD 模拟^[25-26], 从而计算化合物的离子电导率。具体地, 电子相互作用采用投影增强波(PAW)方法描述^[27], 并以 Perdew-Burke-Ernzerhof(PBE)广义梯度近似(GGA)作为电子交换相关泛函进行第一性原理计算^[28]。 K 点设置为以 Γ 点为中心的 $1 \times 1 \times 1$ 网格, 平面波截断能设置为 500 eV, 电子迭代收敛标准为 10^{-5} eV。

由于在室温下进行 AIMD 模拟时, 钠离子的扩散速度过慢, 无法计算得到较为精确的离子电导率值, 因此需要首先在多个高温下进行模拟, 得到各温度下的原子运动轨迹文件, 由此便可以计算出在每个时间步长内所有钠离子位移的平方的平均值, 即均方位移, 再由均方位移除以时间间隔即可得到扩散系数^[29]:

$$D = \frac{1}{2Nd\Delta t} \sum_{i=1}^N \langle |r_i(t + \Delta t) - r_i(t)|^2 \rangle_t \quad (1)$$

式中: N 为扩散微粒的总数; d 为扩散系统的维度; Δt 为时间间隔; $r_i(t)$ 为第 i 个扩散微粒在 t 时刻的位置; $\langle \rangle_t$ 代表对时间 t 取平均值。

然后将不同温度下的扩散系数基于阿伦尼乌斯(Arrhenius)公式进行线性拟合, 即可得到活化能并外推得到 300 K 时化合物的扩散系数, 进而根据 Nernst-Einstein 方程计算得到室温下的离子电导率:

$$\sigma = \frac{nq^2}{k_B T} D \quad (2)$$

式中: n 为扩散粒子(即钠离子)的密度; q 为该电荷所带电荷; k_B 为玻尔兹曼(Boltzmann)常量。

AIMD 模拟在 NVT 系统中进行, 选取了 600、900、1200、1500 K 这 4 种温度模拟离子的扩散。在每个目标温度下, 以 2.0 fs 的时间步长进行 10 000 步模拟, 即总共模拟 20 ps, 为实现精准控温采用能势-胡佛(Nosé-Hoover)恒温器^[30-31]。

为了能够让 AIMD 模拟的结果更加直观, 使用 VESTA 软件实现晶体结构的可视化^[32]。

1.4 CI-NEB

采用 CI-NEB 方法来寻找特定的钠离子扩散路径上的过渡态^[33], 分别计算每个过渡态的体系能量并进行振动自由能修正, 将相邻过渡态的修正能量相减即可得到对应能垒。

需要注意的是, NEB 方法是通过在晶体内人为引入一个离子空位来计算离子迁移的, 这并不符合大多数晶体的实际迁移机制。因此该方法计算得到

的能量势垒只能用于比较同一或相似晶体中沿不同局域路径离子迁移的难易程度^[13]。

2 结果与讨论

比较 4 种机器学习回归模型在训练集和测试集上的表现, 由图 1 可以发现, XGBoost 在训练集和测试集上的 R^2 值均为最大, 明显优于其他模型。值得一提的是, XGBoost 在测试集上的 R^2 值达到了 0.994, 说明该模型的拟合效果非常好, 证明了该模型具有强大的泛化能力。

采用性能最好的 XGBoost 模型对所有的 3631 种含钠无机化合物进行预测, 发现其中有 126 种材料具备成为优秀 SSE 的潜力。通过进一步筛选, 从中选择了 14 种化合物进行 AIMD 模拟, 其中 4 种材料表现出了较高的离子电导率和较低激活能, 分别是 $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$, $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPCO}_7$ 及 $\text{Na}_2\text{LuPWO}_8$ 。这些候选 SSE 材料的晶体结构如图 2 所示, 另外它们的机器学习预测离子电导率以及通过 AIMD 计算得到的离子电导率与激活能在表 1 中给出。

表 1 4 种候选 SSE 的机器学习预测离子电导率(σ_{ML}), AIMD 计算离子电导率(σ_{AIMD})与激活能(E_a)
Table 1 Machine learning predicted ionic conductivity (σ_{ML}), AIMD calculated ionic conductivity (σ_{AIMD}) and activation energy (E_a) for four candidate SSEs

Formula	$\sigma_{\text{ML}}/(\text{mS} \cdot \text{cm}^{-1})$	$\sigma_{\text{AIMD}}/(\text{mS} \cdot \text{cm}^{-1})$	E_a/meV
$\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$	0.152	6.056	186
$\text{Na}_3\text{Bi}(\text{BO}_3)_2$	0.192	5.286	176
$\text{Na}_2\text{LuPCO}_7$	0.703	0.705	237
$\text{Na}_2\text{LuPWO}_8$	0.140	0.601	234

从表 1 可以发现只有 $\text{Na}_2\text{LuPCO}_7$ 离子电导率的机器学习预测值和 AIMD 计算值吻合得较好, 这一结果可能源于机器学习模型预测的误差, 对于复杂的预测任务, 即便是 1 个性能优异的机器学习模型也可能产生 1 个数量级以内的误差。除了机器学习预测误差外, 也有可能是由于同一材料在不同温度下进行 AIMD 模拟时所采用初始结构完全一致, 而实际上随着温度的变化材料晶体结构可能会发生变化, 这就会给离子电导率的 AIMD 计算值引入额外的误差。不过, 由于机器学习预测和 AIMD 计算值之间的差距保持在不大于 1 个数量级, 因此仍然可以认为本工作的高通量筛选方法是可靠的。

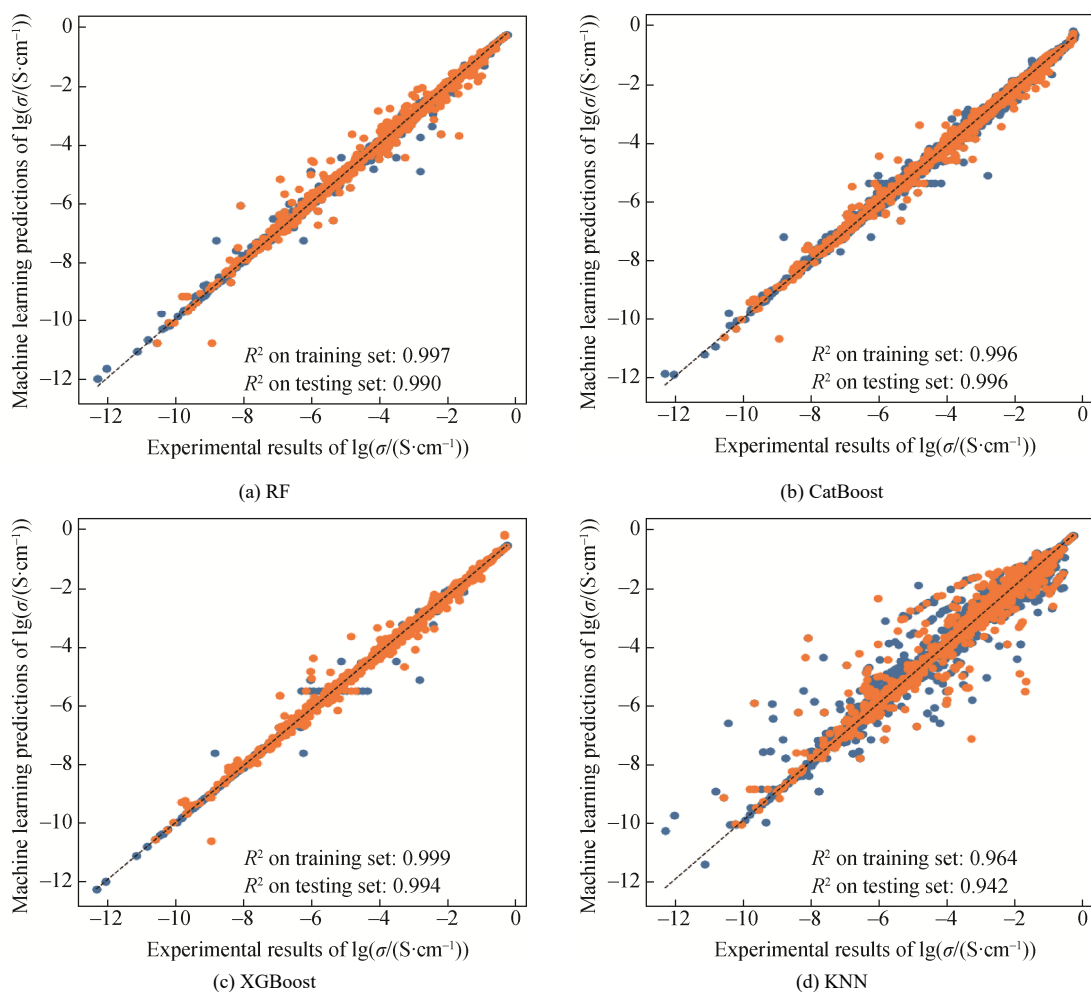


图 1 4 种机器学习模型预测值与真实值的比较

Fig. 1 Comparison of the predictions and true values of four machine learning models

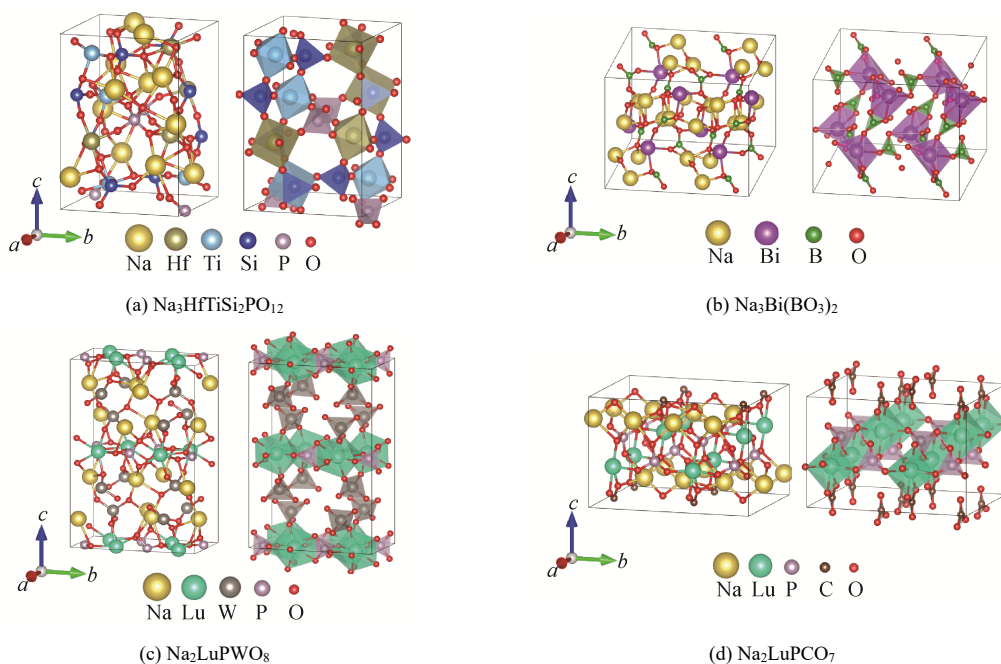


图 2 各材料的晶体结构示意图

Fig. 2 Crystal structure of SSE samples

为了深入探究钠离子的迁移机理,首先观察每种化合物中各类微粒的运动轨迹,能发现在所有的模拟温度下,钠离子的扩散行为都十分明显,而其他离子保持相对稳定,形成可靠的通道以便钠离子进行迁移。进一步可以绘制各个材料的 van Hove 相关函数图。van Hove 相关函数可以分为自相关 G_s 和不同相关 G_d 2 部分。自相关部分描述同一粒子在不同时间点的径向分布,不同相关部分描述的则是不同粒子之间的相对径向分布,这对理解钠离子迁移机理至关重要,揭示钠离子间的相互作用和协同运动。由图 3a 所示,4 种材料均一定程度上展现出了不同离子迁移的相关性,其中 $\text{Na}_3\text{Bi}(\text{BO}_3)_2$ 的相关性相对较弱。不同离子迁移存在相关性说明了钠离子并非独立移动而是协同扩散,这对于提高离子电导率做出重要贡献。通过图 3b 则可以发现 4 种材料在 0~1 Å 之间均存在明显的峰, $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$ 的峰衰减得非常快,意味

着钠离子能够很快地扩散到其他位点中;而 $\text{Na}_2\text{LuPWO}_8$ 的峰却几乎不衰减。

在 4 种候选 SSE 材料中, $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$ 是一种 NASICON 材料^[34],具有典型结构:晶体的主要框架由 HfO_6 , TiO_6 八面体和 PO_4 , SiO_4 四面体组成,它们通过共享顶点形成一个三维的网络,网格中的大量空隙成为钠离子可以通过的通道,使得钠离子能够快速迁移,这便是 NASICON 具有高离子导电性的核心。由于已经存在大量关于 NASICON 中钠离子传导机理的研究^[35-36],本工作将不再对 $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$ 作进一步分析。值得一提的是,在 2023 年的一项关于“基于机器学习的新方法识别有前景的钠离子固态电解质 NASICON 材料”的研究中^[7],确定了 $\text{Na}_3\text{HfZrSi}_2\text{PO}_{12}$ 是具有开发前景的钠 SSE 材料,这与所筛选到的 $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$ 具有极高的相似性,只是将 Ti 取代为 Zr,从而侧面印证了本工作机器学习模型预测的可靠性。

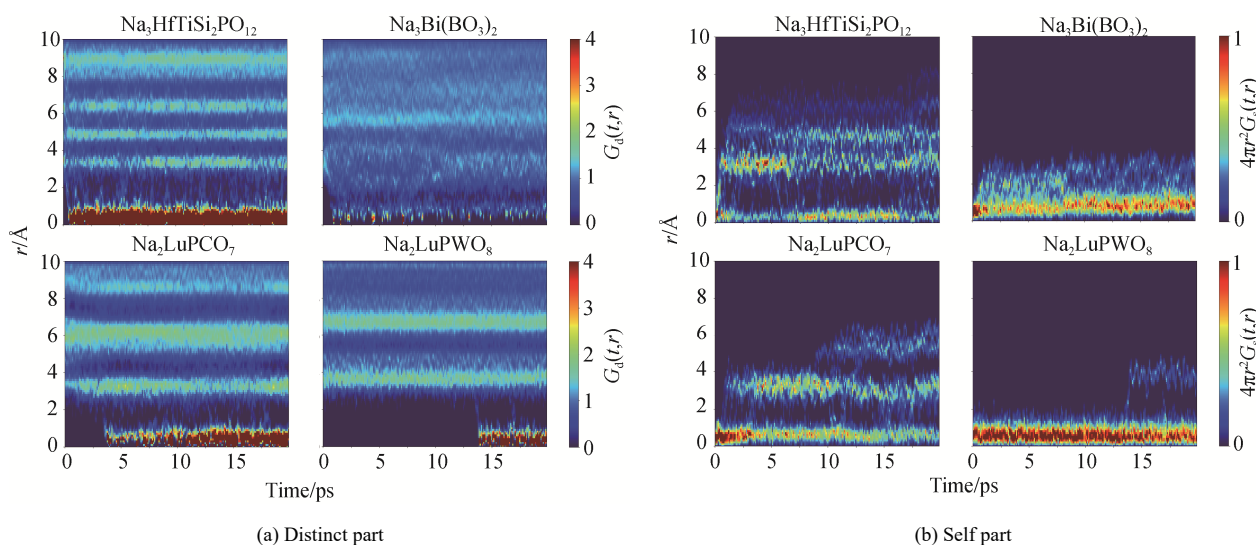


图 3 4 种候选 SSE 的 van Hove 相关函数
Fig. 3 van Hove correlation function of four candidate SSEs

除了 $\text{Na}_3\text{HfZrSi}_2\text{PO}_{12}$ 外, $\text{Na}_3\text{Bi}(\text{BO}_3)_2$ 凭借其较高的离子电导率(σ_{AIMD})和较低的激活能成为最有潜力的 SSE 材料。钠离子在 900 K 的概率密度分布在图 4a 中展示,可以从概率密度分布图中观察到钠离子的迁移路径。为了方便观察,图 4a 中的所有钠离子都被隐藏。可以明显地看到钠离子沿着由 Bi, B, O 所构成的通道在 bc 平面上迁移。值得注意的是,并不是所有的钠离子均能够发生迁移,不在该通道内的钠离子的移动受到限制,从而只能在平衡位置附近振荡,这无疑抑制了离

子电导率的提高。

$\text{Na}_2\text{LuPCO}_7$ 和 $\text{Na}_2\text{LuPWO}_8$ 具有相似的化学式,甚至两者通过 AIMD 模拟得到的离子电导率也都十分相近,但实际上两者在晶体结构上存在不小的差异。如前文中图 2c~图 2d 所示, $\text{Na}_2\text{LuPCO}_7$ 属于单斜晶系 $P12_1/m1$ 空间群,1 个 LuO_6 八面体与 4 个 PO_4 四面体共点连接,构成了主要的扩散框架,为钠离子提供充足且稳定的扩散空间;而 $\text{Na}_2\text{LuPWO}_8$ 属于斜方晶系 $Ibca$ 空间群, LuO_8 多面体以及 PO_4 , WO_4 四面体共同构成扩散框架,其中 PO_4 以共边的

方式与 LuO_8 以及 WO_4 连接。两者晶体结构上的差异也导致了钠离子扩散路径的不同。进而分析钠离子的概率密度分布,如图 4c 所示,在 $\text{Na}_2\text{LuPCO}_7$ 中,所有钠离子都会沿着 b 轴方向进行扩散。由于在扩散路径上存在碳及其相连的 3 个氧所构成的三角形平面,且该平面与 b 轴方向(即钠离子迁移方向)垂直,阻碍了钠离子的扩散,因此钠离子需要绕过 CO_3 平面才能实现迁移。基于 CI-NEB 方法计算了钠离子迁移能垒,计算结果与 AIMD 模拟结果吻合:当钠离子扩散需要穿过 CO_3 平面时,会有向 c 轴正

方向绕过该平面的趋势,并且需要克服较大的能垒,如图 4e 所示;而若无 CO_3 阻碍,如图 4d 所示,则钠离子的迁移会自发开始。 $\text{Na}_2\text{LuPWO}_8$ 中钠离子在由 Lu, P, O 构成的层状结构内沿 a 方向扩散,如图 4b 所示。钨离子分布在层间,连接层与层,以维持结构稳定,但这也在空间上妨碍了钠离子的迁移。并且由于层间钨离子的数量相较于钠离子要少,支撑作用有限,因此钠离子在晶体中也需要起到稳固层与层的作用,因此无法相对自由地移动,这也进一步限制了钠离子的迁移。

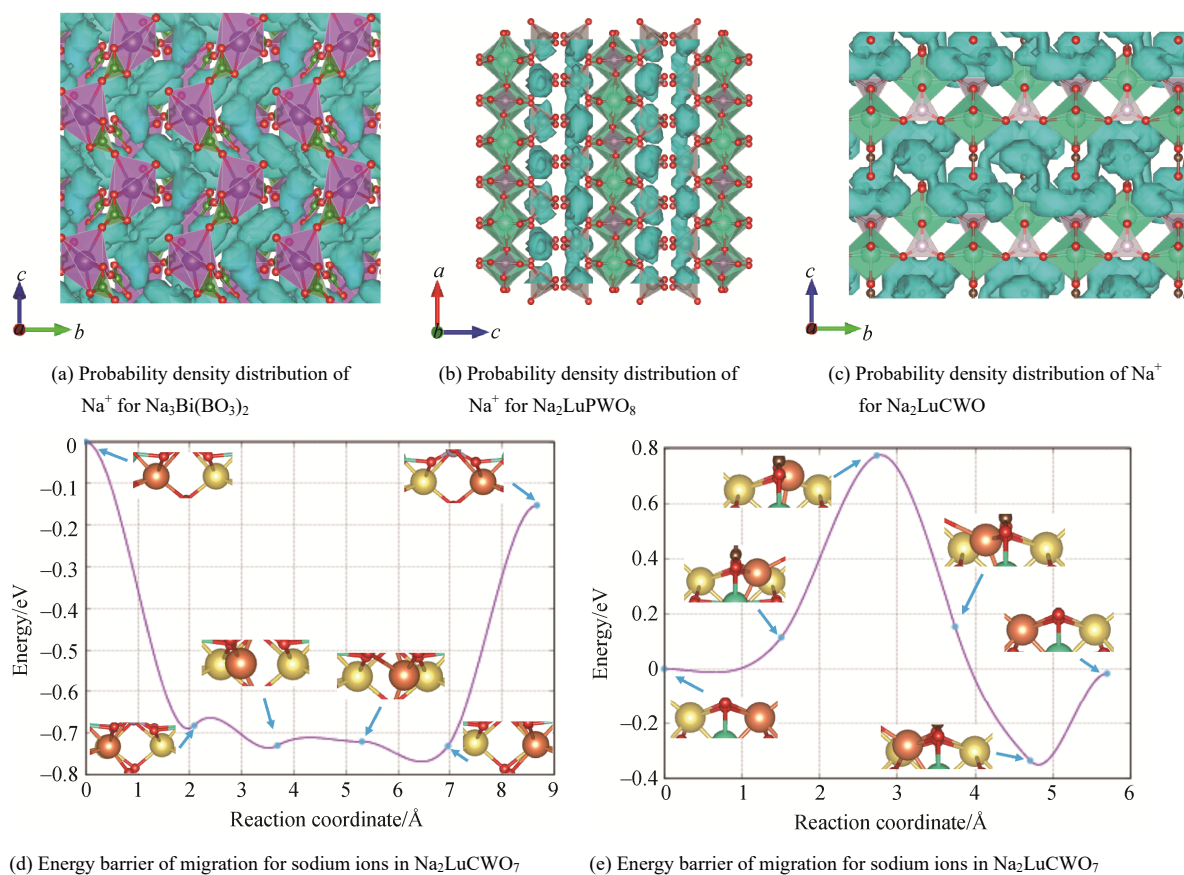


图4 $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPWO}_8$ 和 $\text{Na}_2\text{LuPCO}_7$ 在 900 K 时的钠离子概率密度分布, $\text{Na}_2\text{LuPCO}_7$ 中钠离子的迁移能垒
Fig. 4 Probability density distribution of Na^+ at 900 K for $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPWO}_8$ 和 $\text{Na}_2\text{LuCWO}_7$, energy barrier of migration for sodium ions in $\text{Na}_2\text{LuCWO}_7$

3 结论

1) 从这 126 种化合物中选取 14 种进行 AIMD 模拟, 计算其离子电导率和激活能, 最终筛选得到 4 种具有高离子电导率的无机含钠化合物, 分别为 $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$, $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPCO}_7$ 以及 $\text{Na}_2\text{LuPWO}_8$ 。

2) 基于 AIMD 的模拟结果, 即离子运动轨迹, 还可以得到 van Hove 相关函数和概率密度分布。对

AIMD 模拟结果进行分析, 结果显示 4 种具有高离子电导率材料的共性是具有稳定的钠离子扩散通道, 并且钠离子以协调扩散的方式进行迁移。在这 4 种含钠无机化合物中, $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$ 属于 NASICON 类, 且与先前报道的具有潜力的 $\text{Na}_3\text{HfZrSi}_2\text{PO}_{12}$ 极其相似, 侧面证明了该筛选方法的可靠性。

3) 基于晶体结构对其余 3 种材料中钠离子的迁移机理进行了分析, 希望为 SSE 的结构设计提供一些新的思路。

参考文献:

- [1] KIM S W, SEO D H, MA X H, et al. Electrode materials for rechargeable sodium-ion batteries: Potential alternatives to current lithium-ion batteries[J]. *Adv Energy Mater*, 2012, 2(7): 710–721.
- [2] KIM J J, YOON K, PARK I, et al. Progress in the development of sodium-ion solid electrolytes[J]. *Small Meth*, 2017, 1(10): 1700219.
- [3] ZHAO C L, LIU L L, QI X G, et al. Solid-state sodium batteries[J]. *Adv Energy Mater*, 2018, 8(17): 1703012.
- [4] AHMAD H, KUBRA K T, BUTT A, et al. Recent progress, challenges, and perspectives in the development of solid-state electrolytes for sodium batteries[J]. *J Power Sources*, 2023, 581: 233518.
- [5] GOODENOUGH J B, HONG H Y, KAFALAS J A. Fast Na⁺-ion transport in skeleton structures[J]. *Mater Res Bull*, 1976, 11(2): 203–220.
- [6] LI Z P, LIU P, ZHU K J, et al. Solid-state electrolytes for sodium metal batteries[J]. *Energy Fuels*, 2021, 35(11): 9063–9079.
- [7] KIM J, KANG S, MIN K. Screening platform for promising Na superionic conductors for Na-ion solid-state electrolytes[J]. *ACS Appl Mater Interfaces*, 2023, 15(35): 41417–41425.
- [8] ZHANG Y, ZHAN T, SUN Y, et al. Revolutionizing solid-state NASICON sodium batteries: Enhanced ionic conductivity estimation through multivariate experimental parameters leveraging machine learning[J]. *ChemSusChem*, 2024, 17(6): e202301284.
- [9] ZHOU P F, ZHAO Z R, SUN K T, et al. Machine learning guided cobalt-doping strategy for solid-state NASICON electrolytes[J]. *Eur J Inorg Chem*, 2023, 26(26): e202300382.
- [10] LEE B D, GAVALI D S, KIM H, et al. Discovering virtual Na-based argyrodites as solid-state electrolytes using DFT, AIMD, and machine learning techniques[J]. *J Mater Chem A*, 2025, <https://doi.org/10.1039/D4TA06927G>.
- [11] GUIN M, TIETZ F, GUILLON O. New promising NASICON material as solid electrolyte for sodium-ion batteries: Correlation between composition, crystal structure and ionic conductivity of Na_{3+x}Sc₂Si₆P_{3-x}O₁₂[J]. *Solid State Ion*, 2016, 293: 18–26.
- [12] DENG Z, ZHU Z Y, CHU I H, et al. Data-driven first-principles methods for the study and design of alkali superionic conductors[J]. *Chem Mater*, 2017, 29(1): 281–288.
- [13] XIAO W S, WU M W, WANG H, et al. Li-ion transport mechanisms in selenide-based solid-state electrolytes in lithium-metal batteries: A study of Li₈SeN₂, Li₇PSe₆, and Li₆PSe₅X (X = Cl, Br, I)[J]. *Energy Environ Mater*, 2024, 7(5): e12729.
- [14] JAIN A, ONG S P, HAUTIER G, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation[J]. 2013, 1(1): 011002.
- [15] HONRAO S J, YANG X, RADHAKRISHNAN B, et al. Discovery of novel Li SSE and anode coatings using interpretable machine learning and high-throughput multi-property screening[J]. *Sci Rep*, 2021, 11(1): 16484.
- [16] WARD L, AGRAWAL A, CHOUDHARY A, et al. A general-purpose machine learning framework for predicting properties of inorganic materials[J]. *NPJ Comput Mater*, 2016, 2: 16028.
- [17] BREIMAN L. Random forests[J]. *Mach Learn*, 2001, 45(1): 5–32.
- [18] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA. ACM, 2016.
- [19] Prokhorenkova L, GUSEV G, VOROBEOV A, et al. CatBoost: unbiased boosting with categorical features[C]//32nd Conference on Neural Information Processing Systems (NeurIPS 2018). Montréal, Canada, 2018: 31.
- [20] COVER T, HART P. Nearest neighbor pattern classification[J]. *IEEE Trans Inf Theory*, 1967, 13(1): 21–27.
- [21] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine learning in python [J]. *J. Mach Learn Res*, 2011, 12: 2825–2830.
- [22] CHEN H H, CHEN J P, DING J H. Data evaluation and enhancement for quality improvement of machine learning[J]. *IEEE Trans Reliab*, 2021, 70(2): 831–847.
- [23] LIU Y, YANG Z, ZOU X, et al. Data quantity governance for machine learning in materials science[J]. *Natl Sci Rev*, 2023, 10(7): nwad125.
- [24] YANG F L, CAMPOS DOS SANTOS E, JIA X, et al. A dynamic database of solid-state electrolyte (DDSE) picturing all-solid-state batteries[J]. *Nano Mater Sci*, 2024, 6(2): 256–262.
- [25] KRESSE G, FURTHMÜLLER J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set[J]. *Phys Rev B Condens Matter*, 1996, 54(16): 11169–11186.
- [26] KRESSE G, FURTHMÜLLER J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set[J]. *Comput Mater Sci*, 1996, 6(1): 15–50.
- [27] BLÖCHL P E. Projector augmented-wave method[J]. *Phys Rev B*, 1994, 50(24): 17953–17979.
- [28] PERDEW J P, BURKE K, ERNZERHOF M. Generalized gradient approximation made simple[J]. *Phys Rev Lett*, 1996, 77(18): 3865–3868.
- [29] HE X F, ZHU Y Z, EPSTEIN A, et al. Statistical variances of diffusional properties from *ab initio* molecular dynamics simulations[J]. *NPJ Comput Mater*, 2018, 4: 18.
- [30] HOOVER W G. Canonical dynamics: Equilibrium phase-space distributions[J]. *Phys Rev A*, 1985, 31(3): 1695–1697.
- [31] NOSÉ S. A unified formulation of the constant temperature molecular dynamics methods[J]. 1984, 81(1): 511–519.
- [32] MOMMA K, IZUMI F. *VESTA 3* for three-dimensional visualization of crystal, volumetric and morphology data[J]. *J Appl Crystallogr*, 2011, 44(6): 1272–1276.
- [33] HENKELMAN G, UBERUAGA B P, JÓNSSON H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths[J]. 2000, 113(22): 9901–9904.
- [34] OUYANG B, WANG J, HE T, et al. Synthetic accessibility and stability rules of NASICONs[J]. *Nat Commun*, 2021, 12(1): 5752.
- [35] ZHANG Z Z, ZOU Z Y, KAUP K, et al. Correlated migration invokes higher Na⁺-ion conductivity in NASICON-type solid electrolytes[J]. *Adv Energy Mater*, 2019, 9(42): 1902373.
- [36] LI C, LI R, LIU K N, et al. NaSICON: A promising solid electrolyte for solid-state sodium batteries[J]. *Interdiscip Mater*, 2022, 1(3): 396–416.

Intelligent Model for High Throughput Screening of Inorganic Sodium Solid-State Electrolytes

LIU Yihong, BI Wenzhu, Mohamed Ait Tamerd, YANG Menghao

(Shanghai Key Laboratory for R&D and Application of Metallic Functional Materials, Institute of New Energy for Vehicles, School of Materials Science and Engineering, Tongji University, Shanghai 201804, China)

Extended Abstract

Introduction Development of new energy becomes popular due to the energy shortages and environmental pollution. The existing commercial secondary batteries are primarily lithium-ion batteries, which use organic solvents and lithium salts as electrolytes. These batteries have two significant drawbacks, *i.e.*, global lithium resources are insufficient and unevenly distributed, and organic solvents are flammable, having safety risks. These drawbacks severely affect the further development of lithium-ion batteries. Sodium solid-state electrolytes (SSEs) are emerged as an ideal material for future battery electrolytes due to the high energy density, good thermal stability, strong mechanical rigidity, low cost and improved safety. This paper was to explore a potential sodium SSEs by a high-throughput screening method based on machine learning (ML) and first-principles calculations. The potential sodium SSEs were predicted by machine learning models and were validated through Ab initio molecular dynamics (AIMD) simulations. In addition, the conduction mechanism of sodium ions was also analyzed based on the results of the first-principles calculations.

Methods The dataset of inorganic sodium-containing compounds was established before building ML models. All the data in this dataset were from the Materials Project database. Solid-state electrolytes have good thermodynamic stability and insulation. The materials with a band gap less than 1.5 eV or energy above hull greater than 0.03 eV/atom were excluded, resulting in a final dataset of 3631 sodium-containing inorganic compounds.

To achieve the optimal performance of the regression model, we tried four ML algorithms, *i.e.*, Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and K-Nearest Neighbors (KNN). Magpie (Materials Agnostic Platform for Informatics and Exploration) was used to obtain the crystal structure information into 145 attributes as the input of ML models. To validate the performance of the screened sodium solid-state electrolytes and further analyze the migration mechanism of sodium ions, AIMD simulations and CI-NEB (Climbing Image Nudged Elastic Band) calculations were conducted to calculate their ionic conductivity, activation energy, probability density distribution, van Hove correlation functions and energy barriers along specific diffusion paths.

Results and discussion In the four machine learning regression models, XGBoost has the optimum performance on both training and testing sets with $R^2 = 0.999$ and 0.994 , respectively, indicating its high accuracy and strong generalization ability. The XGBoost model is used to predict 3631 of sodium-containing inorganic compounds, discovering that 126 of them have a potential to be superior SSEs. We further screen these and select 14 compounds for AIMD simulations, among which four materials (*i.e.*, $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$, $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPCO}_7$ and $\text{Na}_2\text{LuPWO}_8$) show a high ionic conductivity and a low activation energy. Since the difference between the machine learning prediction and AIMD calculated value is within an order of magnitude, the high-throughput screening method used is reliable. Among the four candidate SSE materials, $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$ belongs to the NASICON family. A previous study indicates that $\text{Na}_3\text{HfZrSi}_2\text{PO}_{12}$ is a promising sodium SSE material, which is similar to $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$, except that Ti is replaced by Zr, indirectly proving the reliability of the machine learning model prediction.

To clarify the sodium ion migration mechanism, the motion trajectories of various particles in each compound are investigated. The results show that at all simulation temperatures, the diffusion of sodium ions is evident, while other ions remain relatively stable, forming reliable channels for sodium ion migration. From the distinct parts, the four materials exhibit migration correlation, indicating that sodium ions do not move independently but rather diffuse cooperatively, which contributes significantly to the ionic conductivity. The migration pathways of $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPCO}_7$ and $\text{Na}_2\text{LuPWO}_8$ are discussed according to the probability density distributions. Moreover, the energy barrier of migration in $\text{Na}_2\text{LuPCO}_7$ shows that the maximum obstacle is to bypass the triangular plane formed by CO_3 within the diffusion channel.

Conclusions We used the best-performing XGBoost model to search the established dataset of 3631 inorganic sodium-containing compounds, identifying 126 compounds with a high ionic conductivity. Also, we selected 14 of the 126 compounds for AIMD simulations to calculate their ionic conductivity and activation energy, ultimately identifying four inorganic sodium-containing compounds (*i.e.*, $\text{Na}_3\text{HfTiSi}_2\text{PO}_{12}$, $\text{Na}_3\text{Bi}(\text{BO}_3)_2$, $\text{Na}_2\text{LuPCO}_7$ and $\text{Na}_2\text{LuPWO}_8$) with a high ionic conductivity. We derived the van Hove correlation functions and probability density distributions based on the AIMD simulation results. It was indicated that the common characteristic of the four high ionic conductivity materials could be the presence of stable sodium ion diffusion channels with sodium ions migrating in a coordinated manner. Finally, the migration mechanism of sodium ions was analyzed based on the results of AIMD and CI-NEB.

Keywords sodium solid-state electrolytes; ionic conductivity; Na-ion batteries; machine learning; first principle calculations