7 April 2026

From Data to Design: Physics-Informed Symbolic Regression Reveals Ionic Mobility Descriptors in Batteries

Mohsen Sotoudeh[1,2,3], Udaykumar Gajera[3,4], Axel Groß[1,3]

1. Helmholtz Institute Ulm (HIU), Electrochemical Energy Storage
2. Karlsruhe Institute of Technology (KIT
3. Institute of Theoretical Chemistry Ulm University
4. Department of Physics and CSMB Humboldt-Universität zu Berlin

Abstract

Large, high-quality datasets are essential for accelerating materials discovery. However, challenges related to data accessibility and analysis often prevent these datasets from having full impact. Here, we introduce DISCOVER, an open-source descriptor analysis framework designed to facilitate systematic data exploration and descriptor identification for desired or undesired material functions. Using this framework, descriptors can be identified even for complex materials properties, enabling interpretable and data-driven optimization of targeted materials. Alongside the framework, we present an open-access database containing 765 spinel compounds that are relevant to battery applications. These compounds were derived from periodic density functional theory (DFT) simulations, which provide the structural, electronic, and electrochemical properties necessary for understanding and optimizing spinel materials for use in energy storage technologies. Applying this database together with DISCOVER, we uncover a universal, physics-informed descriptor for ionic mobility across oxide, sulfide, and selenide spinels. Specifically, we demonstrate that the migration barrier is fundamentally governed by a direct competition between the thermodynamic site preference and the geometric anion positional parameter. By providing this open-access resource coupled with analytical tools, we aim to promote reproducibility and accelerate the data-driven design of high-performance energy storage materials.

From Data to Design: Physics-Informed Symbolic Regression Reveals Ionic Mobility Descriptors in Batteries

Mohsen Sotoudeh [1,2,3], Udaykumar Gajera[3,4], Axel Groß [3,1]

[1] *Helmholtz Institute Ulm (HIU), Electrochemical Energy Storage, 89081 Ulm, Germany*

[2] *Karlsruhe Institute of Technology (KIT), P.O. Box 3640, D-76021 Karlsruhe, Germany*

[3] *Institute of Theoretical Chemistry, Ulm University, Oberberghof 7, 89081 Ulm, Germany*

[4] *Department of Physics and CSMB, Humboldt-Universität zu Berlin, Berlin, Germany*

Large, high-quality datasets are essential for accelerating materials discovery. However, challenges related to data accessibility and analysis often prevent these datasets from having full impact. Here, we introduce DISCOVER, an open-source descriptor analysis framework designed to facilitate systematic data exploration and descriptor identification for desired or undesired material functions. Using this framework, descriptors can be identified even for complex materials properties, enabling interpretable and data-driven optimization of targeted materials. Alongside the framework, we present an open-access database containing 765 spinel compounds that are relevant to battery applications. These compounds were derived from periodic density functional theory (DFT) simulations, which provide the structural, electronic, and electrochemical properties necessary for understanding and optimizing spinel materials for use in energy storage technologies. Applying this database together with DISCOVER, we uncover a universal, physics-informed descriptor for ionic mobility across oxide, sulfide, and selenide spinels. Specifically, we demonstrate that the migration barrier is

fundamentally governed by a direct competition between the thermodynamic site preference and the geometric anion positional parameter. By providing this open-access resource coupled with analytical tools, we aim to promote reproducibility and accelerate the data-driven design of high-performance energy storage materials.

Main

Accelerated material discovery depends on the extraction of meaningful insights from complex high-dimensional data [1]. Collecting measurements is not enough. The discovery of materials requires extracting the underlying physical relationships—often nonlinear—that control function [2]. Although first-principles models can, in principle, describe these relationships, the intricate interplay of multiple processes often renders explicit modeling impractical. An alternative approach is to identify descriptors—physical parameters that capture correlations between material properties and desired or undesired functionalities [1,3,4]. Once identified, a descriptor bridges raw data and predictive models. This significantly accelerates the search for novel materials with targeted functional or multifunctional properties. Classic examples include the correlation between oxygen binding energy and catalytic activity for the oxygen reduction reaction [5], following the Sabatier principle, or descriptors of migration barrier heights in crystalline materials [6], which enabled the prediction of a novel oxide spinel framework with high Mg conductivity [7].

Artificial intelligence (AI) offers a powerful means to uncover complex correlations. However, conventional approaches often require large datasets and generate "black-box" models. This makes it difficult to identify the underlying mechanisms [4]. Symbolic regression (SR) [8] represents

a promising strategy to overcome these dual limitations. It directly addresses the interpretability challenge by uncovering nonlinear analytical expressions that link a target property to physically meaningful input parameters. Crucially, in contrast to data-hungry AI models, SR is capable of this discovery even from small datasets, making it uniquely suited for scientific domains. Using genetic programming, SR constructs these expressions from combinations of input parameters with mathematical operators.

Despite its promise, the practical application of SR faces a trade-off between flexibility and efficiency [9]. On the one hand, specialized methods such as the Sure-Independence-Screening-and-Sparsifying-Operator (SISSO) approach [10] were developed to identify analytical expressions using compressed sensing [11]. Recent advances, including SISSO++ [12], improve interpretability by incorporating controlled nonlinear optimization. Nevertheless, SISSO and related methods face limitations when applied to heterogeneous datasets or properties that exhibit subtle nonlinearities [4]. However, while more general SR methods are effective in generating highly complex and interpretable models, they are often computationally expensive. A systematic framework is needed to discover interpretable descriptors in broad chemical spaces.

Here we introduce DISCOVER (Data-Informed Symbolic Combination of Operators for Variable Equation Regression), an open-source symbolic regression framework with a modular, physics-informed design that balances flexibility with computational efficiency [13]. By applying DISCOVER, we identify interpretable descriptors that reveal governing physical principles and provide a clear path toward rational materials design.

To showcase the utility of DISCOVER, we targeted the spinel family of compounds, which are ideal candidates for next-generation battery applications [14–16]. However, research in this area is often fragmented between individual studies, preventing the systematic discovery of the overarching structure–property relationships. A centralized, consistently calculated, and open-access database is essential to unlock community-wide progress. Here we present an open-access database of 765 spinel compounds relevant to battery applications, including more than 200 unreported compositions, with consistently computed structural, electronic, and electrochemical properties obtained from periodic density functional theory (DFT) calculations. This database not only serves as a critical resource for the battery community, but also provides an ideal platform for applying DISCOVER to identify physically meaningful descriptors. Due to their structural and chemical versatility, spinels can host a variety of metal cations and facilitate rapid ion transport, motivating their exploration as cathodes [7,17], solid electrolytes [18–21], and protective coatings [22]. While this work focuses on spinel materials, the DISCOVER framework is general and can be extended to other material families, such as perovskites. Although the resulting descriptors may differ depending on the structural characteristics of each system, the methodology itself is broadly applicable for identifying interpretable descriptors across diverse materials.

The present study aims to transform spinel research from a fragmented trial-and-error process into a transparent, collaborative, and insight-driven endeavor by integrating the curated dataset with the analysis capabilities of DISCOVER. As illustrated in Figure 1, the Spinel Database and DISCOVER extend the traditional research cycle in materials science by coupling systematic data curation with descriptor-based analysis. This integration highlights how open-source tools can ac-

celerate hypothesis generation, deepen mechanistic insights, and streamline the discovery of high-performance materials. By ensuring FAIR data principles (Findable, Accessible, Interoperable, Reusable) [2], this approach promotes reproducibility and enables efficient data-driven discovery of energy materials. It also establishes a scalable, generalizable framework for the design of complex functional materials.

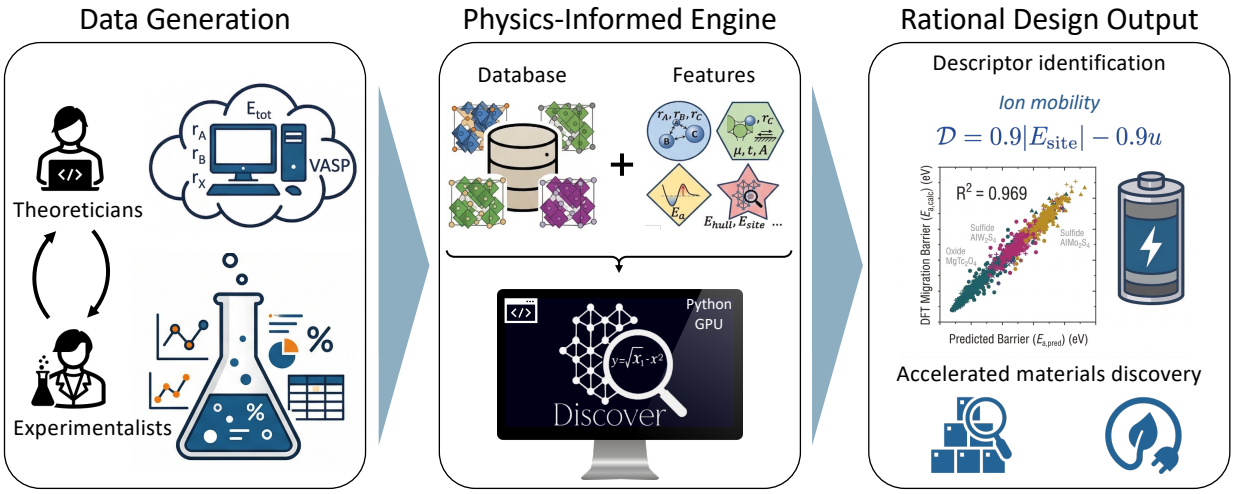

Figure 1: Expanding the research cycle with the Spinel Database and DISCOVER. The schematic contrasts the conventional materials science research cycle, where data are generated through theoretical and experimental studies to populate databases, with the enhanced cycle enabled by DISCOVER (Data-Informed Symbolic Combination of Operators for Variable Equation Regression). By identifying descriptors for target material properties and providing open-source tools for data-driven analysis, DISCOVER facilitates deeper insights, more efficient data integration, and accelerated discovery and design of functional materials.

Details of the Database

The dataset was constructed by generating ternary compounds of the form AB_2X_4 in the spinel crystal structure (space group $Fd\overline{3}m$), as illustrated in Figure 2. In this structure, A-site denotes monovalent, divalent, or trivalent cations occupying the tetrahedral 8a sites, B-site represents transition metal or lanthanide cations at the octahedral 16d sites, and the X anions correspond to oxide, sulfide, or selenide anions at the 32e positions. Each structure in the dataset was optimized using density-functional theory (DFT) [23,24] with the PBE exchange-correlation functional [25] and the Projector Augmented Wave (PAW) method [26], as implemented in the Vienna *Ab* initio Simulation Package (VASP) [27–29]. Full relaxations of the atomic positions and lattice parameters were performed to ensure that the dataset captures energetically stable geometries suitable for subsequent analysis.

To capture correlation effects in strongly localized $3d$ states, the DFT+U correction was applied following the Dudarev [30] approach when oxygen was present as the anion. Magnetic configurations were initialized differently depending on the chemical composition. For oxides containing $3d$ metals, antiferromagnetic (AFM) spin alignments were imposed, while sulfides and selenides were initialized in a ferromagnetic (FM) state. This distinction ensured robust exploration of competing spin orders relevant to strongly correlated systems.

We calculated the structural, electronic, and electrochemical properties relevant to battery performance for each compound in the database. Structural relaxation yielded total energies, lattice parameters, and atomic positions, while electronic analysis provided density of states (DOS),

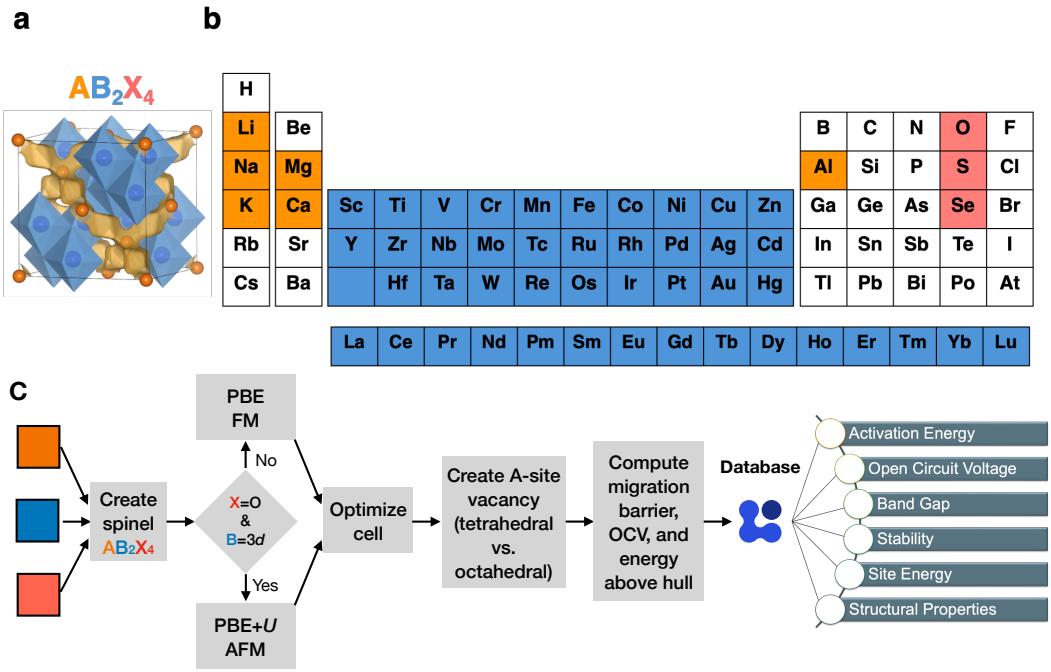

Figure 2: Spinel structure and composition. (**a**) Crystal structure of the spinel AB_2X_4 (space group $Fd\bar{3}m$). A-site cations (orange) occupy the tetrahedral 8a sites, while B-site cations (blue) reside in the octahedral 16d sites. X anions (red) are located at the 32e positions, forming a close-packed anion framework. The migration pathway of the charge carrier through the anion sublattice is illustrated in light orange. (**b**) Map of the elements that occupy the A, B, and X sites within the 765 compounds theoretically simulated as spinel. (**c**) Automated workflow for high-throughput DFT calculations of spinel compounds. Input files are generated with element-specific parameters, magnetic moments (including Ferro Magnetic (FM) and Antiferromagnetic (AFM) configurations), and U corrections for 3d transition metals in oxides. Optimized structures are subsequently used for vacancy studies, migration barrier analysis, and voltage calculations. For each compound, the workflow evaluates migration barriers, open-circuit voltage, band gap, stability, site-preference energy, and structural properties.

Fermi energy, and oxidation state estimates. The site-preference energy (E_{site}), derived from charge carriers occupying tetrahedral and octahedral sites, quantifies cation inversion tendencies. Activation energies for ionic migration (E_a), together with features such as energy above the hull (E_{hull}) and open-circuit voltage (OCV), offer a rigorous basis for assessing transport properties and stability. Figure 3 presents the distributions of four key properties within the spinel dataset, namely activation energy, site preference energy, lattice constant, and open-circuit voltage. The histograms, complemented by kernel density estimates, reveal the spread and characteristic trends across the 765 structures. Activation energies span a wide range with a broad peak, suggesting substantial variability in ionic transport characteristics. The site preference energy distribution is symmetric around zero, consistent with the energetic flexibility of cation arrangements within the spinel lattice. Lattice constants cluster tightly between 8–12 Å, reflecting the structural constraints of the spinel framework, while the open-circuit voltage distribution extends across positive and negative values, highlighting the wide electrochemical tunability of these compounds. Together, these distributions emphasize both the chemical richness and functional variability of spinel materials.

Structural and thermodynamic stability of spinels

The stability of spinel compounds within the database can be evaluated using two complementary approaches: the tolerance factor t [31] and the energy above the convex hull E_{hull}. These quantities capture the geometric and thermodynamic aspects of stability, respectively, and together provide a robust framework for identifying promising compounds.

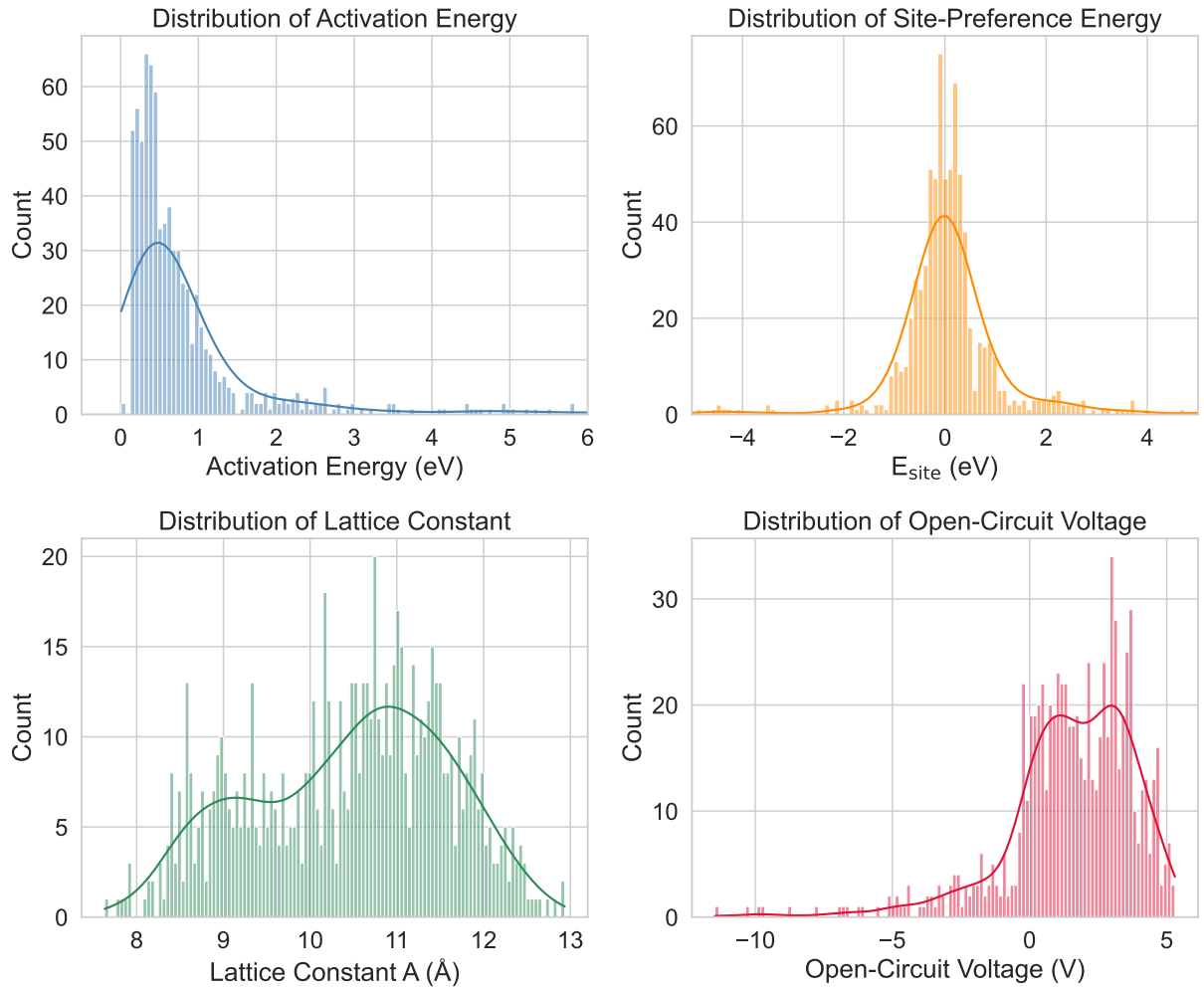

Figure 3: Distributions of key properties across the spinel database. The panels show activation energy (eV) for A-site charge carrier migration, site preference energy (eV), Lattice constant A (Å), and open-circuit voltage for all considered charge carriers A (V). Kernel density estimates are overlaid to highlight the underlying trends. These distributions illustrate the range and diversity of the 765 spinel structures.

The tolerance factor is a dimensionless parameter that describes the geometric compatibility of ionic radii in the spinel structure [31]. It can be expressed as follows:

$$t = \frac{\sqrt{3}}{2} \times \frac{(r_B + r_X)}{(r_A + r_X)}, \tag{1}$$

where r_A is the radius of the cation at the tetrahedral site, r_B is the radius of the cation at the octahedral site, and r_X is the radius of the anion. When $t < 1$, the structure is geometrically favorable and stable. In contrast, values of $t > 1$ suggest lattice strain and potential instability.

Thermodynamic stability is quantified using the energy above the convex hull. The convex hull represents the lowest energy set of phases within a given chemical space. For a compound in a given composition, the energy above the hull, E_{hull}, is the difference between its formation energy and the nearest point on the hull. If $E_{\text{hull}} = 0$, the compound is thermodynamically stable; if $E_{\text{hull}} > 0$, it is metastable or unstable. In practice, compounds with $E_{\text{hull}} < 0.05$ eV/atom are considered thermodynamically stable, while those with $E_{\text{hull}} < 0.1$ eV/atom (but > 0.05 eV/atom) are often considered metastable but experimentally accessible due to entropic effects and kinetic stabilization.

In high-throughput computational screening, both descriptors are applied simultaneously. Compounds that satisfy dual criteria: $t < 1$ and $E_{\text{hull}} < 0.1$ eV/atom are retained as stable candidates. This combined filtering ensures that the database focuses on materials that are both geometrically compatible with the spinel lattice and thermodynamically robust against decomposition.

To visualize the stability landscape of the database, the energy above the convex hull (E_{hull})

was plotted against the tolerance factor (t) in Figure 4. The plot highlights the stability criteria defined above, showing the region of compounds that satisfy the combined requirements for geometric and thermodynamic stability. Shaded areas indicate metastable (0.05–0.1 eV/atom) and unlikely (> 0.1 eV/atom) regimes, as well as the geometric instability region where $t > 1$. Data points are color-coded according to the first element in the chemical formula, enabling a comparison of different chemistries within the stability window.

By applying these criteria, the database is refined to include only stable compounds, thereby increasing the reliability of predictions and aligning the methodology with standards commonly reported in high-impact materials science literature. In the present case, the filtering reduced the dataset from 765 compounds to 496 stable entries.

DISCOVER: A symbolic regression framework for descriptor analysis

To complement the database and enable a deeper understanding of structure–property relationships, we developed a symbolic regression framework named DISCOVER (Data-Informed Symbolic Combination of Operators for Variable Equation Regression) [13]. DISCOVER is a data-driven, physics-informed tool designed to extract interpretable descriptors that correlate strongly with target properties, such as voltage, stability, and ion mobility. The framework is constructed to identify parsimonious expressions that capture essential property trends while balancing accuracy and interpretability.

While conceptually inspired by related compressed-sensing methods like SISSO [10,12], DIS-

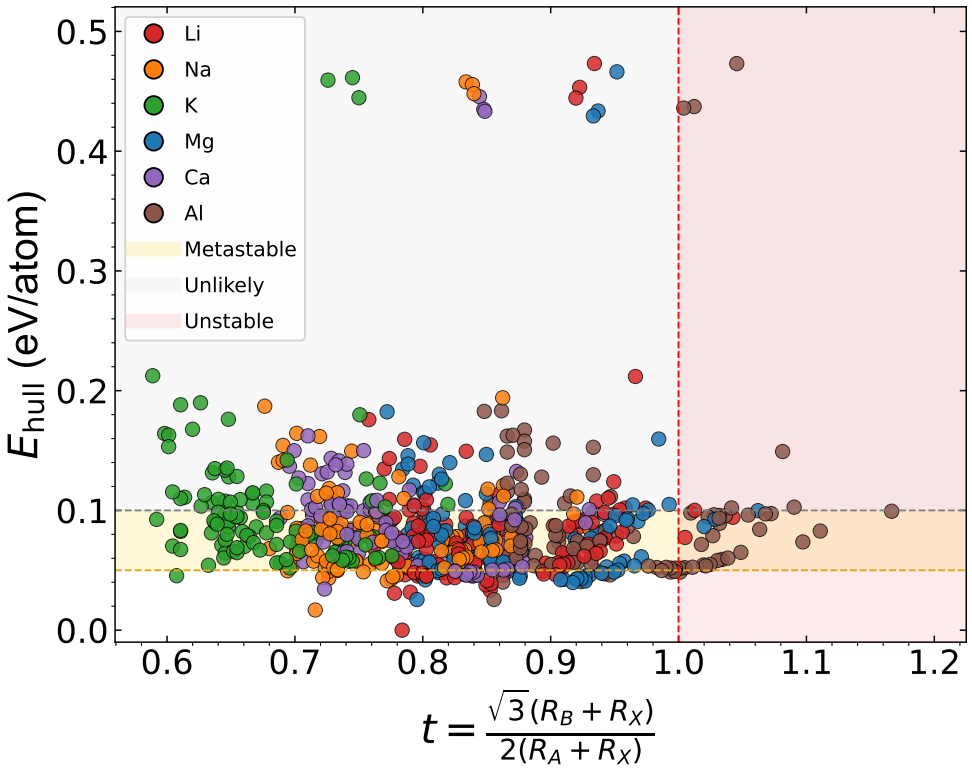

Figure 4: Energy above the convex hull (E_{hull}) versus tolerance factor (t) for all compounds in the database. Stable compounds are identified in the region $t < 1$ and $E_{\text{hull}} < 0.05$ eV/atom. Shaded areas indicate metastable (yellow), unlikely (gray), and geometrically unstable (red) regions. Data points are color-coded by the first element of the chemical formula (Li, Na, K, Mg, Ca, Al).

COVER introduces several key innovations to enhance flexibility, search efficiency, and physical rigor. A central advancement is the expansion of the symbolic search space. DISCOVER utilizes a broader operator set and, crucially, offers improved flexibility in operator nesting and expression tree growth. This allows the framework to uncover more complex and nonlinear functional forms than those accessible by methods limited to simpler operator combinations.

The workflow, illustrated in Figure 5, generates this vast pool of candidate features and then implements a hierarchy of exact, approximate, and heuristic search algorithms to identify the optimal, sparse descriptor. This hybrid search strategy includes fast greedy algorithms like Orthogonal Matching Pursuit (OMP) for initial exploration, metaheuristics to escape local minima, and mathematically exact Mixed-Integer Quadratic Programming (MIQP) solvers for manageable problem sizes. This entire high-throughput exploration of millions of candidate models is made computationally feasible through GPU parallelization.

A further distinguishing innovation is DISCOVER's flexible system for defining feature-space constraints. This allows users to embed *a priori* physical knowledge directly into the search process. This explicit enforcement of physical constraints—such as dimensional consistency, known symmetries, shape constraints (e.g., monotonicity), and other compositionality rules—acts as a physics-informed filter. This pruning ensures that the discovered descriptors not only fit the data but also respect the governing laws of the underlying system. The final output is a set of simple, interpretable descriptors that illuminate structure–property relationships for targeted materials design.

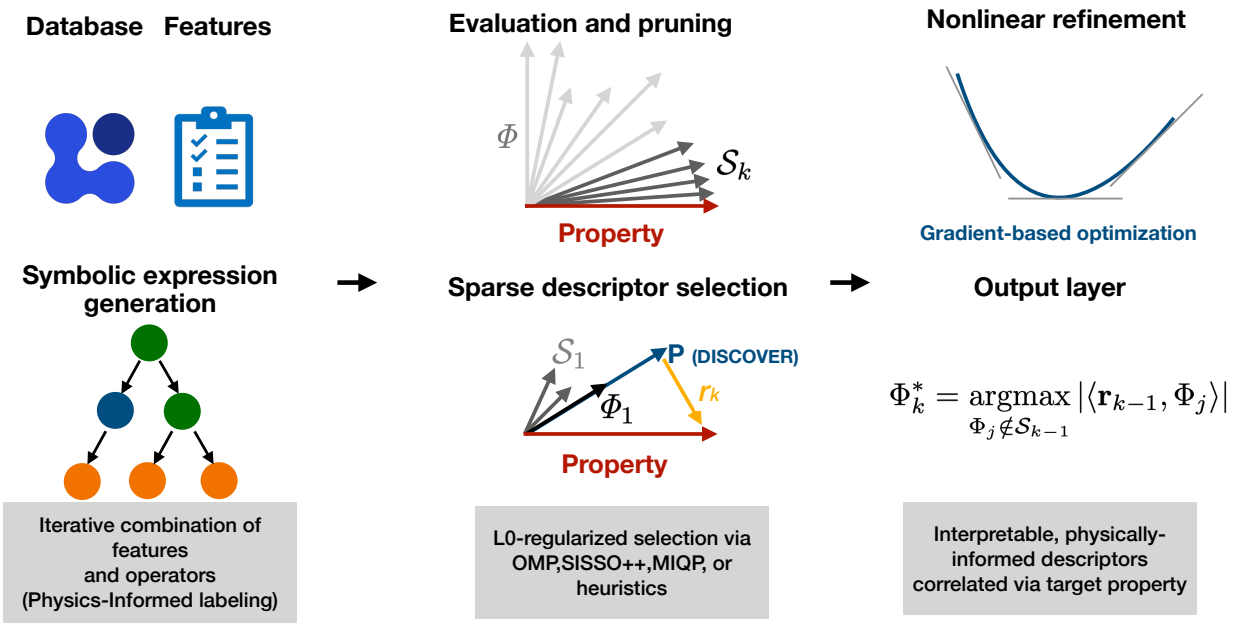

Figure 5: Workflow of the DISCOVER symbolic regression framework. The schematic illustrates how DISCOVER generates interpretable descriptors from a materials database. Primitive input features (structural, compositional, and electronic) are combined with a curated operator library under physics-informed constraints to produce a vast space of candidate symbolic expressions. These expressions are then iteratively evaluated, pruned for physical consistency, and optimized using sparse selection methods to identify a low-dimensional descriptor. GPU acceleration enables an efficient search through millions of candidates. The final output is a compact and physically meaningful equation that correlates strongly with the target property, facilitating data-driven materials design.

DISCOVER Reveals a Universal Descriptor for Ionic Mobility

A central goal in designing high-performance solid-state batteries is to achieve fast ion conduction. This requires a deep understanding of the factors that govern the ionic migration barrier (E_a). For instance, prior descriptors, including our own previous work , have successfully correlated migration barriers with fundamental, pre-tabulated atomic properties such as ionic radii and electronegativity. While valuable for initial screening, those elemental-based models can struggle to capture the complex, system-specific interplay of thermodynamics and structure.

Here, we apply the DISCOVER framework to our comprehensive spinel database to move beyond specific cases and uncover the fundamental physical principles governing ionic mobility. By searching through millions of candidate expressions, DISCOVER identified a remarkably simple and linear relationship that connects the migration barrier to just two fundamental physical quantities: the thermodynamic site-preference energy (E_{site}) and the structural anion positional parameter (u). This relationship reveals that ionic mobility in spinels is controlled by a direct competition between thermodynamics and structure. Unlike previous descriptors that rely on elemental properties like ionic radii, our new physically transparent metric, **Spinel Mobility Descriptor ($\mathcal{D}$)** incorporates DFT-derived properties central to this competition. We formulate this insight as:

$$\mathcal{D} = 0.9|E_{\text{site}}| - 0.9u \tag{2}$$

Here, E_{site} is given in electronvolts (eV) and is defined as the energy difference between the tetrahedral and octahedral sites, $E_{\text{site}} = E_{\text{tet}} - E_{\text{oct}}$. It represents the thermodynamic preference of the ion for occupying the tetrahedral site over the octahedral site. The anion positional parameter u is a

dimensionless quantity describing the deviation of the anion positions from the ideal close-packed arrangement (as defined in Ref. [32]). Consequently, the descriptor $\mathcal{D}$ carries units of energy (eV). The numerical prefactors in Eq. (2) are obtained from the symbolic regression procedure and are therefore dependent on the chosen unit system.

The descriptor encapsulates the core physics of ion migration. The first term, $|E_{\text{site}}|$, represents the thermodynamic penalty for site destabilization. A large energy difference between these sites naturally leads to a higher migration barrier. The second term, involving the anion positional parameter u, quantifies the geometric bottleneck of the migration path. An optimal value of u corresponds to the widest possible channel for the ion to pass through, minimizing the kinetic barrier.

Our analysis reveals that the migration barrier, E_a, exhibits a strong linear correlation with this descriptor, $\mathcal{D}$, that holds universally across oxide, sulfide, and selenide spinels. This discovery provides a unifying principle for ionic mobility that was not apparent from studying individual material classes. It demonstrates that regardless of the specific chemistry, the migration barrier is fundamentally governed by the balance between stabilizing the migrating ion's intermediate position ($|E_{\text{site}}|$) and opening the structural pathway for its transit (u). This insight offers a clear and actionable strategy for materials design: to engineer a spinel with superior ionic conductivity, one must simultaneously tune the composition to minimize $|E_{\text{site}}|$ and optimize the lattice geometry to achieve a favorable anion parameter u.

In our previous work, we established that ion mobility in crystalline battery materials can be

effectively captured by the migration number, N_{migr} [6,33],

$$N_{\mathrm{migr}}^{\mathrm{AX}} = (r_{\mathrm{A}} + r_{\mathrm{X}}) n_{\mathrm{A}} n_{\mathrm{X}} \Delta\chi_{\mathrm{AX}}^{2} \, , \tag{3}$$

where r_i are the ionic radii of the migrating cation A and the counter anion X, n_{i} are the absolute values of their formal oxidation states, and $\Delta\chi_{\mathrm{AX}}^{2}$ is the squared electronegativity difference between the two species. While N_{migr} successfully yields linear scaling relations based entirely on fundamental elemental properties, it requires distinct scaling lines when varying the transition metal (B-site) host in ternary spinels. Transitioning across the immense structural and chemical versatility of the AB_2X_4 spinel family introduces competing thermodynamic and geometric constraints that elemental properties alone cannot universally resolve.

To capture universal transport behavior across diverse oxide, sulfide, and selenide spinels, the DISCOVER framework identifies a descriptor ($\mathcal{D}$) that operates at a higher level of physical abstraction. Rather than contradicting N_{migr}, this new formulation naturally extends our predictive capability into complex ternary spaces. The thermodynamic site preference, $|E_{\mathrm{site}}|$, intrinsically correlates with the maximum binding energy in the host material [34], effectively representing the reaction energy upon intercalation. This directly maps onto Brønsted-Evans-Polanyi (BEP) scaling relations, which link reaction energies to activation energies (diffusion barriers) [3]. Furthermore, the anion positional parameter, u, explicitly encodes the geometric constraints of the migration bottleneck, an effect previously captured implicitly via ionic radii in N_{migr}. Ultimately, the shift from N_{migr} to the DISCOVER-derived $\mathcal{D}$ represents an evolution from element-specific scaling to a generalized physical law that natively unites the complex spinel family.

Refined Models and Descriptors for Additional Properties While the descriptor $\mathcal{D}$ captures the essential physics of ion migration, DISCOVER can produce more complex symbolic expressions for even higher quantitative accuracy. The fully optimized descriptor for the migration barrier is built upon this core physical insight by including a constant offset and a higher-order correction term based on system energies:

$$E_{a,\text{predict}} = \underbrace{-0.9\,u + 0.9\,|E_{\text{site}}|}_{\approx\mathcal{D}} + 0.38 \;-\; 0.09\,\frac{E_{\text{oct}}\,(E_{\text{metal}} - |E_{\text{site}}|)}{E_{\text{tot}}}$$

where $E_{\text{oct}}, E_{\text{metal}}, E_{\text{tot}}$ denote site- or structure-specific energies derived from DFT. This refined model achieves excellent agreement with the reference DFT calculations, as shown in the parity plot in Figure 6. All energies are expressed in eV and all distances in Å. Input features are consistently defined within this unit system, and the symbolic regression coefficients are determined accordingly to ensure dimensional consistency of the resulting expressions.

The framework's versatility was further demonstrated by identifying accurate, physically interpretable descriptors for the lattice constant (A) and open-circuit voltage (OCV). The discovered expressions are:

$$A_{\text{predict}} = 1.59\,\sqrt{R_{AX}} \cdot R_{BX} + 4.03$$

$$\text{OCV}_{\text{predict}} = -0.75\,\frac{r_{\text{ion}} \cdot (E_{\text{tot}} - E_{\text{tet}})}{\sqrt{\text{oxidation state}}} - 1.01 - 0.11\,\frac{(E_{\text{tot}} - E_{\text{tet}})}{r_{\text{ion}}}$$

where R_{AX} and R_{BX} are the cation–anion bond lengths and r_{ion} is the ionic radius of the active metal. These formulas automatically recover compact relationships that capture key nonlinear structure–property trends. For example, the lattice constant is primarily governed by cation–anion bond lengths, while the OCV depends strongly on the ionic radius, oxidation state, and relative

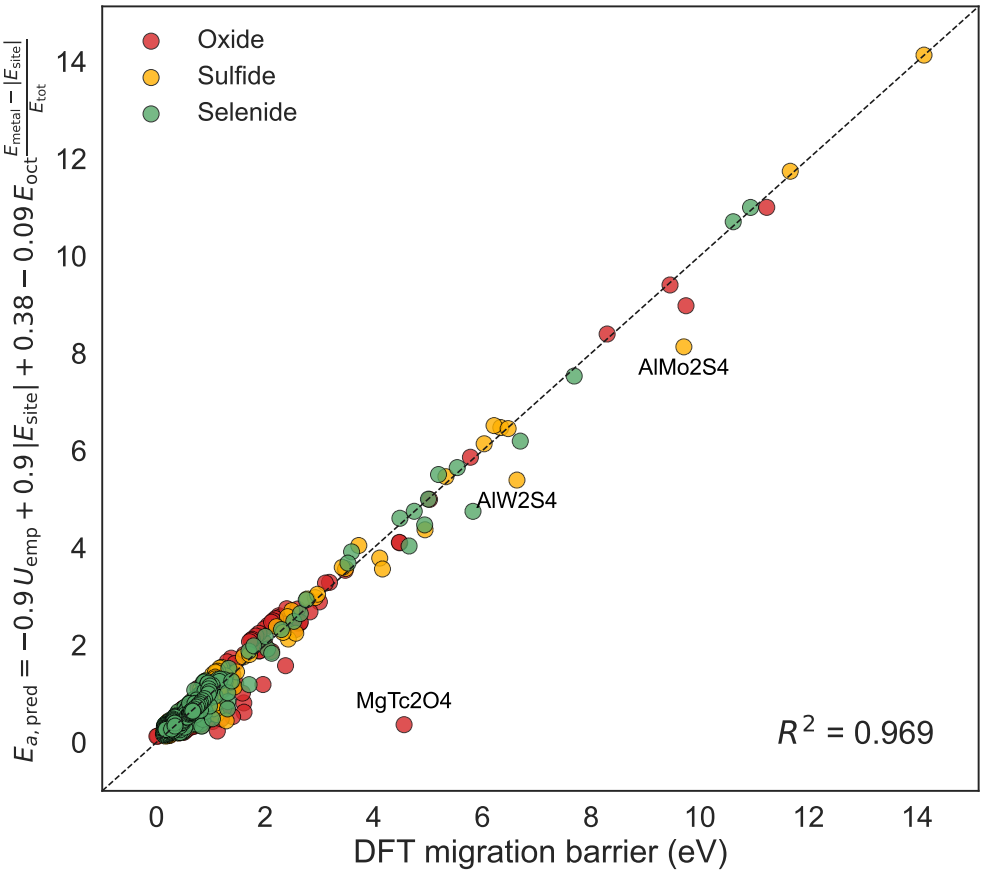

Figure 6: Migration barrier E_a predicted by the full DISCOVER model versus DFT-computed values. Data points are color-coded by chemical type. The dashed line indicates the ideal $y = x$ relationship. The high accuracy ($R^2 = 0.97$) shows the effectiveness of the refined symbolic expression.

system energies. Parity plots in Figures 7 and 8 confirm that these descriptors capture the trends of DFT-calculated properties with high fidelity. The overall low prediction errors for all three properties, shown in Figure 9, underscore the framework's capability to generate reliable and insightful models for diverse material properties.

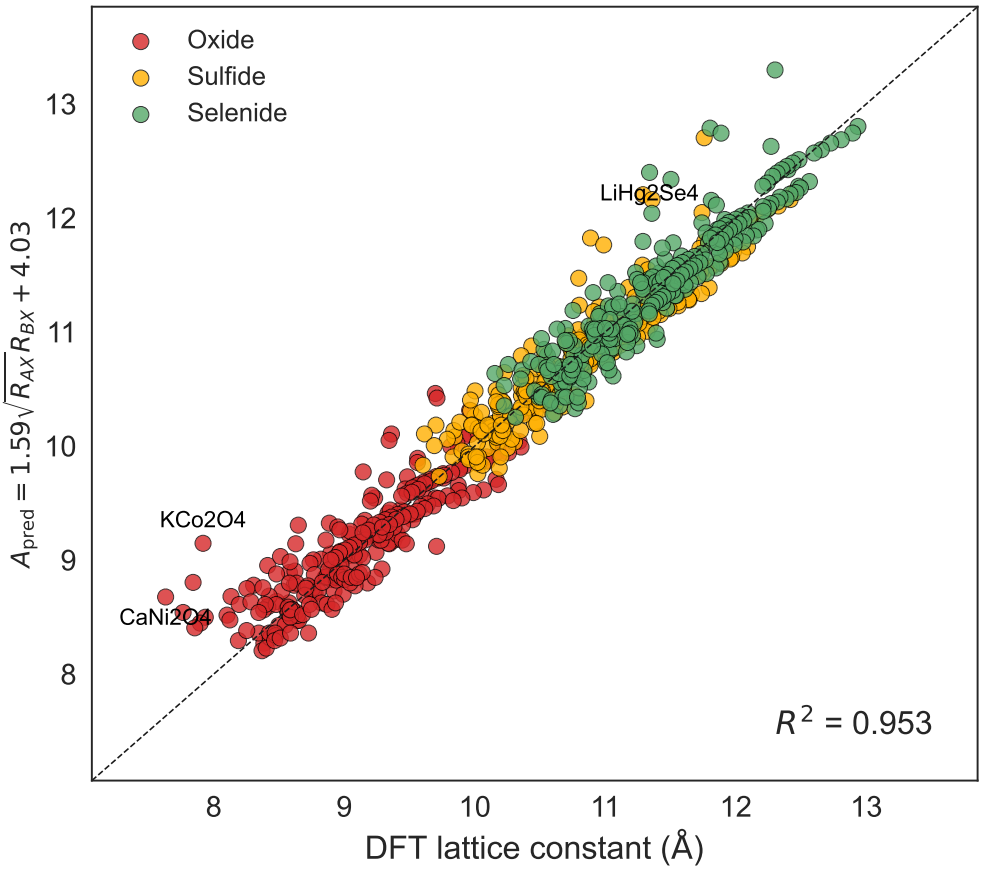

Figure 7: Lattice constant A predicted by DISCOVER versus DFT reference values. Data points are color-coded by chemical type: oxide (crimson), sulfide (gold), and selenide (yellow-green). The dashed line indicates the ideal $y = x$ relationship. Quantitative accuracy is reported using R^2, RMSE, and MAE.

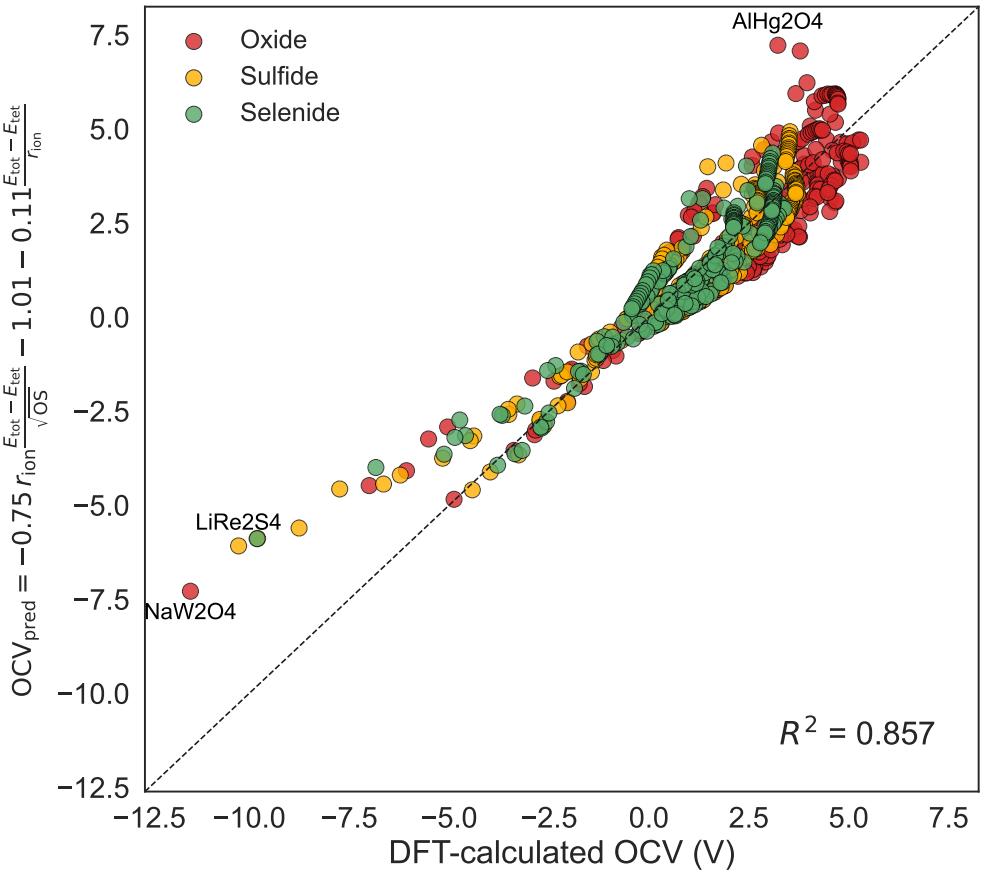

Figure 8: Open-circuit voltage (OCV) predicted by DISCOVER versus DFT reference values. Data points are color-coded by chemical type. The dashed line indicates the ideal $y = x$ relationship. Quantitative accuracy is reported using R^2, RMSE, and MAE.

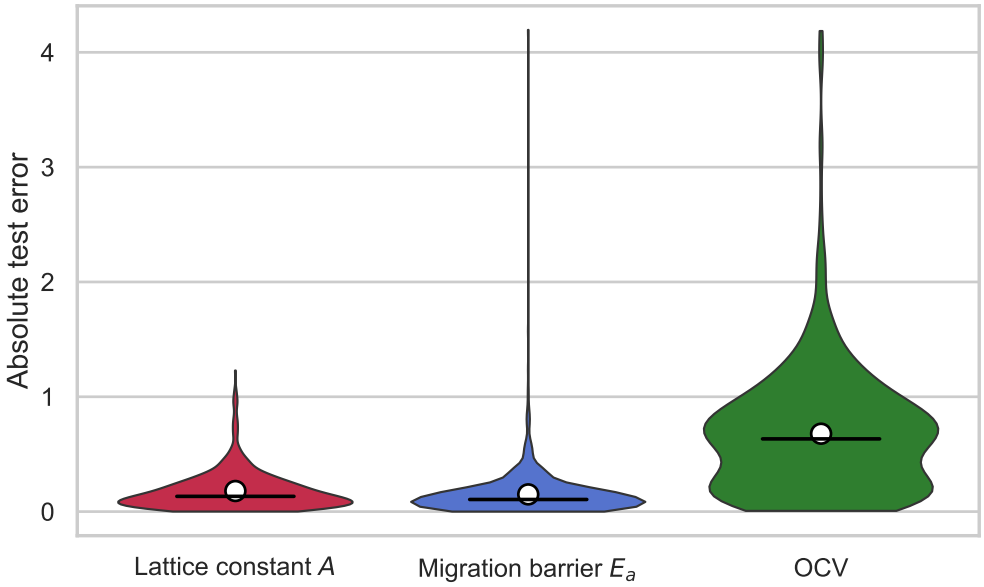

Figure 9: Absolute test error distributions of DISCOVER-predicted descriptors. Violin plots show the distribution of absolute errors between DISCOVER predictions and DFT reference values for the lattice constant A, migration barrier E_a, and open-circuit voltage (OCV). The white circles indicate the mean absolute error (MAE) and the black horizontal lines indicate the median absolute error (MedAE). This figure highlights the high accuracy and reliability of the predictions across the spinel dataset.

Selected compounds for battery applications

Compounds were screened using key descriptors including the activation energy (E_a), open-circuit voltage (OCV), energy above the hull (E_{hull}), and tolerance factor (t). Electronic configuration was also taken into account. Closed-shell compounds, expected to be highly stable with negligible electronic conductivity, were identified as potential solid electrolytes (OCV is not applicable). Open-shell compounds, capable of undergoing redox reactions, were designated as candidate cathode materials, with their OCV values reported.

Potential Solid Electrolytes (Closed-Shell Compounds): Table 1 summarizes the closed-shell compounds identified as promising solid electrolytes. These materials are characterized by low activation energies (E_a = 0.31–0.40 eV), indicating high ionic mobility, and small energies above the hull (E_{hull} = 0.040–0.048 eV), reflecting excellent thermodynamic stability. Their tolerance factors (t = 0.908–0.949) lie within the range expected for stable spinel structures, supporting structural robustness. Among the candidates, $MgNd_2Se_4$ stands out with the lowest activation energy (0.310 eV) and a low hull energy (0.048 eV), making it a particularly promising solid electrolyte for Mg-ion conductors. The full list includes 15 Mg-based sulfide and selenide spinels, providing a diverse set of materials for potential solid-state battery applications.

Cathode Materials (Open-Shell Compounds): Table 2 lists the open-shell compounds identified as promising cathode materials. These compounds exhibit moderate activation energies

Table 1: Closed-shell compounds proposed as solid electrolytes.

Compound	E_a **(eV)**	E_{hull} **(eV)**	t
$MgSc_2S_4$	0.377	0.045	0.876
MgY_2S_4	0.385	0.043	0.927
$MgDy_2S_4$	0.358	0.046	0.930
$MgHo_2S_4$	0.364	0.045	0.927
$MgEr_2S_4$	0.378	0.044	0.924
$MgTm_2S_4$	0.384	0.042	0.920
$MgLu_2S_4$	0.397	0.042	0.910
MgY_2Se_4	0.330	0.041	0.924
$MgNd_2Se_4$	0.310	0.048	0.949
$MgSm_2Se_4$	0.312	0.048	0.940
$MgDy_2Se_4$	0.327	0.041	0.927
$MgHo_2Se_4$	0.329	0.043	0.924
$MgEr_2Se_4$	0.335	0.043	0.921
$MgTm_2Se_4$	0.336	0.040	0.917
$MgLu_2Se_4$	0.343	0.042	0.908

(E_a = 0.170–0.405 eV), suggesting favorable ionic transport, and open-circuit voltages (OCV = 2.07–3.09 V) consistent with practical energy storage applications. Their energies above the hull (E_{hull} = 0.043–0.048 eV) indicate high thermodynamic stability, while tolerance factors (t = 0.726–0.904) are compatible with stable spinel structures. Among the candidates, $LiTm_2Se_4$ emerges as a top performer, combining a low activation energy (0.170 eV) with a high OCV (3.09 V) and favorable structural parameters, making it a particularly attractive cathode material. The dataset encompasses Li- and Na-based oxides, sulfides, and selenides, providing a diverse set of chemically tunable compounds for further experimental and computational exploration.

Future Expansion of the Database

The ambition of the Spinel Battery Materials Database Project is not only to compile a consistent dataset of high-quality DFT-computed properties for existing spinel compounds but also to continuously incorporate new materials, computed descriptors, and experimental annotations into an open, extensible repository. Our long-term vision is to establish a new standard for disseminating spinel-related materials data—an evolving, community-driven resource that serves as the equivalent of a "Wikipedia" for spinel structures in energy storage applications. Achieving this vision requires broad participation from the battery materials community, and a cultural shift toward the regular sharing of computational and experimental results through structured, machine-readable formats. We envision a future where researchers routinely upload their new spinel data to the platform as a complementary practice to traditional publication.

Table 2: Open-shell compounds proposed as cathode materials.

Compound	E_a **(eV)**	**OCV (V)**	E_{hull} **(eV)**	t
$LiPd_2O_4$	0.311	2.835	0.049	0.814
$NaPd_2O_4$	0.181	2.716	0.043	0.726
$LiCr_2S_4$	0.213	2.527	0.048	0.816
$LiRh_2S_4$	0.365	2.712	0.045	0.796
$LiIr_2S_4$	0.405	2.591	0.045	0.799
$LiCr_2Se_4$	0.198	2.207	0.044	0.819
$LiRh_2Se_4$	0.233	2.160	0.048	0.800
$LiIr_2Se_4$	0.360	2.067	0.046	0.803
$LiTm_2Se_4$	0.170	3.089	0.046	0.904
$LiLu_2Se_4$	0.176	3.075	0.047	0.894

We acknowledge that contributing data requires a non-trivial investment of time and effort. To ensure sustained engagement, the Spinel Database Project must provide high utility, ease of use, visibility, longevity, and scientific credibility. In terms of utility, the examples and analyses presented in this work, alongside our interactive web interface and descriptor tools, are intended to demonstrate the advantages of aggregated, high-quality datasets that adhere to the FAIR (Findable, Accessible, Interoperable, Reusable) data principles. Uploading one's own data provides tangible benefits. It is a contribution to the broader scientific community. It increases the visibility and impact of associated publications. It also fulfills growing mandates for open science from funding agencies and journals. Finally, it helps accelerate materials discovery by reducing redundancy and enabling cross-study comparison.

To reduce the barrier to entry, we provide user-friendly, well-documented protocols for data formatting, cleaning, and submission. Our current backend supports data ingestion through a standardized Excel template that is simple, transparent, and adaptable to laboratory-specific needs. Scripts for parsing raw DFT output and automatically populating the database template are available in Python and openly hosted on GitHub. These tools enable rapid conversion of internal datasets into upload-ready formats and offer compatibility with custom workflows and laboratory information management systems (LIMS). Furthermore, the backend infrastructure is built to accommodate diverse data types, including both computed properties and validated experimental measurements.

A key component of future expansion is our incorporation of the DISCOVER (Data-Informed

Symbolic Combination of Operators for Variable Equation Regression) framework into our modeling workflow. DISCOVER enables the derivation of physically meaningful, symbolic descriptors from large pools of primary variables, offering compact, interpretable equations that correlate with target properties such as ionic mobility, open-circuit voltage, and redox stability. While conceptually inspired by SISSO (Sure Independence Screening and Sparsifying Operator), DISCOVER expands the symbolic search space with a broader operator set, introduces physically constrained expression filters, and offers improved flexibility in operator nesting and expression tree growth. We anticipate DISCOVER will become a cornerstone tool for future property modeling and descriptor benchmarking on this platform.

Our goal is to make uploading spinel data to this database a standard practice that strengthens the visibility and reproducibility of computational and experimental work. In the long term, we aim to collaborate with academic publishers, professional societies, and data repositories to embed database integration into the materials publishing workflow. Publicly accessible data not only validates reported results but also enables broader pattern recognition and anomaly detection, such as identifying performance values that deviate from theoretical voltage limits or activation energy distributions.

To ensure long-term availability and trust, we have secured institutional support to maintain the database, web interface, and associated repositories for the coming decade, with an option for extension. The open-source nature of the platform guarantees transparency, encourages community contributions, and allows for resilience in the face of technical disruption.

The database is designed for extensibility and can be readily adapted to related materials, including disordered spinels, doped derivatives, or non-oxide anion chemistries (e.g., sulfides and fluorides). We actively encourage community-led initiatives in this direction. One of the fundamental issues this project seeks to address is the fragmentation of valuable data across the literature in inconsistent and often inaccessible formats. A related challenge is the permanent loss of data not reported in publications—often representing a vast majority of computed or synthesized candidates. By facilitating the collection and standardized reporting of both successful and unsuccessful compounds, we aim to reduce bias and enable a more complete, data-rich picture of the spinel materials landscape.

Conclusions

In this Spinel Battery Materials Project, we have developed an open-access database containing 765 spinel compounds relevant to battery applications, all computed using consistent density functional theory (DFT) protocols. Alongside the database, we provide an open-source descriptor analysis framework that enables interactive exploration and interpretation of structural, electronic, and electrochemical properties. Together, these resources aim to address long-standing challenges in data accessibility, standardization, and reusability in the battery materials community.

We have demonstrated the utility of this framework through example analyses that reveal trends and correlations between key descriptors and battery-relevant performance metrics. These insights enable more targeted materials screening and hypothesis generation, supporting a transi-

tion from trial-and-error approaches to data-driven discovery. While demonstrated here for spinels, the framework is readily transferable to other materials classes, enabling similar descriptor-driven insights across broader chemical spaces.

Our aim is that this work will foster a culture of open data sharing and reproducible computational workflows in the field of energy storage materials. With greater community engagement, the database can expand to include thousands more spinel and related compounds, forming a high-resolution data mesh for battery materials. In turn, this will unlock the potential of machine learning and AI-guided design to accelerate innovation in battery technologies and help meet global energy demands.

Methods

Computational details. First-principles calculations were performed within the framework of density functional theory (DFT) [23,24] for ternary AB_2X_4 spinel compounds. Exchange–correlation effects were treated using the generalized gradient approximation (GGA) in the Perdew–Burke–Ernzerhof (PBE) formulation [25], and the projector augmented wave (PAW) method [26] as implemented in the Vienna *Ab-initio* Simulation Package (VASP) [27–29].

A plane-wave cutoff energy of 520 eV was employed, and electronic self-consistency was achieved to within 1×10^{-5} eV per supercell. Structural optimizations were performed for the $Fd\bar{3}m$ spinel structure with all atomic positions relaxed without symmetry constraints until the forces were below 0.01 eV Å^{-1}. A $2 \times 2 \times 2$ k-point mesh was used for the unit cell containing

eight formula units.

Strong correlations among the transition-metal $3d$ electrons were accounted for using the Hubbard U approach [30] with the following values: $U_{\mathrm{Ti}} = 3.00$ eV, $U_{\mathrm{V}} = 3.25$ eV, $U_{\mathrm{Cr}} = 3.70$ eV, $U_{\mathrm{Mn}} = 3.90$ eV, $U_{\mathrm{Fe}} = 5.30$ eV, $U_{\mathrm{Co}} = 3.32$ eV, and $U_{\mathrm{Ni}} = 6.20$ eV [35]. Magnetic configurations were treated using spin-polarized calculations (ISPIN $= 2$) with initial magnetic moments set to either ferromagnetic or antiferromagnetic alignments depending on the compound.

Ion migration barriers were evaluated using the climbing-image nudged elastic band (CI-NEB) method [36]. To model ionic diffusion we considered vacancy-mediated hops: one A-site atom was removed ($A_{0.875}B_2X_4$), producing a single A-vacancy per cell. We computed the migration barrier using the climbing-image NEB method with four intermediate images between tetrahedral and octahedral positions [37]. Energies were calculated using both PBE and PBE+U to validate the migration barriers. All NEB images were relaxed until residual forces were below 0.05 eV Å^{-1}.

FAIR-compliant Data Management and Dissemination. To ensure the long-term impact and utility of our work, all generated data were curated and disseminated following the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles [38]. We have leveraged the NOMAD (Novel Materials Discovery) Repository [39], a community-driven infrastructure for materials science data, as the primary platform for storage and sharing [40].

A key challenge in creating a FAIR dataset is handling complex, multi-step calculations. To address this, we utilized the NOMAD NEB Workflow plugin (`nomad-neb-workflow`) [41], which standardizes the representation of nudged elastic band (NEB) workflows within NOMAD.

While the VASP parser in NOMAD is responsible for reading and interpreting the raw VASP output files, the NEB plugin automatically organizes these parsed results into a coherent workflow representation. This enables the consistent extraction and storage of overall NEB results, such as migration barriers, reaction pathways, and structural details of intermediate images. The plugin was triggered through the automated workflow connections established between individual VASP calculations corresponding to each NEB image, allowing the complete pathway to be recognized and processed as a single linked workflow.

This specialized tool is part of a broader data management strategy. The general schema captures not only the final computed properties (e.g., voltages, lattice parameters) but also the full computational provenance, including DFT settings, pseudopotentials, and convergence criteria. Each dataset was curated and mapped into a standardized format before being uploaded to NOMAD, ensuring that all data are both human- and machine-readable. Within NOMAD, each dataset (rather than individual entries) is assigned a persistent digital object identifier (DOI), guaranteeing findability and stable referencing. This structured approach, combining general curation with workflow-specific tools, transforms our dataset from a static collection of results into a dynamic, extensible, and community-ready resource designed to accelerate discovery in battery materials.

Energy above the hull calculations. Thermodynamic stability of 765 spinel compounds was assessed via the energy above the convex hull (E_{hull}), computed from DFT (GGA-PBE) total energies. Competing phases in each chemical system were obtained from the *Materials Project* database using `pymatgen`. For systems lacking entries, E_{hull} was approximated relative to the elemental phases. Energies are reported per atom by dividing the energy above the hull per formula

unit by the number of atoms, enabling consistent comparison across different stoichiometries.

Symbolic regression framework details. The symbolic regression task in DISCOVER is formally cast as an L_0-regularized least-squares regression problem. Given a property vector $\mathbf{y}$ and a set of M candidate symbolic features $\Phi = \{\Phi_1, \Phi_2, \ldots, \Phi_M\}$, the objective is to identify a sparse linear model with at most D features:

$$\min_{\beta} \|\mathbf{y} - \Phi\beta\|_2^2 \quad \text{subject to} \quad \|\beta\|_0 \leq D, \tag{4}$$

where $\|\beta\|_0$ denotes the number of nonzero coefficients. This sparsity constraint ensures that the resulting descriptors are compact and interpretable. Since the problem is NP-hard, DISCOVER implements a hierarchy of exact, approximate, and heuristic algorithms to navigate the large search space.

Brute-force strategies can evaluate all $\binom{M}{D}$ combinations to guarantee optimality, but they become impractical for large M or D. Greedy algorithms provide computationally efficient alternatives by sequentially adding features that most reduce the residual error. Orthogonal Matching Pursuit (OMP) improves upon this approach by re-fitting all coefficients at each step, solving a least-squares problem

$$\beta_k = \Phi_{\mathcal{S}_k}^{\dagger}\mathbf{y}, \quad \mathbf{r}_k = \mathbf{y} - \Phi_{\mathcal{S}_k}\beta_k, \tag{5}$$

where $\mathcal{S}_k$ is the active set of features and $\mathbf{r}_k$ is the residual. This refinement helps avoid

redundancy among correlated descriptors.

For large-scale problems, DISCOVER integrates efficient breadth-first search strategies. Here, linear-algebraic updates based on QR decomposition allow rapid evaluation of residual sum of squares (RSS) reductions without refitting full models. If Q_D is the orthonormal basis spanning the current feature set, the orthogonal component of a new candidate feature $\mathbf{x}_{\text{new}}$ is

$$\mathbf{w}_{\text{new}} = \mathbf{x}_{\text{new}} - Q_D Q_D^T \mathbf{x}_{\text{new}}, \tag{6}$$

and the corresponding gain in model accuracy is

$$\Delta\text{RSS} = \frac{(\mathbf{r}_D^T \mathbf{w}_{\text{new}})^2}{\|\mathbf{w}_{\text{new}}\|_2^2}. \tag{7}$$

This procedure enables efficient exploration of millions of candidate models, particularly when accelerated on GPUs. To further enhance exploration, DISCOVER employs stochastic metaheuristics such as Random Mutation Hill Climbing and Simulated Annealing, which allow occasional acceptance of worse solutions to escape local minima. When problem sizes are manageable, Mixed-Integer Quadratic Programming (MIQP) provides mathematically exact solutions by coupling continuous regression coefficients with binary selection variables that enforce sparsity.

A central innovation of DISCOVER is its flexible system for defining feature-space constraints. Users can enforce shape constraints (e.g., monotonicity), dimensional consistency, operator restrictions, and variable interdependence rules. Such physics-informed pruning ensures

that discovered descriptors not only reproduce observed data but also respect governing physical laws. Finally, once a sparse descriptor is identified, DISCOVER offers nonlinear refinement by parameterizing symbolic expressions (e.g., Φ_i^p, $\exp(-p\Phi_i)$) and optimizing both linear coefficients and nonlinear parameters with gradient-based algorithms such as L-BFGS-B. This step extends the framework beyond linear combinations, enabling the discovery of richer functional forms.

Machine-learning approach. To identify interpretable descriptors for battery-relevant properties, we employed the DISCOVER framework using the dataset of 765 spinel compounds. The input features were constructed from tabulated structural, electronic, and chemical quantities obtained from DFT calculations. Non-feature identifiers such as chemical formulae were excluded from the training space.

Feature construction was carried out up to depth 2, using unary and binary mathematical operators including addition, subtraction, multiplication, division, inversion, absolute value, squaring, and square root transformations. Interaction terms were not restricted, and features were filtered to avoid numerical instabilities by bounding the absolute values between 10^{-50} and 10^{50}. All features were assigned physical units (e.g., electron volt, angstrom, or dimensionless) to ensure dimensional consistency during construction.

Model training was performed in a regression setting with the orthogonal matching pursuit (OMP) algorithm as the search strategy. To avoid redundancy, a maximum feature cross-correlation threshold of 0.95 was enforced. Candidate descriptor sets were restricted to size 50, with a maximum descriptor dimensionality of 2. Model selection was conducted using three-fold

cross-validation. The intercept term was not fixed, allowing flexible fitting across the dataset.

All calculations were parallelized across multiple cores (`n_jobs = -4`) with a fixed random seed (`random_state = 42`) to ensure reproducibility. The full feature space was cached and saved to enable efficient re-use in subsequent analyses.

Data Availability. All data is freely available at [url to be added in proof].

Code Availability. The DISCOVER code is freely available at `https://u-gajera.github.io/discover/`.

References

1. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.* **114**, 105503 (2015). URL `https://link.aps.org/doi/10.1103/PhysRevLett.114.105503`.

2. Scheffler, M. *et al.* Fair data enabling new horizons for materials research. *Nature* **604**, 635–642 (2022). URL `https://doi.org/10.1038/s41586-022-04501-x`.

3. Sotoudeh, M. & Groß, A. Computational screening and descriptors for the ion mobility in energy storage materials. *Curr. Opin. Electrochem.* **46**, 101494 (2024). URL `https://www.sciencedirect.com/science/article/pii/S2451910324000553`.

4. Gajera, U., Storchi, L., Amoroso, D., Delodovici, F. & Picozzi, S. Toward machine learning for microscopic mechanisms: A formula search for crystal structure stability based on atomic

properties. *Journal of Applied Physics* **131**, 215703 (2022). URL https://doi.org/10.1063/5.0088177. https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/5.0088177/16508143/215703_1_online.pdf.

5. Nørskov, J. K. *et al.* Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J. Phys. Chem. B* **108**, 17886–17892 (2004). URL https://doi.org/10.1021/jp047349j. https://doi.org/10.1021/jp047349j.

6. Sotoudeh, M. & Groß, A. Descriptor and scaling relations for ion mobility in crystalline solids. *JACS Au* **2**, 463–471 (2022). URL https://doi.org/10.1021/jacsau.1c00505. https://doi.org/10.1021/jacsau.1c00505.

7. Sotoudeh, M. *et al.* Oxide spinels with superior mg conductivity. *Chem. Mater.* **35**, 4786–4797 (2023). URL https://doi.org/10.1021/acs.chemmater.3c00634.

8. Koza, J. R. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **4**, 87–112 (1994). URL https://doi.org/10.1007/BF00175355.

9. Udrescu, S.-M. & Tegmark, M. Ai feynman: A physics-inspired method for symbolic regression. *Science Advances* **6**, eaay2631 (2020). URL https://www.science.org/doi/abs/10.1126/sciadv.aay2631. https://www.science.org/doi/pdf/10.1126/sciadv.aay2631.

10. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an im-

mensity of offered candidates. *Phys. Rev. Materials* **2**, 083802 (2018). URL https://link.aps.org/doi/10.1103/PhysRevMaterials.2.083802.

11. Nelson, L. J., Hart, G. L. W., Zhou, F. & Ozoliņš, V. Compressive sensing as a paradigm for building physics models. *Phys. Rev. B* **87**, 035125 (2013). URL https://link.aps.org/doi/10.1103/PhysRevB.87.035125.

12. Purcell, T. A. R., Scheffler, M. & Ghiringhelli, L. M. Recent advances in the sisso method and their implementation in the sisso++ code. *J. Chem. Phys.* **159**, 114110 (2023). URL https://doi.org/10.1063/5.0156620. https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0156620/18931308/114110_1_5.0156620.pdf.

13. Gajera, U., Sotoudeh, M., Sarkar, K. & Groß, A. Discover: A physics-informed, gpu-accelerated symbolic regression framework (2026). URL https://arxiv.org/abs/2602.06986. 2602.06986.

14. Thackeray, M., David, W., Bruce, P. & Goodenough, J. Lithium insertion into manganese spinels. *Mater. Res. Bull.* **18**, 461–472 (1983). URL https://www.sciencedirect.com/science/article/pii/0025540883901381.

15. Liu, X. *et al.* High stable post-spinel namn2o4 cathode of sodium ion battery. *J. Mater. Chem. A* **2**, 14822–14826 (2014). URL http://dx.doi.org/10.1039/C4TA03349C.

16. Canepa, P. *et al.* High magnesium mobility in ternary spinel chalcogenides. *Nat. Commun.* **8**, 1759 (2017). URL https://doi.org/10.1038/s41467-017-01772-1.

17. Kobayashi, H. *et al.* Ultraporous, ultrasmall mgmn2o4 spinel cathode for a room-temperature magnesium rechargeable battery. *ACS Nano* **17**, 3135–3142 (2023). URL https://doi.org/10.1021/acsnano.2c12392. https://doi.org/10.1021/acsnano.2c12392.

18. Glaser, C. *et al.* High room-temperature magnesium ion conductivity in spinel-type mgyb2se4 solid electrolyte. *Chem. Mater.* **37**, 3353–3362 (2025). URL https://doi.org/10.1021/acs.chemmater.5c00131. https://doi.org/10.1021/acs.chemmater.5c00131.

19. Glaser, C. *et al.* Mgb2se4 spinels (b = sc, y, er, tm) as potential mg-ion solid electrolytes – partial ionic conductivity and the ion migration barrier. *Adv. Energy Mater.* **14**, 2402269 (2024). URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aenm.202402269. https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.202402269.

20. Sotoudeh, M., Dillenz, M. & Groß, A. Mechanism of magnesium transport in spinel chalcogenides. *Adv. Energy Sustainability Res.* **2**, 2100113 (2021). URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aesr.202100113. https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aesr.202100113.

21. Dillenz, M. *et al.* Unravelling charge carrier mobility in d_0-metal-based spinels. *Batter. Supercaps* **5**, e202200164 (2022).

22. Sotoudeh, M. & Groß, A. Stability of magnesium binary and ternary compounds for batteries determined from first principles. *J. Phys. Chem. Lett.* **13**, 10092–10100 (2022). URL https://doi.org/10.1021/acs.jpclett.2c02316. https://doi.org/10.1021/acs.jpclett.2c02316.

23. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964). URL http://link.aps.org/doi/10.1103/PhysRev.136.B864.

24. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965). URL http://link.aps.org/doi/10.1103/PhysRev.140.A1133.

25. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996). URL http://link.aps.org/doi/10.1103/PhysRevLett.77.3865.

26. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994). URL http://link.aps.org/doi/10.1103/PhysRevB.50.17953.

27. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993). URL https://link.aps.org/doi/10.1103/PhysRevB.47.558.

28. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996). URL https://link.aps.org/doi/10.1103/PhysRevB.54.11169.

29. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999). URL https://link.aps.org/doi/10.1103/PhysRevB.59.1758.

30. Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: An lsda+u study. *Phys. Rev. B* **57**, 1505–1509 (1998). URL https://link.aps.org/doi/10.1103/PhysRevB.57.1505.

31. Song, Z. & Liu, Q. Tolerance factor and phase stability of the normal spinel structure. *Cryst. Growth Des.* **20**, 2014–2018 (2020). URL https://doi.org/10.1021/acs.cgd.9b01673. https://doi.org/10.1021/acs.cgd.9b01673.

32. Sotoudeh, M., Dillenz, M. & Groß, A. Mechanism of magnesium transport in spinel chalcogenides. *Adv. Energy Sustainability Res.* **2**, 2100113 (2021). URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aesr.202100113. https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aesr.202100113.

33. Sotoudeh, M. *et al.* Ion mobility in crystalline battery materials. *Adv. Energy Mater.* **14**, 2302550 (2024). URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aenm.202302550. https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.202302550.

34. Dillenz, M., Sotoudeh, M., Euchner, H. & Groß, A. Screening of charge carrier migration in the $MgSc_2Se_4$ spinel structure. *Front. Energy Res.* **8**, 584654 (2020). URL https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2020.584654.

35. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013). URL https://doi.org/10.1063/1.4812323. https://pubs.aip.org/aip/apm/article-pdf/doi/10.1063/1.4812323/13163869/011002_1_online.pdf.

36. Sheppard, D., Terrell, R. & Henkelman, G. Optimization methods for finding minimum energy paths. *J. Chem. Phys.* **128**, 134106 (2008). URL https://doi.org/10.1063/1.2841941.

37. Mehrer, H. *Diffusion in Solids: Fundamentals, Methods, Materials, Diffusion-Controlled Processes*. Springer Series in Solid-State Sciences (Springer Berlin Heidelberg, 2007). URL https://doi.org/10.1007/978-3-540-71488-0.

38. Wilkinson, M. D. *et al.* The fair guiding principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016). URL https://doi.org/10.1038/sdata.2016.18.

39. Draxl, C. & Scheffler, M. Nomad: The fair concept for big-data-driven materials science (2018). URL https://arxiv.org/abs/1805.05039. 1805.05039.

40. Scheidgen, M. *et al.* Nomad: A distributed web-based platform for managing materials science research data. *Journal of Open Source Software* **8**, 5388 (2023). URL https://doi.org/10.21105/joss.05388.

41. Schumann, J., Pizarro, J. M., Gajera, U., Rudzinski, J. F. & Näsström, H. NOMAD NEB workflows (2025). URL https://zenodo.org/records/17533661.

Acknowledgements This work contributes to the research carried out within CELEST (Center for Electrochemical Energy Storage Ulm-Karlsruhe) and was supported by the German Research Foundation (DFG) under Project ID 390874152 (POLiS Cluster of Excellence). U.G. acknowledges support from the NFDI consortium FAIRmat, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project 460197019, and from the Dr. Barbara Mez-Starck-Foundation. Computational resources were provided by the state of Baden-Württemberg through bwHPC, and by the German Research Foundation (DFG) through grant no. INST 40/575-1 FUGG (JUSTUS 2 cluster), which are gratefully acknowledged. We also wish to thank Julia Schumann, Joseph F. Rudzinski, Hampus Näsström, José M. Pizarro, and the NOMAD team for their valuable support in implementing the nomad-neb-workflow.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to

Mohsen Sotoudeh (email: mohsen.sotoudeh@kit.edu).