**MRS Communications**

**Early Career Distinguished Presenter Prospective**

# Specialized modeling, database, and AI for Li-ion and Na-ion battery materials

**Benjamin Cahill, Lin Wang, and Bin Ouyang**[iD]**,** Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32304, USA

Address all correspondence to Bin Ouyang at **bouyang@fsu.edu**

## Abstract

Li-ion and Na-ion battery materials have experienced rapid growth over the past decades and have become emblematic of the clean energy industry. Driven by advances in computational methods and data-driven research, this review provides an overview of key developments in physical modeling, data mining, and machine learning. In particular, it highlights how to understand the complex interplay of various types of disorder, accurately predict ionic conductivity, and develop specialized databases and machine learning frameworks.

## Introduction

The growing demand for next-generation rechargeable Li-ion and Na-ion batteries has led researchers toward previously dismissed avenues of research.[1–9] Despite all kinds of innovation made in this field, developing low-cost and high-performance Li-ion and Na-ion batteries remains a persistent pursue.[10] With the rise of data science, how much it can benefit the development of Li-ion and Na-ion battery materials is an open question. Several important problems to solve are (a) How to reduce the cost of materials and the usage of critical metals with minimum performance trade-offs; (b) How to leverage the existing modeling tools to enhance mechanistic understanding of electrochemical process; (c) Expanding and refining existing battery databases to address more specialized problem rather than offering only a broad overview build on generic materials databases.

These prospectives aim to present an overview of progress in tackling the above-mentioned challenges. Particularly, it will summarize several key specialized modeling tools available, as well as the specialized problems they can solve.[11–23] Examples can be represented by the modeling of disordering in batteries, understanding the key local structures for enhancing ion diffusion, as well as predicting low temperature diffusion behaviors that deviate from Arrhenius relation. Moreover, the development of specialized Li-ion and Na-ion dataset contrasting with the Moore's scaling law of transistor development will be reviewed, aiming at highlighting the capability of data mining nowadays for Li-ion and Na-ion battery development. Last but not least, a perspective of how to adapt the state-of-the-art machine learning into acceleration of Li-ion and Na-ion battery development will be presented, with particular emphasis on what is overlooked in the current surging foundation AI development. It is hoped that this work can offer a quick refreshment of where we are in terms of data-driven Li-ion and Na-ion battery design.

## Specialized modeling tools

Disorder is a common feature in Li-ion and Na-ion materials, particularly for those battery materials with high performance. In such systems, ionic conductivity is influenced by both local structural disorder and the formation of percolation networks. Recent advances have shown that compositional disorder and short-range order play key roles in shaping electrochemical performance. Percolation analysis can be used to optimize transport properties. When combined with computational methods such as *Ab initio* Molecular Dynamics (AIMD) and Machine Learning Force Field Molecular Dynamics (MLFF-MD), structural analysis enables the evaluation of ionic conductivity. Together, these approaches offer a framework for engineering disorder for enhancing ion transport.

### The modeling of disordering and percolation network

Disordering is a common phenomenon in both Li-ion and Na-ion battery materials.[24,25] In most cases, this disorder is compositional, where the overall crystal structure remains largely intact, but ionic mixing occurs well beyond the dilute limit. Modeling and understanding such disorder require specialized tools from both theoretical and experimental perspectives, with a strong emphasis on local structural analysis. Furthermore, the structure–property relationship in battery materials critically depends on understanding how local structures influence electrochemical properties, particularly ion transport.

Substantial efforts have been devoted to studying metal oxides with coupled FCC sublattices for both cations and anions. Notably, common cathode materials such as layered and spinel structures feature distinct Wyckoff positions for transition metal (TM) ions and lithium sites, effectively separating their sublattices.[11] More recently, cation-disordered rock-salt (DRX) Li-ion battery materials have attracted increasing attention.[26–28] In these structures, lithium and transition
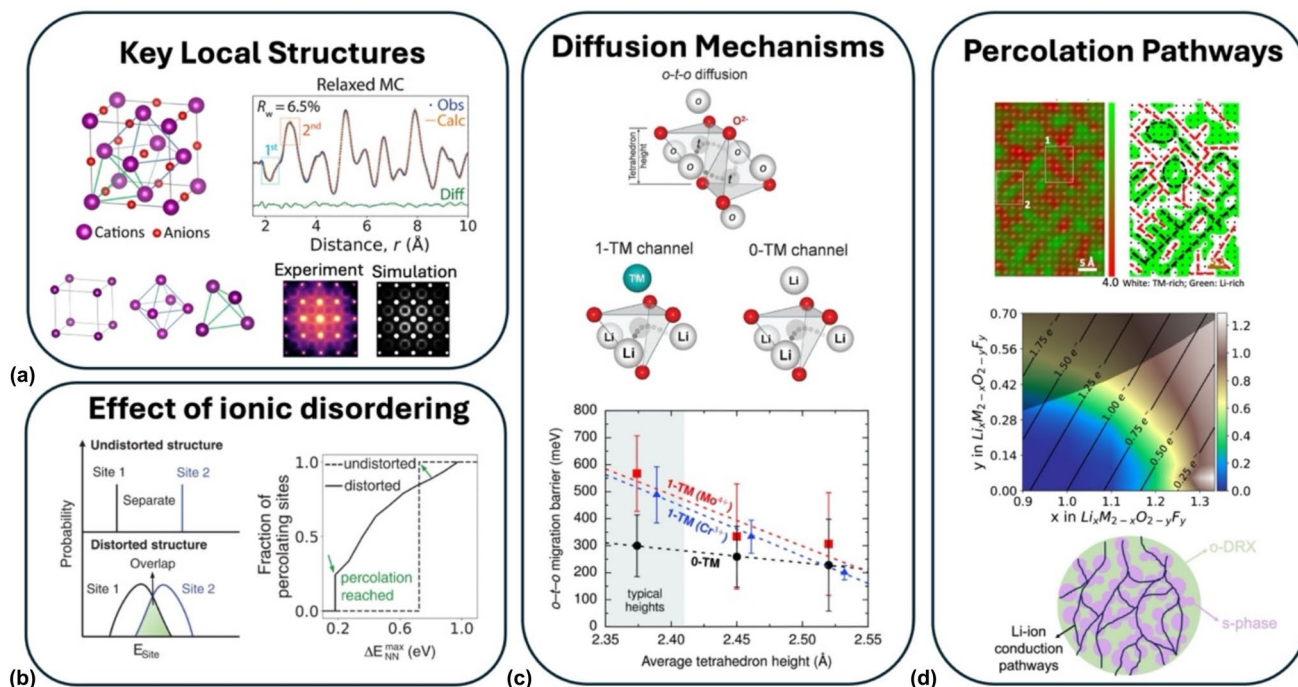
**Figure 1.** (a) Illustration of the key local structures and representative characterization results in Li-ion cathode materials. The top-left crystal structure is that of a typical DRX crystal structure, with the cations being Li or TMs and the anions typically being oxygen or fluorine.[15] The bottom-left crystal structure is related to the types of CSRO that has been observed in cation sublattices of DRX.[15] The top-right graph is a comparison between a calculated PDF pattern and an observed pattern.[14] The bottom right figure is a comparison between the experimental electron diffraction with the simulated one from cluster-expansion Monte Carlo simulation.[15] (b) The schematic on the left demonstrates how lattice distortion caused by disordering causes overlapping energy densities.[32] The graph on the right shows the fraction of Li percolating sites as a function of energy for both distorted and undistorted structures.[32] (c) Illustration of the diffusion mechanisms that occur in Li-ion cathode materials. The o-t-o diffusion is the general ion hopping pathway.[11] The 1-TM channel signifies that the Li in the tetrahedral while hopping will face-share with 1 TM.[11] The 0-TM channel signifies that the Li in the tetrahedral site while hopping will face-share with 0 TMs. The graph at the bottom is the calculated Li migration barriers 0-Tm and 1-TM channels.[11] (d) Top graph is the processed atomic resolution STEM image that shows the Li distribution because of CSRO.[15] The middle graph shows the computationally predicted percolation map in Li–M–O–F system.[13] The bottom graph shows a schematic of how percolation happens at particle size level in a composite.[33]

metal ions share the same sublattice, resulting in an "average" rocksalt-like crystal structure.[13] A demonstration of the typical rocksalt-like crystal structure and the locations of cations and anions is given in the top left of Fig. 1(a). The periodic bond topology in DRX Li-ion battery materials is referred to as long-range order (LRO). However, these materials often exhibit remnants of cation pair preferences, leading to a variety of local environments known as chemical short-range order (CSRO).[29] The identification of these local structures can be seen through the analysis of the pair distribution function (PDF).[14] By inspecting the Bragg peaks of the scattering spectrum, one can gain insight into the LRO of the average structure. As a contrast, by looking at the diffuse scattering from different total scattering tools, one can gain a quantitative description of CSRO.[14,30] This is seen in the graph in Fig. 1(a), where a change can be observed between the observed and calculated peaks, giving insight into the appearance of CSRO.[30] By further identification and understanding of the local structures within a DRX Li-ion crystal structure,

one can investigate the control handles of disordering as well as the ion transport.[31]

More specifically, a generalized theoretical framework that explains the synergistic effect of disordering is presented in Fig. 1(b). Particularly, the random distribution of local distortion will effectively create a spectrum of site energies, which could lead to the potential of similar site energies of nearest neighboring sites.[32] By controlling the population and distribution of local structures, we can effectively control the site energy overlaps, which eventually lead to tailorable microscopic diffusivity.

Beyond the general theory of disorder's impact on ion diffusion, a specific example is illustrated in Fig. 1(c). Studies on DRX-type cathode materials have shown that ion transport is primarily governed by hops between octahedral sites *via* intermediate tetrahedral sites—referred to as o–t–o diffusion—as demonstrated in Fig. 1(c).[11] The activated Li in the tetrahedral site face shares with four octahedral sites, those being the site that the Li previously occupied, the site that the Li is moving into, and two other sites that can be occupied by

TMs or Li. In Li-TM oxide layered structures, this hopping mechanism occurs when there is one transition metal face-sharing with the tetrahedral site that the Li-ion is hopping through and is referred to as a 1-TM channel, Fig. 1(c).[11] The main diffusion mechanism for DRX cathode materials differs from layered compounds. Instead, Li diffusion occurs the easiest when there is no TMs face-sharing the tetrahedral site, which is referred to as 0-TM channels, Fig. 1(c).[11] The graph located in Fig. 1 shows that the energy barrier for Li diffusing through a 0-TM site is less than if it diffused through a 1-TM site for DRX materials. Therefore, 0-TM channels have the potential to enable facile diffusion of Li through DRX Li-ion cathode materials. However, this is not without its limitations, mainly due to the frequency of these channels that occur in structures.

With an understanding of key local structures, the rational design of electrochemical properties can be achieved by controlling the percolation of these critical features—for example, the 0-TM channel. Representative examples across different length scales, supported by both experimental and theoretical studies, are presented in Fig. 1(d). The top panel highlights work by Li et al., which combines aberration-corrected scanning transmission electron microscopy (STEM) with selected area electron diffraction (SAED), cluster-expansion Monte Carlo (CEMC) simulations, and simulations of both STEM imaging and electron diffraction.[15] The middle panel presents a theoretical percolation analysis based on Monte Carlo-simulated structures. The color contour illustrates the extent of Li percolation as a function of Li and F content.[13,34] These simulated structures align well with experimentally measured pair distribution functions (PDF) and SAED patterns, as shown in Fig. 1(a).[14] Furthermore, percolation behavior can extend to much larger length scales, as illustrated schematically in Fig. 1(d). While 0-TM channels offer reasonable diffusivity for high-rate cathode materials, they are still insufficient to meet the conductivity requirements for solid electrolyte applications.[35] To further boost the ionic conductivity, introducing an over-stoichiometric amount of Li into the formula has been shown as an effective strategy. With this, it unlocks a new range of chemistries that were previously thought unavailable. Looking at the bottom panel of Fig. 1(d), the extra Li that makes cation/anion ratio larger than one will lead to the formation of a composite at nanoscale. The composite is made from DRX phase as well as spinel-like phase with cation/anion ratio larger than one. The spinel-like phase, referred to as the s-phase in Fig. 1(d), shows orders of magnitude higher ionic conductivity than DRX phase. It has been shown in the bottom of Fig. 1(d) that if the s-phase can percolate well enough in the composite, the ionic conductivity at room temperature can reach mS/cm level.[36]

Understanding local structures and percolation in Li-ion battery materials has enabled the tuning of their electrochemical performance. Similarly, local structure has emerged as a critical design parameter in Na-ion battery materials. A prominent example is the well-known NASICON family, which exhibits high ionic conductivity and electrochemical stability, making it suitable for both electrode and solid-state electrolyte applications. Designing NASICONs for improved performance requires a deep understanding of how local structural features influence material behavior. Key factors include bottleneck size and Na-site ordering, both of which are typically affected by the mixing of polyanions or cations.[37] The bottleneck size governs the transition state through which Na ions migrate between sites, while Na-site ordering shapes the overall energy landscape for ion diffusion. Together, these factors define the key local structures that control ionic transport in NASICON-type materials.

To investigate how bottleneck size influences Na-ion diffusion, several metrics for quantifying the bottleneck geometry are presented in the top half of Fig. 2(a). The bottleneck is characterized in three ways: (1) the height (LBN) between the metal cation site (M) and the adjacent polyanion site that shares an edge with the bottleneck, (2) the area (SBN) of the triangular cross-section traversed by Na ions, and (3) the volume (VBN) of the bottleneck tetrahedron.[37] As demonstrated by He et al., the volume of the bottleneck tetrahedron shows the strongest correlation with ionic conductivity.[37] It is important to note that a minimum bottleneck size is required to achieve ionic conductivities greater than $10^{-4}$ S cm$^{-1}$. While the bottleneck volume (VBN) depends on composition, it is not solely determined by the average anion radius ($R_{Anion}^{Avg}$) or average cation radius (($R_{Cation}^{Avg}$) alone. As shown in the bottom half of Fig. 2(a), VBN can vary significantly even when either of these radii is held constant. Therefore, both radii must be considered together when optimizing VBN. Specifically, a combination of large cation and anion sizes is necessary to achieve larger bottleneck volumes.[37]

In some cases, a large bottleneck volume does not translate to high ionic conductivity. This discrepancy is primarily due to the second key descriptor—Na-site ordering—which governs the local hopping mechanism in NASICON materials. Depending on the local ordering, two distinct diffusion mechanisms can occur: single-ion hopping (SIH) and occupancy-conserved hopping (OCH). As illustrated in Fig. 2(b), the energy difference between the 6b and 18e sites varies with the specific diffusion mechanism.[37] When there is a strong ordering tendency of the charge carrier ion, the energy barrier for OCH is lower than that for SIH, as shown in the energy landscape of Fig. 2(b). This is because, in OCH, the 6b site is never fully vacated during hopping, whereas in SIH, it is completely vacated, resulting in a higher barrier. However, when the ordering tendency is minimal, the energy difference between the two diffusion mechanisms becomes negligible. This understanding also informs the design of high-entropy NASICON-type Li and Na conductors, where maximizing disorder effectively flattens the energy landscape and promotes more uniform ion transport.[32] These insights highlight the importance of understanding key local structures within a material class to effectively tune desired properties through engineered disorder. By elucidating and manipulating ionic conduction frameworks, new research
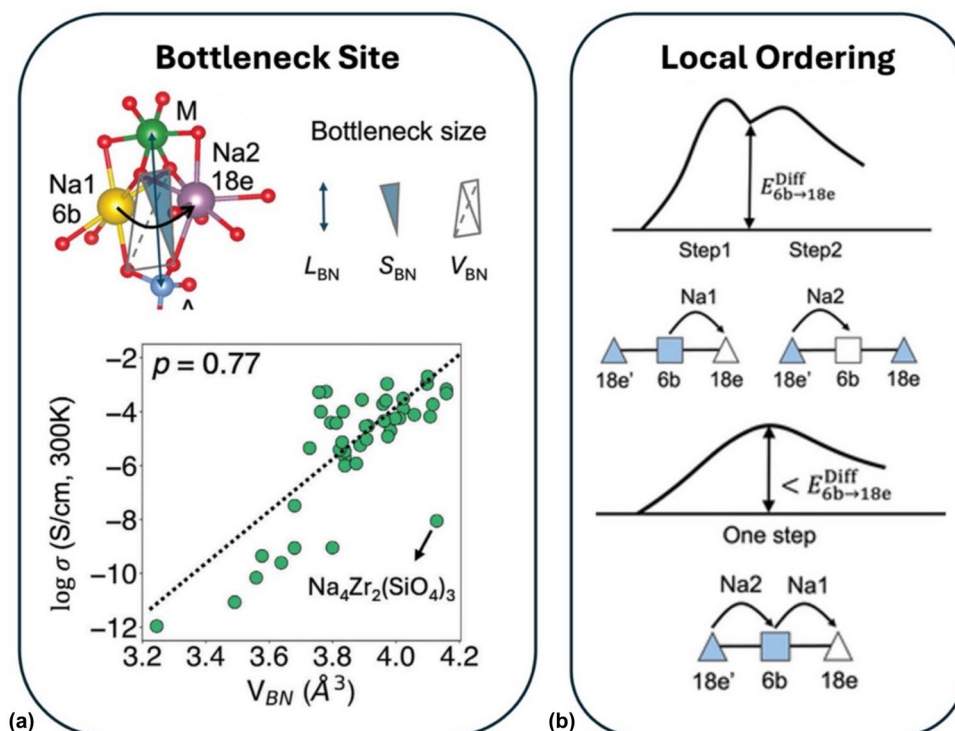
**Figure 2.** (a) The diagram at the top shows the local structures at which diffusion occurs along with the factors that are related to bottle-neck size and where they are.[37] The graph at the bottom is the 2D correlation between the bottleneck volume and the size of the anion and cation.[37] (b) Related to the diffusion mechanisms of Na in NASICONs.[37] The top possible diffusion mechanisms being SIH and OCH, with the energy landscapes for both mechanisms below.[37]

directions can emerge, enabling the discovery and design of next-generation Li-ion and Na-ion materials.

## Understand ion transport with AIMD and MLFF-MD

In addition to topological analysis of local structures and per-colation networks, techniques such as *Ab Initio* Molecular Dynamics (AIMD) and Machine Learning Force Field Molecu-lar Dynamics (MLFF-MD) offer direct estimates of diffusivity for a given structure. Moreover, these methods can help iden-tify key local structural features that govern ionic conductiv-ity. Both approaches have recently been employed to uncover previously overlooked interactions in a range of Li-ion and Na-ion battery materials. A schematic summary of these insights is presented in Fig. 3.

Taking Li-ion conductors as an example, AIMD has been used to analyze face-sharing configurations of lithium, inves-tigate complex diffusion, and study Li-ion interactions with other molecules.[33,38,39] When it came to probing Na-ion bat-tery materials, AIMD simulations were used to investigate Na-ion diffusion in various superionic conductors for helping guide the compositional optimization, while also probing if it is possible to use first-principle calculations to validate if a certain framework is viable for use as a solid-state superi-onic conductor.[39–41] In addition to directly simulating ionic

conductivity, AIMD also serves as a powerful tool for uncov-ering the underlying diffusion mechanisms. Two illustrative examples are shown in the top-left and top center panels of Fig. 3. In the top-left panel, AIMD was used to investigate the relationship between polyanion rotation and Li diffusion. The results indicate that the rotation of the lightweight $(BH_4)^-$ group has minimal impact on Li diffusion due to the mismatch in their characteristic mobility frequencies.[39] In the top center panel, the influence of polyanion species and Na-site occupancy on diffusion mechanisms was examined in Na argyrodites.[41] The Na probability distribution reveals that switching from exclu-sive occupancy of the 4c site to partial occupancy of both the 4a and 4c sites significantly alters the diffusion pathway, lead-ing to orders-of-magnitude enhancement in ionic conductivity.

AIMD usually has its drawback of being limited in pre-dicting at extended time and length scales. It thus becomes a challenge when comes to low temperature predictions, e.g., room temperature predictions. People will usually assume Arrhenius behavior from high T, e.g., 1500 K or more to low T, e.g., 300 K. As being demonstrated in the top-right panel of Fig. 3, such endeavor will not always work due to the poten-tial of structural evolution. The type I mechanism is what usu-ally occur, while type II and Type II behaviors are also widely observed in many materials.[42] Therefore, it may not always be reliable to extrapolate high-temperature AIMD simulations to room temperature. Machine learning force field that enables
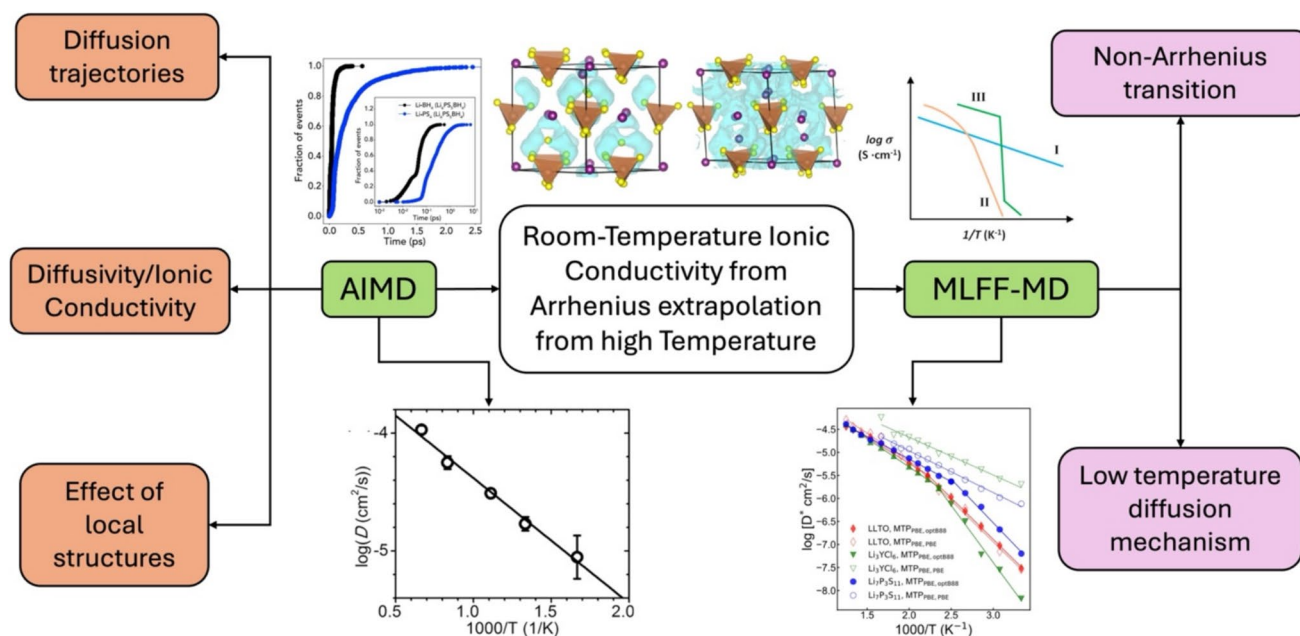
**Figure 3.** Illustration of what AIMD is useful for and what it struggles to predict. Also illustrated is MLFF-MD, which is used to make up for the pitfalls of AIMD. The illustration and graph above the green AIMD box refer to the analysis of the ankling effect of local structure on Li diffusion.[39] The graph located in the top center indicates a comparison between different diffusion mechanisms in sodium argyrodites with different disordering.[41] The graph under the green AIMD box is meant to show that it must assume Arrhenius behavior to calculate ionic conductivities at room temperature.[38] The graph on top of the green MLFF-MD box is meant to illustrate other types of trends that can be seen, and that purely assuming Arrhenius behavior is not ideal.[42] The graph below the MLFF-MD greenbox is meant to signify that MLFF-MD at low temperature can capture these non-Arrhenius transitions.[43]

longer time and larger scale simulations will be crucial to address those challenges.

As demonstrated by Qi et al., moment tensor potentials (MTP) were trained on DFT values, and for all three Li superionic conductors (LSCs) that were investigated, non-Arrhenius transitions were observed, which is also shown in the bottom right panel in Fig. 3.[43] The three LSCs that were selected had previously not shown non-Arrhenius behavior. The past five years have observed the rapid growth of various types of machine learning force fields. To give a few more examples on LSCs: perovskite lithium-ion conductors have also been investigated using MLFFs. More specifically, the influence of Li-site occupancy on ionic conductivity was investigated.[44] According to the ionic conductivities extrapolated from MLFF based MD simulations, depending on the site at which Li is located, the ionic conductivity would differ greatly.
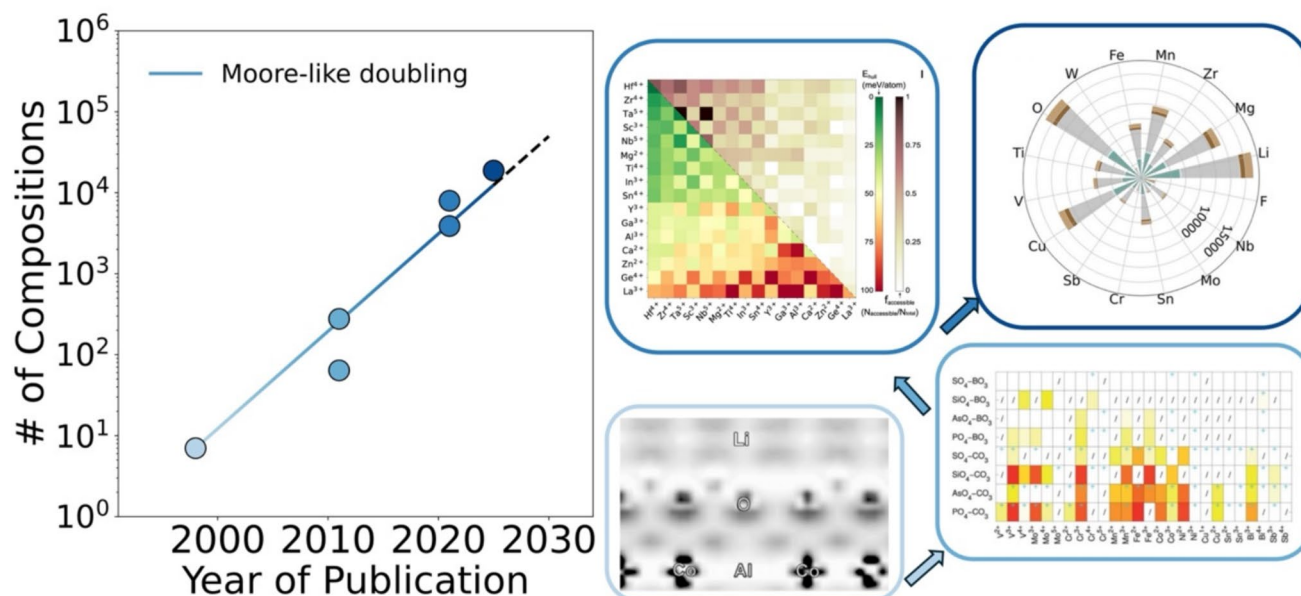
## Specialized database for Li-ion and Na-ion batteries

With all the specialized scientific problems identified in the section above, building specialized databases customized for Li-ion and Na-ion based systems will be the best way to pivot the materials design challenge for both electrode and electrolyte. The past three decades have witnessed the development of

battery design through first principle and the rising of materials genome approach. In this section, the development of specialized database for Li-ion and Na-ion batteries as well as the chemical science intuition will be reviewed.

## Implication of Moore's law in high-throughput battery discovery

The vast compositional space of Li-ion and Na-ion battery materials necessitates high-throughput screening, rather than relying on empirical searches. Visualizing the evolution of high-throughput efforts over the past three decades reveals significant progress in screening strategies for these materials. A landmark study in computational screening can be traced back to Ceder et al. in 1998,[45] which demonstrated the feasibility of using first-principles methods to identify viable dopants for $LiCoO_2$. This pioneering work led to the prediction and subsequent experimental validation of Al-doped $LiCoO_2$ with promising performance,[45] as illustrated in Fig. 4 above the year 1998. This study stands out as one of the earliest successful examples of computationally guided battery materials design. Its publication in Nature and coverage on the MIT website underscore its importance. However, nearly 30 years ago, computational screening was still in its infancy, typically limited to evaluating only a few dopants at a time. Roughly a decade later, the launch of the Materials Genome Initiative catalyzed the first large-scale

**Figure 4.** Visual illustration of the number of compositions reported in papers and the year those papers were published for solving specialized Li-ion and Na-ion battery materials[18,45–49] A Moore-like doubling line is also plotted in the graph meant to represent the progression that high-throughput DFT has improved over time. The color of the box next to the graph corresponds to the paper the figure is from.

computational screenings of battery materials. Around 2011, several notable high-throughput studies led by Ceder et al., as highlighted in Fig. 4, marked a turning point in the field. These efforts enabled the screening of hundreds of polyanion-based compounds for Li-ion and Na-ion electrodes, with a focus on identifying optimal trade-offs between phase stability and energy density.[46,47]

Tracing the Moore-like doubling trend shown as the dashed lines in Fig. 4, the throughput of DFT screening has surged to several thousand by around 2021. Ouyang et al. conducted large-scale screening of high-entropy disordered rocksalt-type Li-ion battery cathodes[48] and NASICON-type solid-state electrolytes,[18] covering 7965 and 3881 distinct compositions, respectively. More recently, in 2025, Wang et al. reported the largest specialized screening to date, with 18,810 DFT-computed disordered rocksalt-type cathode materials.[49] It is important to note that Fig. 4 does not aim to show the limit of high-throughput DFT calculations at each time period. It should be acknowledged that the true upper limit of high-throughput materials screening is much larger, as exemplified by initiatives like the Materials Project,[50] which aim to explore the global design space of materials. Instead, Fig. 4 is intended to illustrate what constitutes a practically achievable throughput when addressing a focused problem—such as the design of specific Li-ion and Na-ion battery materials. Encouragingly, as demonstrated by Lin et al.,[49] it is now feasible to screen the entire viable element space in combination with several prototype stoichiometries, encompassing tens of thousands of compositions. Moreover, it is worth mentioning that Moore's Law has reached a bottleneck in recent years due to the unlikely feasibility to further scale down transistors. It thus highlights

also that new paradigms of doing high-throughput battery materials screening will be needed in the near future.

## Data mining stability rules from specialized database

High-throughput screening, along with machine learning models and experimental synthesis, also allows for the identification of elemental stability rules. One representative elemental stability rule of cation-disordered oxides for earth abundant cathode materials in lithium-ion batteries is presented by Urban et. al.[17] as illustrated in Fig. 5(a). Cation-disorder is necessary for lithium transport in these lithium excess materials for eliminating or minimizing critical metals, such as Co and Ni.[51,52] Whether these lithium metal oxides formed an ordered or disordered structure was tested by using the materials Order–Disorder (OD) transition temperature ($T_{oroer}$).[17]

On the other side, the elemental stability rules for NASICON electrolyte have been established by text mining the 3881 DFT-computed compositions, with a visual representation of the rules in Fig. 5(b). The first key rule determined was the electronegativity of a metal requires a very careful balance, as it is bounded by both sides in a NASICON structure.[18] When a metal is too electronegative, it will favor competing phases where it is bound in a more covalent bonding environment. If it is too electropositive, it will also be unstable due to the removal of the necessary sodium-metal competition for ionization. The second key rule that was identified was that a large size difference among the M-sites as well as the mixing of the polyanions would lead to instability of the crystal structure.[18]

Last but not the least, data mining of elemental stability rules has been further extended to high-entropy Li-ion battery
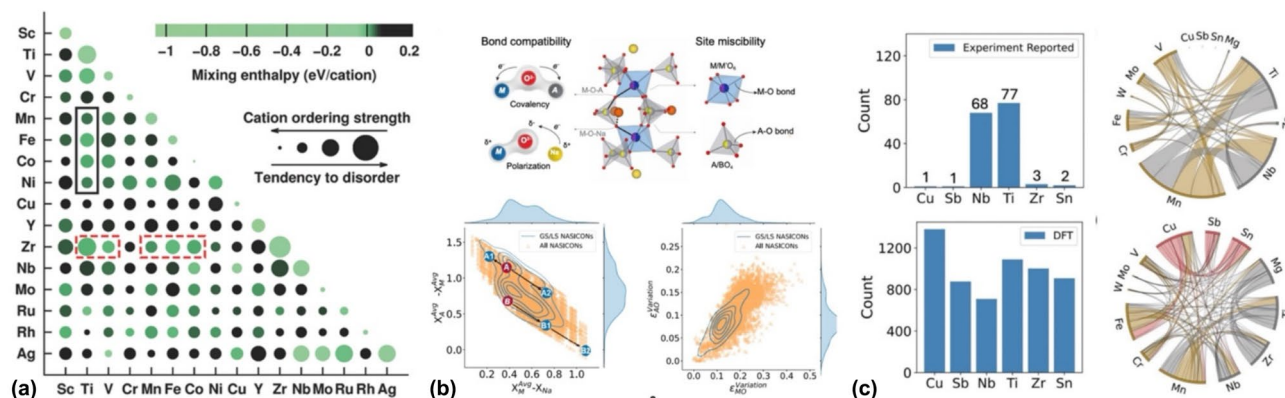
**Figure 5.** (a) This graph represents the screening results that Urban et al. gathered from screening the composition space $LiA_{0.5}B_{0.5}O_2$[17] The color of the dot corresponds to the stability of the compound. The size of the dot indicates the likelihood that the cation will disorder. (b) The two schematics above the graphs represent the two major contributing factors to NASICON stability, bond compatibility and site miscibility[18] The graphs below show the distribution all the NASICON compositions reported (c) The top-left graph is the number of HE-DRX compounds experimentally reported which contains the metals seen, while the bottom-left graph is the number of HE-DRX compounds with the metals that are experimentally feasible ($E_{hull, 1473 K} < 0$) in the dataset generated by Wang et al.[49] The top chord diagram shows the elemental compatibility for the experimentally reported compounds, while the bottom chord diagram shows the elemental compatibility of HE-DRX compounds that are experimentally feasible ($E_{hull, 1473 K} < 0$) in the dataset generated by Wang et al.[49]

materials.[49,53] The vast composition space urges larger scale databases, while it also values the insights from elemental stability. One recent effort from Wang et.al. has charted the elemental space and established elemental stability rules by data mining 18,810 DFT computed HE-DRX compounds.[49] Such a large database was required to chart the large-composition space that exists for HE-DRX compounds. Not only that, but as seen in the top-left and top-right graph of Fig. 5(c), as of the date of publishing, there was a limited number of compounds experimentally reported that contained those metals. While the graphs below demonstrate that the elemental space charted by Wang et al. has been comprehensive enough to propose elemental stability rules of HE-DRX compounds.

There were three main considerations when it came to the stability rules of HE-DRX compounds. The first key consideration is the redox compatibility of the compound, meaning that the electron configurations of the metals must be able to coexist in a single phase. The two next considerations fall hand in hand with each other, those being the stability of the redox centers and charge compensators. Typically, the redox centers and charge compensators that were considered stable were the metals that had similar Shannon radii in an octahedral environment to reported stable HE-DRX compounds.[49,54] This offers the potential of further investigation of stability of this composition space and replacement or minimization of cobalt and nickel in HE-DRX compounds.

## Specialized AI

As demonstrated above, with the ever-improving computational tools to assess a wide-range of compositional spaces, the adoption of models that offer insight into a wide variety of chemical properties is a desired outcome. Particularly, exploring the capability of proposing phenomenological theories for compounds, along with the prediction of viable candidates for desired use.

## Symbolic machine learning for phenomenological theory

Phenomenological models offer not only predictive capabilities but also valuable physical insights into stability mechanisms and have been widely adopted throughout the history of chemistry and materials science. Notable examples in materials discovery include the Goldschmidt tolerance factor for perovskite and the Zintl–Klemm rule for Zintl phases.[55] However, formulating such phenomenological models is often nontrivial due to the structural and chemical complexity inherent in crystalline materials. In parallel with advances in deep learning efforts, symbolic machine learning has emerged as a powerful tool for accelerating the discovery of phenomenological theory.[18–20] Recently, considerable efforts have been devoted to extending symbolic machine learning approaches to the design of Li-ion and Na-ion batteries.

The SISSO method, which is an abbreviation of Sure Independent Screening (SIS), in combination with sparsifying operator (SO) has emerged as an important symbolic machine learning protocol for materials science, with a simplified workflow in Fig. 6(a).[20] In addition to the original framework proposed by Ouyang et. al., the SIS part can be further interfaced with different classic machine learning models, such as decision tree, or machine learning ranking to achieve different results.[20,56] Outcomes are seen in the graphs of Fig. 6(b, c) after the simplified workflow of the respective processes. Particularly, SIS+MLR has been used to find a tolerance factor that can be used to evaluate the synthesizability of NASICON solid-state electrolytes. Developed by Ouyang et al., the reason MLR was
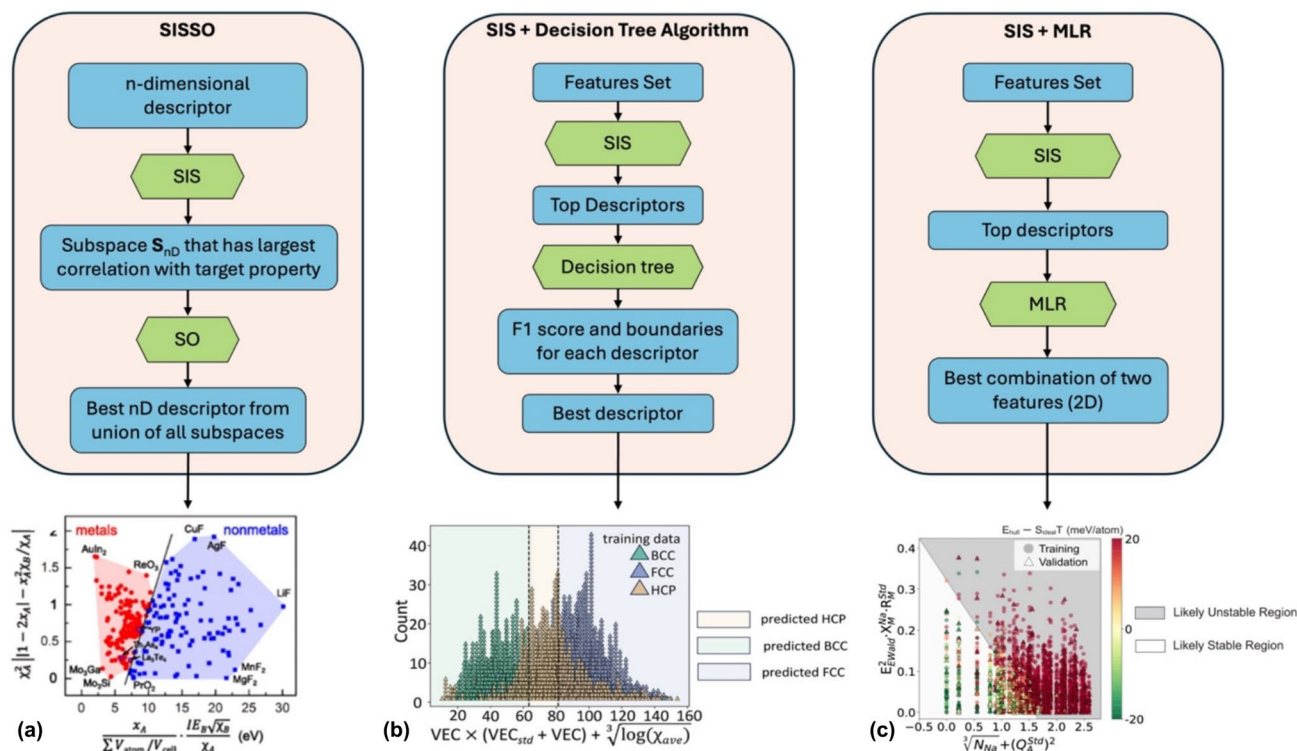
**Figure 6.** (a) Workflow for SISSO, with the associated graph being an example of what SISSO was used to model[20] (b) Workflow for using SIS + a Decision Tree Algorithm, with the associated graph being one example of how it can be used to predict phase stability[19] (c) Workflow for SIS + MLR, with the associated being an example of how SIS + MLR can be used to model stability of NASICONs[18]

utilized as the ranking of relative stability wanted to be the ordering of $E_{hull} - S_{Ideal}T$, not the absolute value that would be generated through the utilization of only SISSO.[18] Like previously, SIS was first applied to the large feature space proposed, and the top 2000 1D features were identified. From the 2000 1D features, MLR was used to find the best combination of two features that describes the stability of NASICONs. Driven by the successful application of symbolic machine learning in materials discovery, great opportunities are available to further translate the high-throughput computational dataset into phenomenological theory and chemical intuition.[18–20,56]

### Specialized AI versus general AI

In contrast to the relatively less explored domain of symbolic machine learning, deep learning based on neural networks has significantly advanced the field of materials chemistry in recent years. Substantial efforts have been devoted to developing universal machine learning models capable of rapidly predicting material properties across broad chemical spaces.[21,22,57–61] While benchmark datasets such as the Materials project[50] continue to yield progressively lower test mean absolute errors (MAEs), there is still no definitive claim that an all-compassing AI model exists—one that can accurately predict the properties of any Li-ion or Na-ion battery composition or structure. The gap highlights a key challenge in materials AI: rather than

relying solely on the promise of a hypothetical artificial general intelligence (AGI) capable of solving all materials problems at once, it may be more effective to develop specialized AI agents tailored to specific materials domain.[62] Such specialized AIs, trained on domain-specific databases, are better suited to capture the intricacies of targeted materials challenges Fig. 7. Moreover, as shown in Fig. 7, when a specialized database is employed, the influence of model complexity on performance becomes much less pronounced compared to training data size.

Indeed, with modern high-throughput DFT capabilities, we can now generate millions of materials data points.[10,50] However, the true chemical design space is vastly larger—potentially exceeding millions of millions. Taking Li-ion batteries as an example, if we consider 20 viable metals and the exploration of high-entropy compositions (e.g., a 9-dimensional compositional space), even with a 0.1 increment in atomic fraction and one structure type, the number of possible materials already exceeds 25.6 billion. Attempting to learn from a few million data points—most of which may not be battery-relevant—within a generic database poses clear limitations for battery design. In such cases, using a universal model is akin to learning about the ocean from a million droplets: unless the algorithm possesses truly exceptional generalization capabilities, its utility remains questionable. While the transformative potential of AI should not be dismissed, we remain cautious in expecting generic models to effectively accelerate the discovery of Li-ion and Na-ion batteries at this stage.
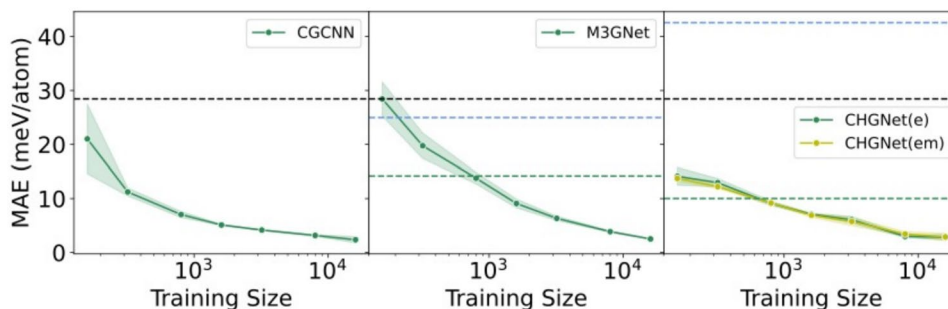
**Figure 7.** Performance of all three models on HE-DRX dataset[23] The black dotted line refers to the dummy model. The blue dotted line refers to the pretrained model. The green dotted line refers to the shifted pretrained model.

Based on this analysis, we advocate for the development of specialized AI systems rather than placing all our hopes on artificial general intelligence (AGI). To use an analogy: a million data points can yield far more insight if you're studying a swimming pool or a pond, rather than attempting to characterize the entire ocean. Recent progress in specialized AI lends strong support to this perspective. For example, applying a standard graph neural network to a domain-specific database—such as high-entropy disordered rocksalt (HE-DRX) materials used in Li-ion battery cathodes—has achieved testing mean absolute errors (MAEs) as low as 2 meV/atom, approaching the intrinsic noise level of DFT calculations. Crucially, these MAEs are orders of magnitude lower than those reported for universal machine learning models trained on general-purpose datasets.[21,22,57–61] Interestingly, this performance trend holds across different deep learning architectures, suggesting that model complexity is less critical than the quality and specialization of the training data. These findings further emphasize the value of developing specialized AI agents focused on targeted research domains. Rather than a single AI to solve all of materials science, it is far more promising to build tailored AI tools—for instance, ones focused specifically on battery design, or even on a narrower subset such as layered oxide cathodes.

## Conclusion

Advances in modeling, databases, and AI has led to the broadening of design landscape for Li-ion and Na-ion battery materials. Through the incorporation of AIMD and MLFF-MD into structure and percolation analysis, underlying mechanisms could be understood. This, along with the creation of specialized databases and advancements in high-throughput screening, has led to the establishment of stability rules that were not previously understood. Moreover, the use of AI in this same field has opened the door to potential futures where domain-specific models are able to accurately predict desired chemical properties. Together, these tools offer a framework for engineering disorder for future battery applications. As a contrast to the artificial general AI work, it is suggested that more emphasize should also be put on developing specialized AI framework.

## Author contributions

B.C. led the writing of this paper under the guidance of B.O. L.W. contributed to the writing and provided feedback.

## Declarations

### Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

1. P. Canepa, G.S. Gautam, D.C. Hannah, R. Malik, M. Liu, K.G. Gallagher, K.A. Persson, G. Ceder, Chem. Rev. (2017). https://doi.org/10.1021/acs.chemrev.6b00614
2. Y. Tian, G. Zeng, A. Rutt, T. Shi, H. Kim, J. Wang, J. Koettgen, Y. Sun, B. Ouyang, T. Chen, Z. Lun, Z. Rong, K. Persson, G. Ceder, Chem. Rev. (2021). https://doi.org/10.1021/acs.chemrev.0c00767
3. S.-W. Kim, D.-H. Seo, X. Ma, G. Ceder, K. Kang, Adv. Energy Mater. (2012). https://doi.org/10.1002/aenm.201200026
4. P. Nzereogu, A. Omah, F. Ezema, E. Iwuoha, A. Nwanya, Appl. Surf. Sci. Adv. (2022). https://doi.org/10.1016/j.apsadv.2022.100233
5. J. Xu, X. Cai, S. Cai, Y. Shao, C. Hu, S. Lu, S. Ding, Energy Environ. Mater. (2023). https://doi.org/10.1002/eem2.12450
6. P. Gupta, S. Pushpakanth, M.A. Haider, S. Basu, ACS Omega (2022). https://doi.org/10.1021/acsomega.1c05794
7. K. Lemoine, A. Hémon-Ribaud, M. Leblanc, J. Lhoste, J.-M. Tarascon, V. Maisonneuve, Chem. Rev. (2022). https://doi.org/10.1021/acs.chemrev.2c00247
8. C. Vaalma, D. Buchholz, M. Weil, S. Passerini, Nat. Rev. Mater. (2018). https://doi.org/10.1038/natrevmats.2018.13
9. H. Zhang, X. Gao, Q. Cai, X. Zhang, Y. Tian, M. Jia, W. Xie, Y. Du, X. Yan, J. Mater. Chem. A (2023). https://doi.org/10.1039/D3TA00852E
10. L. Barroso-Luque, M. Shuaibi, X. Fu, B.M. Wood, M. Dzamba, M. Gao, A. Rizvi, C.L. Zitnick and Z.W. Ulissi: Open materials 2024 (omat24)

inorganic materials dataset and models, in arXiv preprint arXiv:2410.12771 (2024).

11. J. Lee, A. Urban, X. Li, D. Su, G. Hautier, G. Ceder, Science (2014). https://doi.org/10.1126/science.1246432

12. L. Wang, Z. He, B. Ouyang, Comput. Mater. Sci. (2023). https://doi.org/10.1016/j.commatsci.2023.112513

13. B. Ouyang, N. Artrith, Z. Lun, Z. Jadidi, D.A. Kitchaev, H. Ji, A. Urban, G. Ceder, Adv. Energy Mater. (2020). https://doi.org/10.1002/aenm.201903240

14. N.J. Szymanski, Z. Lun, J. Liu, E.C. Self, C.J. Bartel, J. Nanda, B. Ouyang, G. Ceder, Chem. Mater. (2023). https://doi.org/10.1021/acs.chemmater.2c03827

15. L. Li, B. Ouyang, Z. Lun, H. Huo, D. Chen, Y. Yue, C. Ophus, W. Tong, G. Chen, G. Ceder, C. Wang, Nat. Commun. (2023). https://doi.org/10.1038/s41467-023-43356-2

16. J.B. Boyce, B.A. Huberman, Superionic conductors: transitions, structures, dynamics. Phys. Rep. (1979). https://doi.org/10.1016/0370-1573(79)90067-X

17. A. Urban, I. Matts, A. Abdellahi, G. Ceder, Adv. Energy Mater. (2016). https://doi.org/10.1002/aenm.201600488

18. B. Ouyang, J. Wang, T. He, C.J. Bartel, H. Huo, Y. Wang, V. Lacivita, H. Kim, G. Ceder, Nat. Commun. (2021). https://doi.org/10.1038/s41467-021-26006-3

19. L. Wang, B. Ouyang, Adv. Mater. (2024). https://doi.org/10.1002/adma.202307860

20. R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L.M. Ghiringhelli, Phys. Rev. Mater. (2018). https://doi.org/10.1103/PhysRevMaterials.2.083802

21. C. Chen, S.P. Ong, Nat. Comput. Sci. (2022). https://doi.org/10.1038/s43588-022-00349-3

22. B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C.J. Bartel, G. Ceder, Nat. Mach. Intell. (2023). https://doi.org/10.1038/s42256-023-00716-3

23. L. Wang, T. He, B. Ouyang, ACS Mater. Lett. **7**, 2708–2715 (2025)

24. C. Oses, C. Toher, S. Curtarolo, Nat. Rev. Mater. (2020). https://doi.org/10.1038/s41578-019-0170-8

25. R.-Z. Zhang, M.J. Reece, J. Mater. Chem. A (2019). https://doi.org/10.1039/C9TA05698J

26. Z. Zhou, Y. Ma, T. Brezesinski, B. Breitung, Y. Wu, Y. Ma, Energy Environ. Sci. (2025). https://doi.org/10.1039/D4EE03708A

27. L. Huang, J. Zhu, J.-X. Liu, H. Wu, G.-J. Zhang, J. Adv. Ceram. (2024). https://doi.org/10.26599/JAC.2024.9220913

28. Q. Wang, A. Sarkar, D. Wang, L. Velasco, R. Azmi, S.S. Bhattacharya, T. Bergfeldt, A. Düvel, P. Heitjans, T. Brezesinski, H. Hahn, B. Breitung, Energy Environ. Sci. (2019). https://doi.org/10.1039/C9EE00368A

29. J.M. Cowley, J. Appl. Phys. (1950). https://doi.org/10.1063/1.1699415

30. M.W. Terban, S.J.L. Billinge, Chem. Rev. (2022). https://doi.org/10.1021/acs.chemrev.1c00237

31. Z. Cai, B. Ouyang, H.-M. Hau, T. Chen, R. Giovine, K.P. Koirala, L. Li, H. Ji, Y. Ha, Y. Sun, J. Huang, Y. Chen, V. Wu, W. Yang, C. Wang, R.J. Clément, Z. Lun, G. Ceder, Nat. Energy (2024). https://doi.org/10.1038/s41560-023-01375-9

32. Y. Zeng, B. Ouyang, J. Liu, Y.-W. Byeon, Z. Cai, L.J. Miara, Y. Wang, G. Ceder, Science (2022). https://doi.org/10.1126/science.abq1346

33. Y. Chen, Z. Lun, X. Zhao, K.P. Koirala, L. Li, Y. Sun, C.A. O'Keefe, X. Yang, Z. Cai, C. Wang, H. Ji, C.P. Grey, B. Ouyang, G. Ceder, Nat. Mater. (2024). https://doi.org/10.1038/s41563-024-01800-8

34. Z. Lun, B. Ouyang, D.-H. Kwon, Y. Ha, E.E. Foley, T.-Y. Huang, Z. Cai, H. Kim, M. Balasubramanian, Y. Sun, J. Huang, Y. Tian, H. Kim, B.D. McCloskey, W. Yang, R.J. Clément, H. Ji, G. Ceder, Chem. (2020). https://doi.org/10.1016/j.chempr.2019.10.001

35. H. Ji, A. Urban, D.A. Kitchaev, D.-H. Kwon, N. Artrith, C. Ophus, W. Huang, Z. Cai, T. Shi, J.C. Kim, H. Kim, G. Ceder, Nat. Commun. (2019). https://doi.org/10.1038/s41467-019-08490-w

36. Y. Chen, X. Zhao, K. Chen, K.P. Koirala, R. Giovine, X. Yang, S. Wang, N.J. Szymanski, S. Xiong, Z. Lun, H. Ji, C. Wang, J. Bai, F. Wang, B. Ouyang, G. Ceder, Adv. Mater. (2025). https://doi.org/10.1002/adma.202416342

37. Y. He, E. Scivally, A. Shaji, B. Ouyang, Y. Zeng, Adv. Energy Mater. (2025). https://doi.org/10.1002/aenm.202403877

38. Y. Mo, S.P. Ong, G. Ceder, Chem. Mater. (2012). https://doi.org/10.1021/cm203303y

39. Y. Sun, B. Ouyang, Y. Wang, Y. Zhang, S. Sun, Z. Cai, V. Lacivita, Y. Guo, G. Ceder, Matter (2022). https://doi.org/10.1016/j.matt.2022.08.029

40. J. Wang, T. He, X. Yang, Z. Cai, Y. Wang, V. Lacivita, H. Kim, B. Ouyang, G. Ceder, Nat. Commun. (2023). https://doi.org/10.1038/s41467-023-40669-0

41. B. Ouyang, Y. Wang, Y. Sun, G. Ceder, Chem. Mater. (2020). https://doi.org/10.1021/acs.chemmater.9b04541

42. J.B. Boyce, B.A. Huberman, Phys. Rep. (1979). https://doi.org/10.1016/0370-1573(79)90067-X

43. J. Qi, S. Banerjee, Y. Zuo, C. Chen, Z. Zhu, M.L. Holekevi Chandrappa, X. Li, S.P. Ong, Mater. Today Phys. (2021). https://doi.org/10.1016/j.mtphys.2021.100463

44. D. Sun, N. Wu, Y. Wen, S. Sun, Y. He, K. Huang, C. Li, B. Ouyang, R. White, K. Huang, J. Mater. Chem. A (2025). https://doi.org/10.1039/D5TA01157D

45. G. Ceder, Y.M. Chiang, D.R. Sadoway, M.K. Aydinol, Y.I. Jang, B. Huang, Nature (1998). https://doi.org/10.1038/33647

46. T. Mueller, G. Hautier, A. Jain, G. Ceder, Chem. Mater. (2011). https://doi.org/10.1021/cm200753g

47. G. Hautier, A. Jain, H. Chen, C. Moore, S.P. Ong, G. Ceder, J. Mater. Chem. (2011). https://doi.org/10.1039/C1JM12216A

48. Z. Lun, B. Ouyang, D.-H. Kwon, Y. Ha, E.E. Foley, T.-Y. Huang, Z. Cai, H. Kim, M. Balasubramanian, Y. Sun, J. Huang, Y. Tian, H. Kim, B.D. McCloskey, W. Yang, R.J. Clément, H. Ji, G. Ceder, Cation-disordered rocksalt-type high-entropy cathodes for Li-ion batteries. Nat. Mater. **20**, 214 (2021)

49. L. Wang, N. Sunariwal, Y. He, D.-H. Kim, D.-H. Yeon, Y. Zeng, J. Cabana, B. Ouyang, Adv. Energy Mater. (2025). https://doi.org/10.1002/aenm.202404982

50. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, APL Mater. (2013). https://doi.org/10.1063/1.4812323

51. J. Reed, G. Ceder, A. Van Der Ven, Electrochem. Solid-State Lett. (2001). https://doi.org/10.1149/1.1368896

52. J. Reed, G. Ceder, Chem. Rev. (2004). https://doi.org/10.1021/cr020733x

53. O. Kononova, T. He, H. Huo, A. Trewartha, E.A. Olivetti, G. Ceder, iScience (2021). https://doi.org/10.1016/j.isci.2021.102155

54. R. Shannon, Acta Crystallogr. Sect. A (1976). https://doi.org/10.1107/S0567739476001551

55. E. Zintl, *Angewandte Chemie.* (1939)

56. C.J. Bartel, C. Sutton, B.R. Goldsmith, R. Ouyang, C.B. Musgrave, L.M. Ghiringhelli, M. Scheffler, New tolerance factor to predict the stability of perovskite oxides and halides. Sci. Adv. **5**, eaav0693 (2019)

57. I. Batatia, P. Benner, Y. Chiang, A.M. Elena, D.P. Kovács, J. Riebesell, X.R. Advincula, M. Asta, M. Avaylon and W.J. Baldwin, *arXiv preprint* arXiv:2401.00096. *(2023)*

58. H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen and C. Zeni, *arXiv preprint* arXiv:2405.04967. *(2024)*

59. A. Merchant, S. Batzner, S.S. Schoenholz, M. Aykol, G. Cheon, E.D. Cubuk, Nature (2023). https://doi.org/10.1038/s41586-023-06735-9

60. K. Choudhary, B. DeCost, NPJ Comput. Mater. (2021). https://doi.org/10.1038/s41524-021-00650-1

61. T. Xie, J.C. Grossman, Phys. Rev. Lett. (2018). https://doi.org/10.1103/PhysRevLett.120.145301

62. X. Liu, B. Ouyang, Y. Zeng, Nat. Comput. Sci. (2025). https://doi.org/10.1038/s43588-025-00769-x