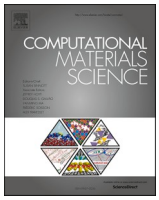
Computational Materials Science 230 (2023) 112513

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci



Data driven design of compositionally complex energy materials

Lin Wang, Zhengda He, Bin Ouyang *

Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32304, USA

ARTICLE INFO

Keywords:
Compositionally complex energy materials
Predictive synthesis
Short-range order
Machine learning

ABSTRACT

Compositionally complex materials have emerged as new frontier for sustainable energy storage and conversion. There are many unique features of compositional complex energy materials (CCEMs), include but not limited to less dependency of critical elements, the potential for enhanced ionic conduction, and the capability to prevent chemo-mechanical degradation. However, the design space of CCEMs is intricate due to higher dimensionality of compositional space, as well as convoluted interplay between long range order and short-range structures. This review aims at providing a concise summation of research frontier in CCEMs, covering aspects from synthesis, manipulation of local structures and property control. Given the rapid advancement of the battery field in recent years, this review will particularly emphasize on battery related CCEMs as a key area of representation. Furthermore, the common challenges and benefits of CCEMs will also be extrapolated to other fields of energy storage and conversion. Among the various prospects for utilizing data science in studying CCEMs, this work will primarily concentrate on the physical interpretation based on massive data generated by high-throughput computation. However, it will also encompass the cutting-edge progress of machine learning algorithms and their potential applications in the study of CCEMs.

1. Introduction

Compositionally complex materials refer to materials with multiple components and certain degree of disorder. These materials encompass not only high entropy alloys and ceramics, but also inorganic materials that have multiple principle elements. The existence of compositionally complex materials challenges the Pauling's fifth rule, e.g., the rule of parsimony [1] as well as extend the Hume-Rothery rules [2]. Such materials are usually stabilized by entropy (both configurational and vibrational), such as the case of high entropy materials and (or) special bond topology. The high entropy material field is initially driven by the great advancement of high entropy alloys [3] and structural ceramics [4]. Later, increasing attentions have been attracted for applying compositionally complex materials for clean energy applications. However, in contrast with compositionally complex structural materials, the understanding of the underlying mechanisms and the role of data science in the design of compositionally complex energy materials (CCEMs) remains notably scarce.

Regardless of the limited understanding so far for CCEM, this field is rising at horizon with multiple insightful research angles emerging, including promotion of ion conductivity [5–8], reduction of volumetric change during electrochemical cycling [9–11], as well as ease or elimination of reliance on critical metals [5,8,12]. The applications of CCEMs have spread into designing of battery materials [5,8], catalysts [13], thermoelectricity [14] and recently also into optoelectronic materials [15]. Despite the emerging promises on energy storage and conversion, the data driven design of CCEM is challenging, not only due to much larger compositional space as well as complex reaction pathway, but also the potential existence of various local structures and short-range structural features. In this review, we will summarize several key topics on designing compositionally complex materials for energy applications, as well as utilizing data science tools for exploring the structure–property-synthesis relationship of CCEMs. Particularly, we will concentrate on summarizing the important discovery for energy storage applications, and then extrapolate the underlying physics into other fields of energy applications. Moreover, it is worth noting that there are limited machine learning work focusing on CCEMs in contrast to physical interpretation from high throughput computation. This review will thus emphasize more on presenting the underlying physical intuitions but will also cover the potential opportunities in terms of machine learning. In the end, outlook for coming research frontier in this field will also be presented.

* Corresponding author.
E-mail address: bouyang@fsu.edu (B. Ouyang).

https://doi.org/10.1016/j.commatsci.2023.112513
Received 23 May 2023; Received in revised form 10 September 2023; Accepted 13 September 2023

2. Unique feature of compositionally complex energy materials

The uniqueness of CCEM can be represented by several recently established concepts. The first concept is that the disordering in CCEM can potentially facilitate ion diffusion. The absence of principal element in such materials prevents the presence of a dominating coordination environment, leading to diverse ensemble of local environments. The diverse local environment can potentially provide optimal ion diffusion network by engineering the chemical short-range order [5–8]. Such mechanistic understanding has inspired rapid development of superionic conductors as well as high-rate materials for energy storage applications [5–8]. As one example, commercialized Li-ion battery cathodes are dominated by Co and Ni based layered oxides, as layered structure with other metals will have migration issues which prohibit the stable cycle of cathodes [16]. However, when disordering and multiple element are cooperated into such materials, the ionic conduction mechanism will change and rely only on local structures [17]. As being demonstrated in Fig. 1(a), the ion diffusion is happening through local structure called "0TM" channel, while the first electrochemically active material with this type is demonstrated by $Li_{1.211}Mo_{0.467}Cr_{0.3}O_2$ (right panel of Fig. 1(a)) [14]. Moreover, with increasing number of elements, the local atomic ordering can be further optimized and lead to even better rate performance [8]. In addition to battery electrode application, it has also been found that the lattice distortion in CCEM can also enhance ionic conduction [5]. This is due to the perturbation of energy landscape because of lattice distortion, which facilitate the ion percolation with low activation barrier (Fig. 1(b)). Such a perturbation can lead to multicomponent superionic conductors that have orders of magnitude higher ionic conductivity than any of its single component counterpart, which is also shown in Fig. 1(b).

Additionally, the diverse local structures means that the functional behavior of CCEMs relies not so much on specific metals but on the collective ensemble effect and the interconnectivity of distinct bonding configurations. Consequently, the efficacy of a CCEM hinges less on critical metals, particularly when juxtaposed with typical binary or ternary energy materials [5,8]. Such characteristic introduces a new strategy for resolving supply chain apprehension concerning energy storage and conversion, particularly within the content of rapid electrification process. To give a few examples, consider the instance of commercialized Li-ion battery cathodes, which predominantly consist of cobalt (Co) and nickel (Ni) based layered oxides. These materials, however, suffer from limited production capabilities and a constrained supplier base [18]. Conversely, through the deliberate design of CCEMs, the reliance on Co and Ni can be significantly curtailed, or even eliminated entirely. For example, the reported $Li_{1.3}Mn^{2+}_{0.1}Co^{2+}_{0.1}Mn^{3+}_{0.1}Cr^{3+}_{0.1}Ti_{0.1}Nb_{0.2}O_{1.7}F_{0.3}$ has only 5 % Co and no Ni, the

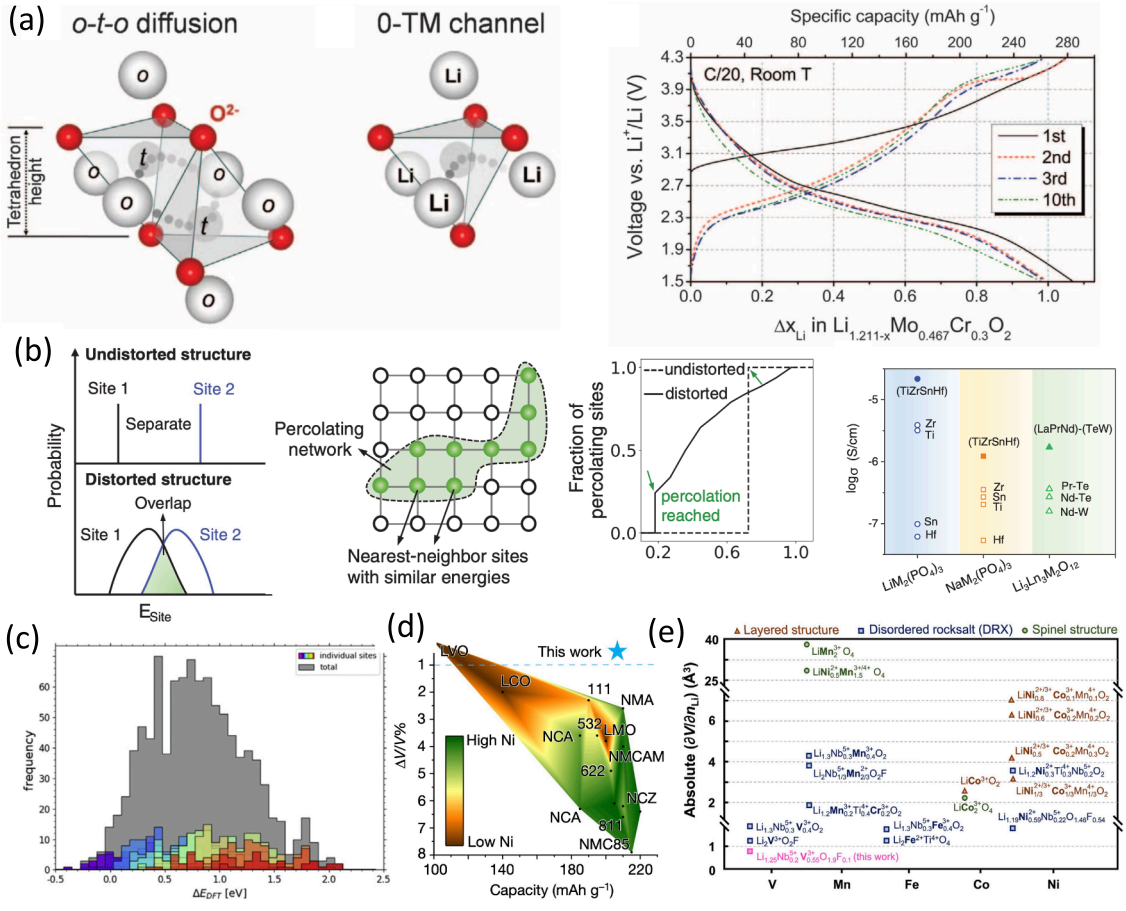


Fig. 1. (a) The diffusion pathway of Li in disordered rocksalt cathode and the critical local structure with only Li in tetrahedron vertices (0TM channel), the right panel shows the first electrochemical active material presented experimentally under such concept [17]; (b) The first panel from the left shows the schematic demonstration of the energy landscape with and without lattice distortion, the second panel from left shows the idea of percolation among lattice sites with similar energy. The third panel from the left shows the DFT calculated Li percolation for the prototype materials $LiTi_2(PO_4)_3$. The fourth panel from the left illustrates the experimental verification of fact that mixing of multiple principle elements can increase ionic conductivity by orders of magnitudes [5]; (c) The computational evidence that shows the multiple principle element effect can distribute broadly of the adsorption energy of molecules [12]; (d) Experimentally reported materials that shows almost zero volumetric change during electrochemical cycling. "This work" indicates the material composition of $LiNi_{0.8}Mn_{0.13}Ti_{0.02}Mg_{0.02}Nb_{0.01}Mo_{0.02}O_2$ [10]; (e) The table that summarize the partial mole volume change of different materials that are frequently investigated as Li-ion battery cathodes [9].

materials can surpass the capacity of commercialized battery capacity while shows the capability of fast charge in less than 10 min [8]. Moreover, it has been demonstrated by Thomas et al. [12], that by designing CCEM, it will effectively distribute the adsorption energy of molecules, shown in Fig. 1(c), which provide potential of reaching optimal catalytic performance with minimized noble metals [10].

Last but not the least, one other special feature of CCEM is the lack of collective volumetric change due to disordering. This is beneficial for many energy storage applications as less collective volumetric change can lead to release of chemo-mechanical degradation during the cycling of energy devices. Moreover, due to disordering and dilution of dominating metal redox, CCEM can be designed to show very little volumetric change. As being demonstrated by Fig. 1(d), the volumetric change of $LiNi_{0.8}Mn_{0.13}Ti_{0.02}Mg_{0.02}Nb_{0.01}Mo_{0.02}O_2$ can be greatly minimized when multiple elements are cooperated into the cation sites. Moreover, the mechanisms governing the influence of lattice disorder on volumetric changes during electrochemical cycling have been thoroughly discussed by Zhao et al. [9,11]. From such design principle, Zhao et al., have both predicted and experimentally synthesized several multi-component materials with low volumetric change during electrochemical cycling in the compositional space of Li-V-Nb-O-F [11]. Moreover, the idea of enhancing disorder in reducing volumetric change is also further supported by Konuma et al. [19], materials engineered within the Li-V-Ti-O-F space shows extraordinary electrochemical performance.

3. High throughput screening and synthetic accessibility

Accompanied by the great potential of using compositional complex energy materials (CCEM), understanding of synthetic accessibility of such materials became critical as such materials usually have larger chemical and synthesis space due to its multicomponent composition and disordered structure. The exploration of synthetic accessibility of CCEMs can be decomposed into three parts, shown in Fig. 2, e.g., (a) the identification of materials with low thermodynamic driving force for decomposition; (b) the design of viable precursors that tailors the reaction pathway; (c) the choose of appropriate synthesis conditions.

To determine the thermodynamic driving force for decomposition, compositional phase diagrams are usually utilized to be coupled with high throughput density function theory calculations [8,20]. The establishment of compositional phase diagrams is mathematically convex hull optimization, which involves the enumeration of all known materials, particularly ground states in a certain compositional space. The convex envelope of lowest energy phases from density function theory (DFT) calculation is, by definition, the ground state of the compositional space at 0 K. Therefore, the thermodynamic driving force of such phase to decompose can then be used to estimate relative metastability of materials. Such compositional phase diagram is not only useful to estimate materials stability at low temperature and room temperature but can also be predictive for estimating materials stability at high temperature with simple assessment for entropy [8,20]. The capability of transferring compositional phase diagram across different temperature ranges and other state variables can be supported by systematic evaluations of all reported inorganic materials from inorganic crystallography structure database (ICSD) [21,22]. With such theoretical foundation established, the compositional phase diagram has been applied broadly and enabled quick assessment of synthetic accessibility.

Together with the development of automated high throughput computational workflow [23–25], the past decade has witnessed many work that enumerate the complete combinatorial space across the periodic table for various types of materials [8,20,26,27]. Taking two high throughput screening heatmaps as examples, the NASICON stability heatmap [20] (Fig. 3 (a)) and the high entropy disordered rocksalt (HE-DRX) heatmap [8] (Fig. 3(b)). In the NASICON case, a high throughput screening of 3881 NASICON compositions with 21 metals and three types of polyanions have been performed that screens all possible chemistry that can potentially be introduced into NASICON structure and make it a superionic Na-ion conductor. Such a screen predicts 396 new NASICON compositions that is likely to be synthetic accessible, which triples the reported NASICON compositions [20]. Moreover, for the case of HE-DRX, 7965 compounds from 23 cations in a prototype formula $Li_{1.3}TM_{0.7}O_{1.7}F_{0.3}$ are screened, which leads to the discovery that high entropy will generally lead to less chemical short-range order in DRX structure. The capability of applying such approaches provide a quick solution to the task of finding a needle under the sea, as hypothetic compounds with low E_{hull} value is usually the minority of the complete dataset [8,20,26,27]. Therefore, the phase diagram tools greatly facilitate the quick identification of promising compositions.

In addition to the thermodynamic driving force, precursor is the most widely used experimental handle for attempting new compounds [28–38]. Different choices of precursor set up the starting points of energy landscape for chemical reaction [30]. Consequently, different precursors lead to the formation of various intermediate states before reaching the thermodynamic ground states [30,31,36–38]. The appearance of different intermediate states thus offer potential of "freezing" a metastable state by stopping the chemical reaction or reducing the kinetic reaction rate. In contrast with the organic synthesis, where retrosynthesis is well established [39] so that many reaction pathways and reaction intermediates can be well predicted, the reaction pathway and potential intermediate is largely unknown for inorganic synthesis.

Recently, theoretical hypothesis has emerged that predicts the re-

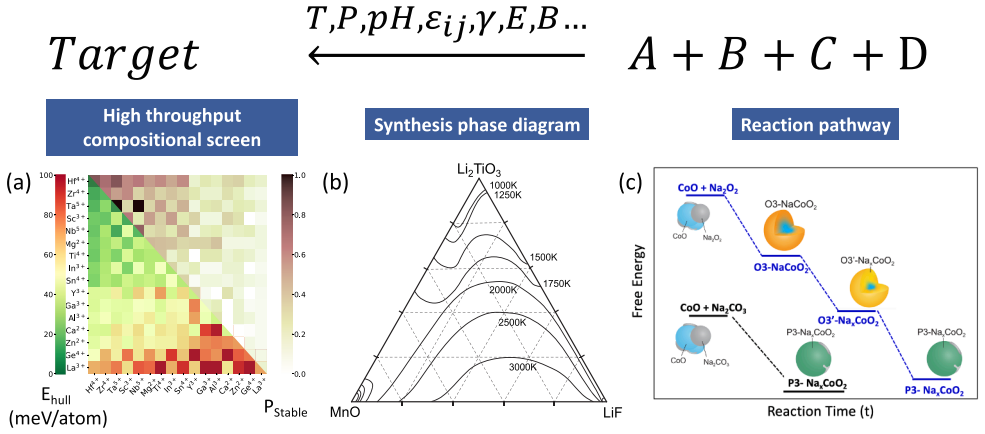


Fig. 2. Schematics of the computational driven inverse design of CCEM; (a) High throughput materials screening; (b) Phase diagram cooperating realistic synthesis conditions; (c) Prediction of chemical reaction pathways.

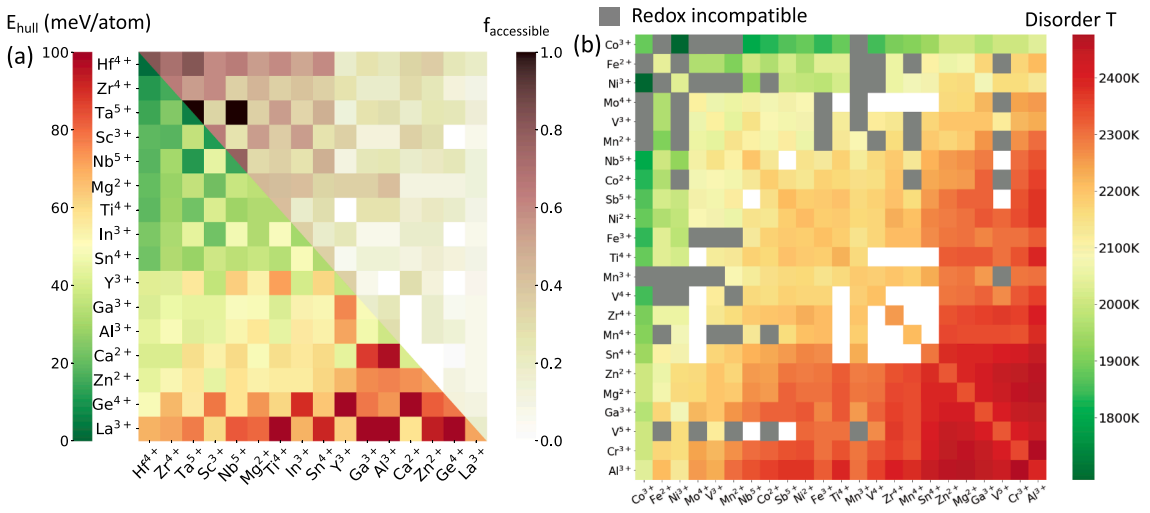


Fig. 3. Examples of element compatibility map of (a) NASICON [20] with prototype formula of $Na_xM_yM'_{2-y}(AO_4)_z(BO_4)_{3-z}$. The colormap on bottom left triangle indicates the computed E_{hull} values while the colormap on top right tri-angle indicates the probably of finding an synthetic accessible NASICON composition ($E_{hull} - S_{Ideal}T(1000\ K) \leq 0$) with two specific metals; (b) High entropy disordered rocksalt [8] with six transition metals in the prototype formula $Li_{1.3}M_{0.7}O_{1.7}F_{0.3}$. The color map indicated the average disordering temperature, which is defined as E_{hull} value of the special quasi random structure (SQS) at specific temperature.

action pathway with different choice of precursors. One example of such hypothesis is the "max ΔG" theory [30,31,37,38], which propose that many solid-state reactions will first form the phase with largest thermodynamic driving force at the convex hull. As being shown by Fig. 4 (a), the "max ΔG" theory assumes that solid-state reaction usually nucleate at interfaces with small thickness so that the bulk stoichiometry will be irrelevant to the product generated after solid-state reaction. Instead, the first phase formed will usually be the phase that has the deepest formation energy in the reaction convex hull. There have been quite a few works which support the validity of "max ΔG" theory at least for several categories of solid-state reactions, which are demonstrated in Fig. 4(b)–(e). It has been found out that for all four groups of synthesis

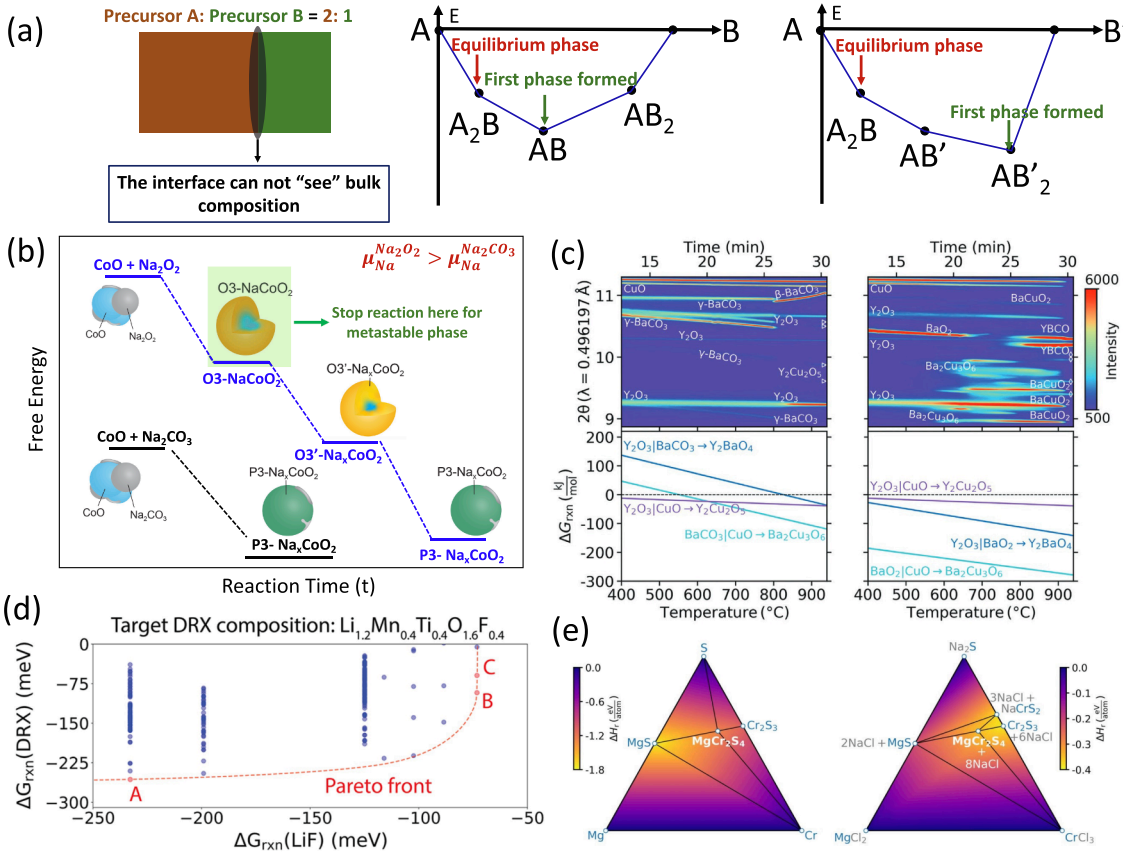


Fig. 4. (a) Schematic of the "max-Δ G" hypothesis taking a stoichiometry of A:B = 2:1 as one example; The precursor tailored reaction pathway at different synthesis conditions for (b) The synthesis of $NaCoO_2$ [27]; (c) The synthesis of $YB_2Cu_3O_{6+x}$ [28]; (d) The synthesis of $Li_{1.2}Mn_{0.4}Ti_{0.4}O_{1.6}F_{0.4}$ [34]; (e) The synthesis of $MgCr_2S_4$ [35].

attempts, the change of precursor will effectively tailor the synthetic reaction pathway. More specifically, in the case of synthesizing $NaCoO_2$ (Fig. 4(b)), the metastable O3-$NaCoO_2$ will be formed if switching from the classic Na_2CO_3 precursor to Na_2O_2 or Na_2O, as the elevation of Na chemical potential will effective change the synthesis convex hull which leads to the formation of Na rich O3 phase rather than the ground state [30]. Similar observation has also been observed by Miura et al. [31] in the case of synthesizing perovskite based $YB_2Cu_3O_{6+x}$, it has been find that the utilization of BaO_2 rather than $BaCO_3$ will effectively increase the reactivity, which enhance the pair-wise reaction and accelerate greatly of the formation of targeted $YB_2Cu_3O_{6+x}$ phase with neglected impurities (Fig. 4(c)). On the other hand, the "max Δ G" theory is also used as a computational descriptor for high throughput screening of potential precursors, Szymanski and Zeng et al. [37] have designed a variety of potential precursors with synthetic verifications (Fig. 4(d)), which paves in-depth understanding of principles for maximizing F solubility in disordered rocksalt based Li-ion battery cathodes [37]. Moreover, Miura et al. [31] have found that the utilization of elemental

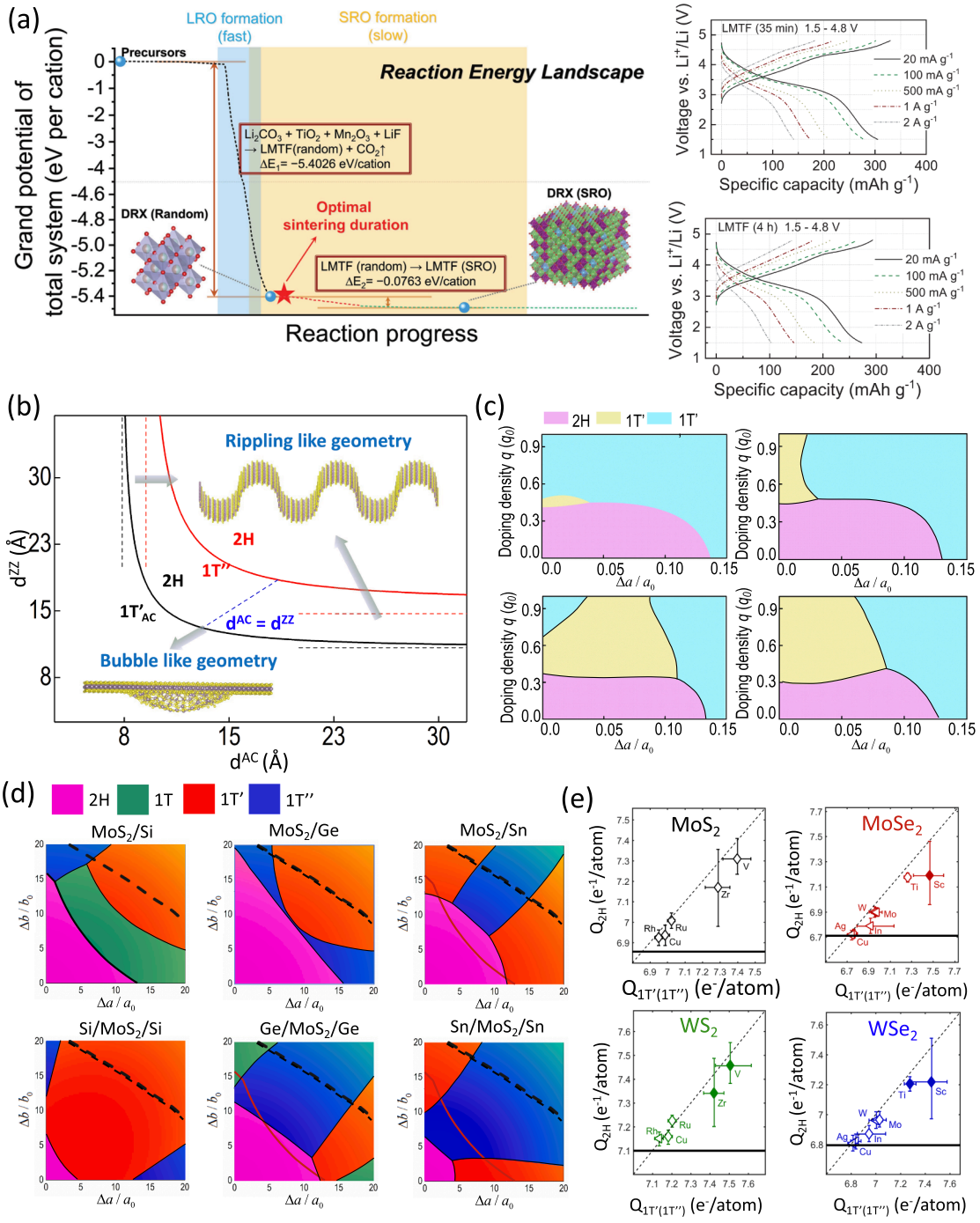


Fig. 5. (a) The impact of cooling rate in cation ordering and electrochemical properties of disordered rocksalt cathode materials [40]; (b) The impact of curvature on phase stability of low dimensional transition metal dichalcogenides [42]; (c) The impact of biaxial strain coupled with electron doping on phase stability of low dimensional transition metal dichalcogenides [46]; (d) The impact of interfacial transfer coupled with biaxial strain on phase stability of low dimensional transition metal dichalcogenides [44]; (e) The impact of interfacial charge transfer on phase stability of low dimensional transition metal dichalcogenides [43].

phase as precursor, in contrast to the conventional binary sulfide and chloride precursors, will establish an "empty" convex hull that get rid of many typical impurities compared with the standard recipe of synthesizing $MgCr_2S_4$ [38], (Fig. 4(e)).

The synthetic accessibility of CCEM is also highly depending on synthesis conditions. Depends on the type of synthesis method, computational modeling can provide instructions on actual synthesis conditions through quantitative or rudimentary phase diagrams [6,30,40–48]. The construction of synthesis phase diagram can be regarded as a Legendre transformation of the compositional phase diagram that is directly computed from DFT calculations [22,49]. It will thus require the estimation of remanent contribution of real synthesis conditions to the free energy of the synthesis reaction. Some of the representative synthesis conditions can be pH, redox potential, and particle size in the case of aqueous synthesis; epitaxy strain, interfacial charge transfer and curvature in the case of thin film growth; as well as temperature and mechanical deformation in the case of solid-state synthesis. For all these types of the conventional synthesis method, phase diagrams under realistic synthesis conditions have shown great potential in calibrating the synthesis window of specific compounds.

The synthesis phase diagram can be demonstrated by the examples in Fig. 5. As being shown by Fig. 5(a), the cooling speed can be used as an effective tool for tailoring the disordered phases for solid state synthesis of disordered rocksalt Li-ion battery cathode materials [37]. Different cooling rate can lead to very different rate performance as shown in the right panel of Fig. 5(a). In addition to typical synthesis phase diagram, exotic phase diagrams can also be established for exploring novel synthesis space. Specifically, the influence of curvature has been explored as potential synthesis handles during epitaxy growth (Fig. 5(b)) of transition metal dichalcogenides have been investigated [39] as shown in Fig. 5(b); the elastic strain, on the other hand, can be coupled with electrostatic gating to control the 2H → 1 T (1 T' and 1 T'') structural phase transformation [43] (Fig. 5(c)). Moreover, elastic strain can also be coupled with different kind of interfaces, including low dimensional interfaces arise from heterostructures [41] (Fig. 5(d)) as well as substrate induced interfacial charge transfer [40] (Fig. 5(e)).

4. Chemical short-range order and local structure engineering

One unique structure feature of CCEM is that the property of such materials is not only related to periodic bond topology, which is often referred to as long range order, but also tied up with the existence of various local structures, which correspond to the chemical short-range order (CSRO) [50]. CSRO is widely observed in all kinds of CCEM, ranging from alloys [51–55] to ceramics [56–58]. The nature of CSRO is essentially the remanent of pair preference as in the case of long range ordered structure. As can be derived from Patterson function [50,59], a Fourier transformation of the occupancy of the corresponding local

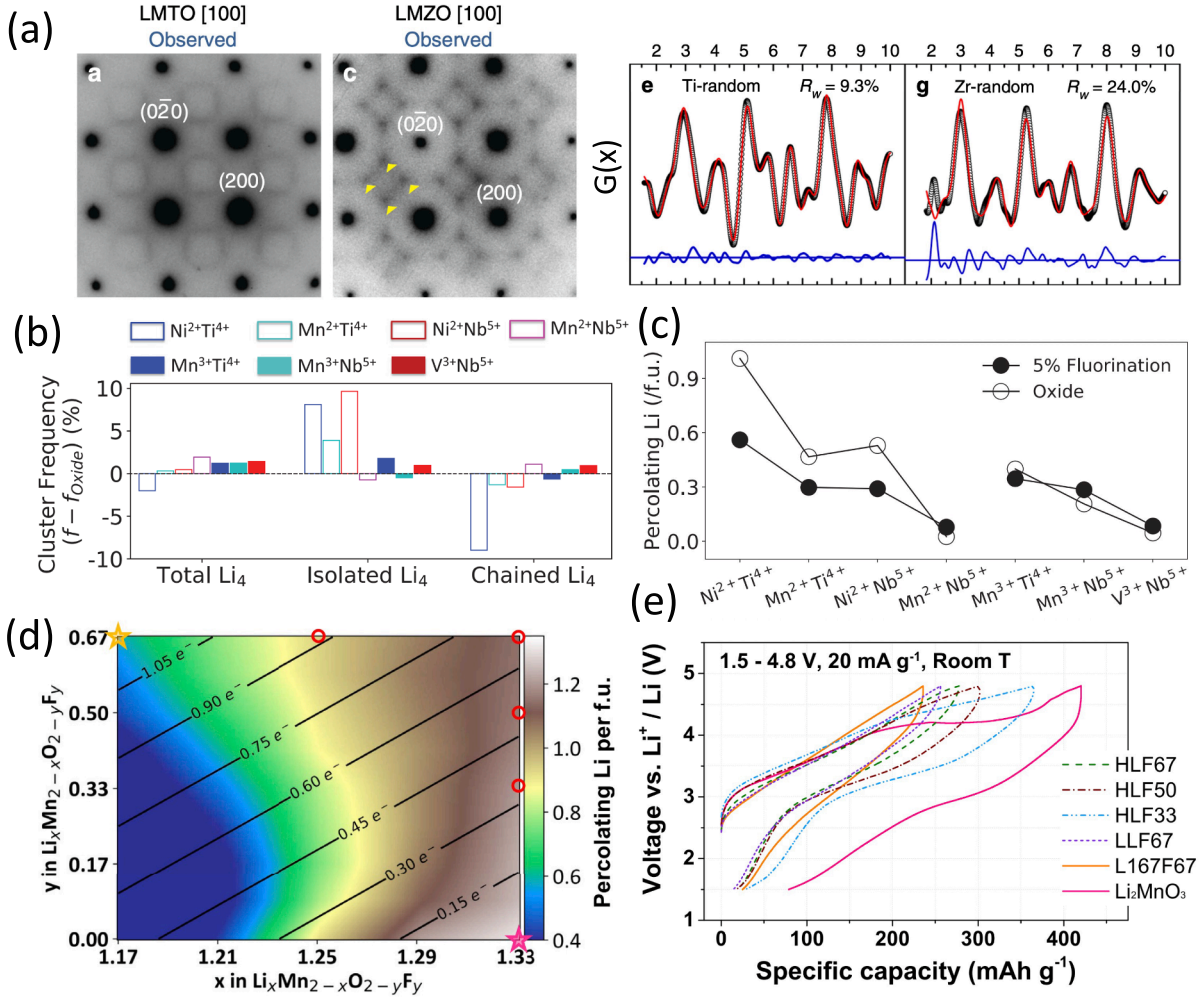


Fig. 6. (a) Observed CSRO from disordered rocksalt type oxide and oxyfluoride from both SAED and neutron diffraction pair distribution analysis [56]; (b) CSRO and local structure analysis from large scale Monte Carlo sampling for local tetrahedron coordination [6]; (c) The percolating amount of Li of different oxides and oxyfluorides with different CSRO and local structures [6]; (d) The percolation map at the complete compositional space of Li-Mn-O-F with Mn at + 2/+3/+4 [7]; (e) The experimental measured initial capacity which is consistent with the theoretical prediction shown in (d) [7].

structure will lead to the observed diffusive scattering features, which is often regarded as the "fingerprint" of CSRO.

The investigation of CSRO is challenging in both theory and experiment. In experiment, the CSRO features collected from electron, neutron diffraction or X-ray diffraction are usually convoluted with information of both structural feature and corresponding probability, while the elements that mixed can also have very similar scattering factors [51,52,60]. As a result, the interpretation of characterization features will become challenging. On the other hand, theoretical modeling of CSRO requires efficient sampling of large configurational space [61]. Particularly for CCEM, such a space can be too large to be enumerated with DFT calculations [7,8,61].

Recently, the emerging of disordered battery electrodes have catalyzed the development of modeling tools and theoretical understanding of CSRO in CCEM. As being indicated in Fig. 6(a), difference CSRO features have been discovered and reported in cation disordered type of oxides and oxyfluorides with a FCC type cation sublattice [56]. Such CSRO is proved to be sensitively related to electrochemical performance [56] thus enhanced the developments of simulation tools for reproducing that theoretically. The capability of modeling Hamiltonian across high dimensional compositional space enables in-depth understanding of CSRO structures [7,8,61]. With sparse cluster expansion models that can capture high entropy compositional space up to 10 elements [8,61–67], we can develop atomistic modelling via Monte Carlo sampling with size equivalent to a few tens to hundreds nanometers [6–8,68] and the capability of obtaining converged statistics as shown in Fig. 6(b and c) [6]. Such sampling enables various study on ion percolating network as well as the consequential electrochemical performance due to CSRO [6–8]. As shown by Fig. 6(d), the percolation map, which is essentially the connection of specific local structure that have low diffusion barrier can be well captured by Monte Carlo simulations [6,7], which lead to consistent predictions of initial capacity of battery, shown in Fig. 6(e), when we tailor the CSRO through compositional change.

In addition to CSRO, another emerging concept is that some local structures that are loosely bonded with the rest of the crystalline framework can also be crucial for energy storage and conversion properties. Such local structures are widely discussed in the field of hybrid organic/inorganic materials [69] and metal organic frameworks [70], as in both category of materials we have the existence of molecules that is more flexible in terms of rotation and even migration. Such rotational degree of freedom can lead to more tunable space of functionality [69,70]. Recently, such doping strategy has been applied even in pure inorganic crystalline framework, particularly for the application of making superionic conductors. The origin of those local behavior is usually phrased as "dynamic disorder", which has been thoroughly discussed with respect to the application as superionic Akali metal conductors since 1990s (shown by Fig. 7a) [71]. More recently, the role of cluster ion has been widely studied as a new route for engineering properties related to clean energy applications [71–77].

To give a few representative examples, it has been proposed in many recent works that a highly rotating polyanion framework can lead to enhanced ion diffusion in materials. There has been many discussions concentrating on whether the "paddle wheel" mechanism will potentially enhance ionic conductivity [71–78]. However, it is worth noting that it is not easy to have solid evidence about the existence or absence of paddle wheel effect, from both computation and characterization. The challenge major comes from the fact that quantifying the paddle wheel mechanism requires the separation of correlation between ion diffusion and polyanion rotation at both time and spatial scale. However, in both simulation and experiment, this is not easy and requires careful setup. Moreover, it is worth noting that the paddle wheel effect of cluster ion is sometimes decoupled with diffusive hopping as they are in different time scale. Taking $Li_6PS_5(BH_4)$ as an example, the rotating of cluster ion $(BH_4)^-$ is very pronounced as being shown by the trajectory in Fig. 7(b), which is visualized from 10 ps of AIMD simulations. However, when decompose such rotation into more elementary bond bending and bond stretching mode as shown in Fig. 7(c), it has been found that the

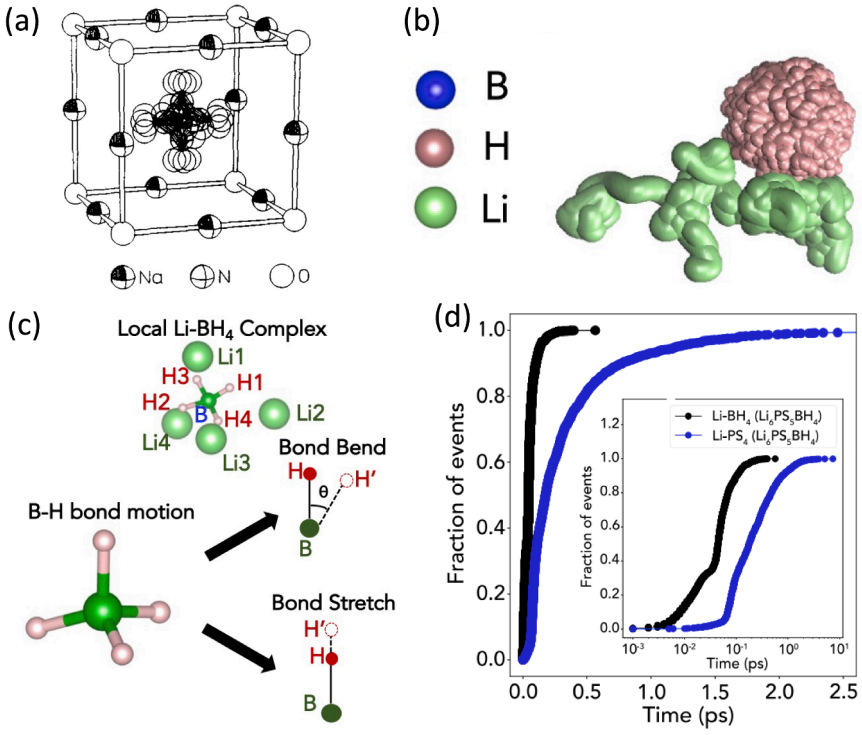


Fig. 7. (a) Demonstration of the dynamic disorder of nitrate ion in H-$NaNO_3$ [71]; (b) AIMD simulated trajectory of dynamic disorder of $(BH_4)^-$ in $Li_6PS_5(BH_4)$; (c) Decomposition of the rotational motion into bond bending and bond stretching; (d) The dynamic correlation between Li motion with $(BH_4)^-$ and $(PS_4)^{3-}$.

enhanced ionic conductivity is more related to the weak interaction between $(BH_4)^-$ and the rest of the framework rather than paddle wheel effect [78,79]. According to the trajectory analysis as shown in Fig. 7(d), the weak interaction between $(BH_4)^-$ and the rest of the framework enables less residual time between $(BH_4)^-$ and the framework. The absence of paddle wheel mechanism in this case is simply because the rotation degree of freedom has much higher frequency compared with the ion hopping attempt frequency. Moreover, freezing the rotating ion could cause troubles of analyzing such phenomenon as both the rotation degree of freedom and vibrational degree of freedom are frozen with such set up. Therefore, the lowering of ionic conductivity can be not identified as the losing of paddle wheel effect. In conclusion, novel ways of performing selective dynamic simulation are urged to decouple the multiple dynamic process in such phenomenon.

5. Machine learning and beyond

Benefiting from the Moore's law, the computational capability has scaled exponentially compared with a few decades ago. Even with quantum chemistry level of efficiency, it is relative trivial to compute hundreds or thousands of materials from a few weeks to a few months. As a result of that, we observe the rapid growth of materials genome materials database in the past decades, which can be represented by Materials Project [80], OQMD [81], A-flow [82] in US, as well as NOMAD [83] and MARVEL [84] in Europe. The existence and rapid expansion of materials dataset thus catalyze the development of data science and machine learning protocols. Particularly for CCEMs, the major challenging for machine learning is to (1) Develop effective encoding algorithms that can capture the energy fluctuation from subtle structural and compositional change of CCEMs; (2) Harvest and establish non-biased dataset to overcome the scarcity of CCEMs data in current generic materials genome database. For addressing both challenging, there is very little work so far focusing on CCEMs. However, there are indeed plausible development from other field of computational materials science that can potentially benefit the machine learning of CCEMs. Such emerging efforts will be summarized in the following paragraphs.

For encoding algorithms, it has been shown that compositional models that lack structural information will fail greatly in predicting decomposition pathway as it can only capture the average chemical potential at certain composition [85]. Such estimation will perform badly at a compositional space with rich structural evolution due to subtle compositional change or pure polymorphic transition. Such benchmark has been demonstrated by Bartel et al. [83], as shown in Fig. 8 (a). Roost is the best compositional model among six types of composition based convoluted neural network model benchmarked. On the other hand, the crystal graph convolutional neural network [85,87], which captures not only compositional, but also structural information of crystalline materials, turn out to be much better at capturing decomposition energy than Roost, as shown in Fig. 8(a). More specifically, CGCNN captures the decomposition energy much better compared with Roost not only for lower MAE, but also the capability of getting rid of predictions with huge discrepancy of predicted energy from the DFT computed energy [85]. Therefore, it has been confirmed that with consideration of structural features, the prediction of phase decomposition can be greatly improved. However, it is also worth mentioning that the state-of-art MAE or RMSE of such global energy models is around 30–50 meV/atom in general [85,87,88], which is comparable with the configurational energy range of CCEM with different atomic configurations. Therefore, it remains questionable that such model will necessarily be useful when discussing about phase transformation across different disordering states. To make further advances in utilizing machine learning energy models for CCEMs, the intrinsic limitation of CNN based methods need to be better understood and benchmarked with customized dataset, rather than the dataset from above-mentioned generic database, which typically have scarce data of CCEMs.

Beyond the idea of machine learning, there are also other scientific challenges to solve rather than prediction-driven data science. One important challenge is data assimilation and processing. There are majorly two challenges in this theme: (a) Harvesting data from scientific literature; b) Combine and use data with different source and fidelity. Natural language processing (NLP) has emerged in the past few years that harvesting data from millions of scientific literatures

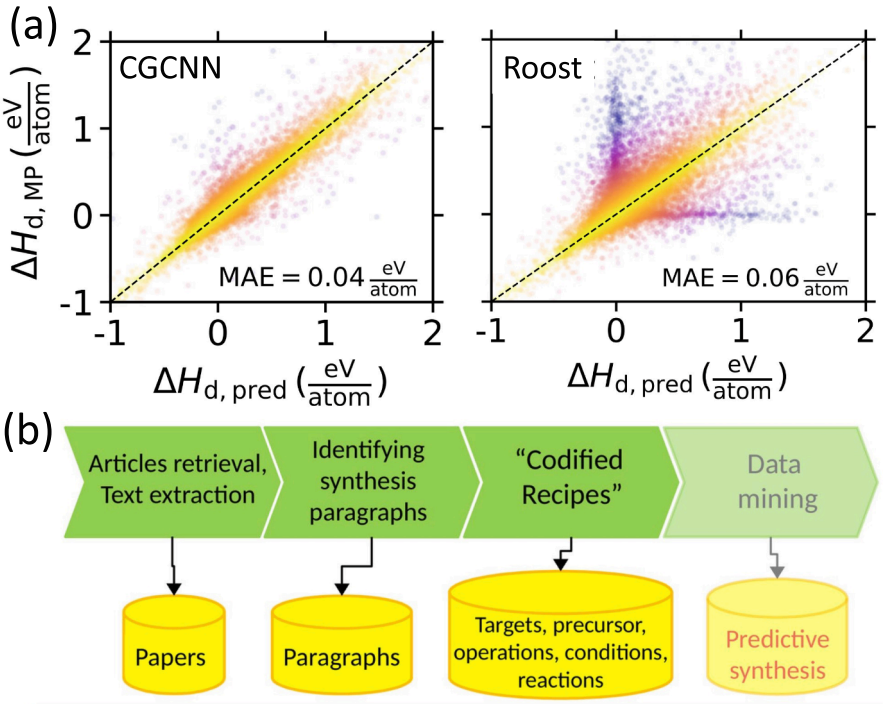


Fig. 8. (a) Comparison between crystal graph convolutional neural network (CGCCN) and the best composition-based neural network model among 6 benchmarked models [85]; (b) Natural language processing (NLP) pipeline that lead to the establishment of SynTERRA inorganic synthesis database [86].

[36,86,89–92]. With such experimental database from NLP, while the data assimilation pipeline demonstrated in Fig. 8(b), the predictive synthesis can be greatly enhanced with experimental input [86]. Moreover, the existence of data with different source, accuracy and references also introduce difficulty for data analysis. Transfer learning techniques will be suitable for blending data with different fidelity [93,94], moreover, it helps the establishment of pre-trained machine learning model that is capable of accelerating the materials discovery and prediction [88].

6. Outlook and summary

The realm of Compositionally Complex Energy Materials (CCEMs) holds immense promise and potential, especially in light of the substantial material requirements to achieve carbon neutrality. Drawing upon the outlined prospects, there are two significant impending challenges within the domain of data driven CCEM design. Firstly, amidst the rapid evolution of materials genome databases, numerous databases now contain millions of distinct materials [80–82,84]. However, it is worth mentioning that the known materials are still very sparse compared with the global space of materials [95]. This holds particularly true for CCEMs, where current state-of-the-art material genome databases exhibit limited or even non-existent data from both theoretical predictions and experimental validations. This underscores the pressing need for the development of robust data infrastructure alongside automated data mining workflows.

Moreover, the exploration territory for CCEMs is vast, encompassing not only the chemical space but also the synthesis space. To expedite progress, the advancement of tools facilitating swift iterations between computational predictions and experimental validations becomes paramount. The amalgamation of robotic synthesis techniques and algorithmic decision-making frameworks emerges as a pivotal step towards closing the loop on the autonomous discovery and exploration of CCEMs [96].

In conjunction with the challenges, a multitude of unexplored opportunities await discovery within the realm of CCEMs. Specifically, one notable advantage of CCEMs originates from the diverse local bonding environments, which has the potential to catalyze numerous applications extending beyond the scope of energy storage and catalysts, as highlighted in this review. To fully unlock this potential, meticulous design and robust justifications will be essential to extend this concept to a broader spectrum of energy storage and conversion applications [14,15]. Furthermore, the distinct property profile of CCEMs, derived largely from the distribution of local structures, offers the exciting prospect of diminishing reliance on critical metals. This potential avenue holds promise for alleviating concerns surrounding critical metal supply during the electrification process. Furthermore, this inherent capability may spur the development of predictive synthesis theories tailored specifically for CCEMs, while also driving the innovation of novel experimental setups that enable rapid synthesis and expedited iteration cycles.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by startup funding from Florida State University.

References

[1] Nature 148 (1941) 677–677.
[2] R. Hultgren, L. Tarnopol, Nature 141 (1938) 473–474.
[3] E.P. George, D. Raabe, R.O. Ritchie, Nat. Rev. Mater. 4 (2019) 515–534.
[4] C. Oses, C. Toher, S. Curtarolo, Nat. Rev. Mater. 5 (2020) 295–309.
[5] Y. Zeng, B. Ouyang, J. Liu, Y.-W. Byeon, Z. Cai, L.J. Miara, Y. Wang, G. Ceder, Science 378 (2022) 1320–1324.
[6] B. Ouyang, N. Artrith, Z. Lun, Z. Jadidi, D.A. Kitchaev, H. Ji, A. Urban, G. Ceder, Adv. Energy Mater. 10 (2020) 1903240.
[7] Z. Lun, B. Ouyang, Z. Cai, R.J. Clément, D.-H. Kwon, J. Huang, J.K. Papp, M. Balasubramanian, Y. Tian, B.D. McCloskey, H. Ji, H. Kim, D.A. Kitchaev, G. Ceder, Chem 6 (2020) 153–168.
[8] 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, Nat. Mater. 20 (2021) 214–221.
[9] X. Zhao, Y. Tian, Z. Lun, Z. Cai, T. Chen, B. Ouyang, G. Ceder, Joule 6 (2022) 1654–1671.
[10] R. Zhang, C. Wang, P. Zou, R. Lin, L. Ma, L. Yin, T. Li, W. Xu, H. Jia, Q. Li, S. Sainio, K. Kisslinger, S.E. Trask, S.N. Ehrlich, Y. Yang, A.M. Kiss, M. Ge, B.J. Polzin, S. J. Lee, W. Xu, Y. Ren, H.L. Xin, Nature 610 (2022) 67–73.
[11] X. Zhao, G. Ceder, Joule 6 (2022) 2683–2685.
[12] T.A.A. Batchelor, J.K. Pedersen, S.H. Winther, I.E. Castelli, K.W. Jacobsen, J. Rossmeisl, Joule 3 (2019) 834–845.
[13] Y. Sun, S. Dai, Sci. Adv. 7 (2021) eabg1600.
[14] B. Jiang, Y. Yu, J. Cui, X. Liu, L. Xie, J. Liao, Q. Zhang, Y. Huang, S. Ning, B. Jia, B. Zhu, S. Bai, L. Chen, S.J. Pennycook, J. He, Science 371 (2021) 830–834.
[15] M.C. Folgueras, Y. Jiang, J. Jin, P. Yang, Nature (2023).
[16] J. Reed, G. Ceder, Chem. Rev. 104 (2004) 4513–4534.
[17] J. Lee, A. Urban, X. Li, D. Su, G. Hautier, G. Ceder, Science 343 (2014) 519–522.
[18] X. Fu, D.N. Beatty, G.G. Gaustad, G. Ceder, R. Roth, R.E. Kirchain, M. Bustamante, C. Babbitt, E.A. Olivetti, Environ. Sci. Tech. 54 (2020) 2985–2993.
[19] I. Konuma, D. Goonetilleke, N. Sharma, T. Miyuki, S. Hiroi, K. Ohara, Y. Yamakawa, Y. Morino, H.B. Rajendra, T. Ishigaki, N. Yabuuchi, Nat. Mater. 22 (2023) 225–234.
[20] B. Ouyang, J. Wang, T. He, C.J. Bartel, H. Huo, Y. Wang, V. Lacivita, H. Kim, G. Ceder, Nat. Commun. 12 (2021) 5752.
[21] W. Sun, S.T. Dacek, S.P. Ong, G. Hautier, A. Jain, W.D. Richards, A.C. Gamst, K. A. Persson, G. Ceder, Sci. Adv. 2 (2016) e1600225.
[22] M. Aykol, S.S. Dwaraknath, W. Sun, K.A. Persson, Sci. Adv. 4 (2018) eaaq0148.
[23] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K.A. Persson, G. Ceder, Comput. Mater. Sci 68 (2013) 314–319.
[24] K. Mathew, J.H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I.-H. Chu, T. Smidt, B. Bocklund, M. Horton, J. Dagdelen, B. Wood, Z.-K. Liu, J. Neaton, S.P. Ong, K. Persson, A. Jain, Comput. Mater. Sci 139 (2017) 140–152.
[25] A. Jain, S.P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, K.A. Persson, Concurr. Comput.: Pract. Exp. 27 (2015) 5037–5059.
[26] W. Sun, C.J. Bartel, E. Arca, S.R. Bauers, B. Matthews, B. Orvañanos, B.-R. Chen, M. F. Toney, L.T. Schelhas, W. Tumas, J. Tate, A. Zakutayev, S. Lany, A.M. Holder, G. Ceder, Nat. Mater. 18 (2019) 732–739.
[27] G. Hautier, C.C. Fischer, A. Jain, T. Mueller, G. Ceder, Chem. Mater. 22 (2010) 3762–3767.
[28] A. Navrotsky, Proc. Natl. Acad. Sci. 101 (2004) 12096–12101.
[29] S. Widgeon, G. Mera, Y. Gao, S. Sen, A. Navrotsky, R. Riedel, J. Am. Ceram. Soc. 96 (2013) 1651–1659.
[30] M. Bianchini, J. Wang, R.J. Clément, B. Ouyang, P. Xiao, D. Kitchaev, T. Shi, Y. Zhang, Y. Wang, H. Kim, M. Zhang, J. Bai, F. Wang, W. Sun, G. Ceder, Nat. Mater. 19 (2020) 1088–1095.
[31] A. Miura, C.J. Bartel, Y. Goto, Y. Mizuguchi, C. Moriyoshi, Y. Kuroiwa, Y. Wang, T. Yaguchi, M. Shirai, M. Nagao, N.C. Rosero-Navarro, K. Tadanaga, G. Ceder, W. Sun, Adv. Mater. 33 (2021) 2100312.
[32] M. Wen, E.W.C. Spotte-Smith, S.M. Blau, M.J. McDermott, A.S. Krishnapriyan, K. A. Persson, Nat. Comput. Sci. 3 (2023) 12–24.
[33] D. Barter, E.W. Clark Spotte-Smith, N.S. Redkar, A. Khanwale, S. Dwaraknath, K. A. Persson, S.M. Blau, Dig. Disc. 2 (2023) 123–137.
[34] M.J. McDermott, S.S. Dwaraknath, K.A. Persson, Nat. Commun. 12 (2021) 3097.
[35] A. Wustrow, G. Huang, M.J. McDermott, D. O'Nolan, C.-H. Liu, G.T. Tran, B. C. McBride, S.S. Dwaraknath, K.W. Chapman, S.J.L. Billinge, K.A. Persson, K. Thornton, J.R. Neilson, Chem. Mater. 33 (2021) 3692–3701.
[36] H. Huo, C.J. Bartel, T. He, A. Trewartha, A. Dunn, B. Ouyang, A. Jain, G. Ceder, Chem. Mater. 34 (2022) 7323–7336.
[37] N.J. Szymanski, Y. Zeng, T. Bennett, S. Patil, J.K. Keum, E.C. Self, J. Bai, Z. Cai, R. Giovine, B. Ouyang, F. Wang, C.J. Bartel, R.J. Clément, W. Tong, J. Nanda, G. Ceder, Chem. Mater. 34 (2022) 7015–7028.
[38] A. Miura, H. Ito, C.J. Bartel, W. Sun, N.C. Rosero-Navarro, K. Tadanaga, H. Nakata, K. Maeda, G. Ceder, Mater. Horiz. 7 (2020) 1310–1316.
[39] E.J. Corey, Angew. Chem. Int. Ed. Eng. 30 (1991) 455–465.
[40] Z. Cai, Y.-Q. Zhang, Z. Lun, B. Ouyang, L.C. Gallington, Y. Sun, H.-M. Hau, Y. Chen, M.C. Scott, G. Ceder, Adv. Energy Mater. 12 (2022) 2103923.
[41] H. Kim, D.-H. Kwon, J.C. Kim, B. Ouyang, H. Kim, J. Yang, G. Ceder, Chem. Mater. 32 (2020) 4312–4323.
[42] B. Ouyang, P. Ou, J. Song, Adv. Theory Simul. 1 (2018) 1800003.
[43] B. Ouyang, S. Xiong, Y. Jing, NPJ 2D Mater. Appl. 2 (2018) 13.
[44] B. Ouyang, S. Xiong, Z. Yang, Y. Jing, Y. Wang, Nanoscale 9 (2017) 8126–8132.
[45] B. Ouyang, P. Ou, Y. Wang, Z. Mi, J. Song, PCCP 18 (2016) 33351–33356.

[46] B. Ouyang, G. Lan, Y. Guo, Z. Mi, J. Song, Appl. Phys. Lett. 107 (2015), 191903.
[47] S. Patil, D. Darbar, E.C. Self, T. Malkowski, V.C. Wu, R. Giovine, N.J. Szymanski, R. D. McAuliffe, B. Jiang, J.K. Keum, K.P. Koirala, B. Ouyang, K. Page, C. Wang, G. Ceder, R.J. Clément, J. Nanda, Adv. Energy Mater. 13 (2023) 2203207.
[48] J. Wang, B. Ouyang, H. Kim, Y. Tian, G. Ceder, H. Kim, J. Mater. Chem. A 9 (2021) 18564–18575.
[49] R.K.P. Zia, E.F. Redish, S.R. McKay, Am. J. Phys 77 (2009) 614–622.
[50] J.M. Cowley, J. Appl. Phys. 21 (1950) 24–30.
[51] R. Zhang, S. Zhao, C. Ophus, Y. Deng, S.J. Vachhani, B. Ozdol, R. Traylor, K. C. Bustillo, J.W. Morris, D.C. Chrzan, M. Asta, A.M. Minor, Sci. Adv. 5 (2019) eaax2799.
[52] R. Zhang, S. Zhao, J. Ding, Y. Chong, T. Jia, C. Ophus, M. Asta, R.O. Ritchie, A. M. Minor, Nature 581 (2020) 283–287.
[53] P.C. Clapp, S.C. Moss, Phys. Rev. 142 (1966) 418–427.
[54] P.C. Clapp, S.C. Moss, Phys. Rev. 171 (1968) 754–763.
[55] S.C. Moss, P.C. Clapp, Phys. Rev. 171 (1968) 764–777.
[56] 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. 10 (2019) 592.
[57] D.J. Werder, C.H. Chen, R.J. Cava, B. Batlogg, Phys. Rev. B 37 (1988) 2317–2319.
[58] G.A. Waychunas, W.A. Dollase, C.R. Ross, Am. Min. 79 (1994) 274–288.
[59] R. De Ridder, G. Van Tendeloo, S. Amelinckx, Acta Crystallogr. Sect. A 32 (1976) 216–224.
[60] R.L. McGreevy, L. Pusztai, Mol. Simul. 1 (1988) 359–367.
[61] L. Barroso-Luque, P. Zhong, J.H. Yang, F. Xie, T. Chen, B. Ouyang, G. Ceder, Phys. Rev. B 106 (2022), 144202.
[62] F. Xie, P. Zhong, L. Barroso-Luque, B. Ouyang, G. Ceder, Comput. Mater. Sci 218 (2023), 112000.
[63] J. Burns, B. Ouyang, J. Cheng, M.K. Horton, M. Siron, O. Andriuc, R. Yang, G. Ceder, K.A. Persson, Chem. Mater. 34 (2022) 7210–7219.
[64] J. Cheng, B. Ouyang, K.A. Persson, ACS Energy Lett. (2023) 2401–2407.
[65] P. Zhong, Z. Cai, Y. Zhang, R. Giovine, B. Ouyang, G. Zeng, Y. Chen, R. Clément, Z. Lun, G. Ceder, Chem. Mater. 32 (2020) 10728–10736.
[66] Z. Lun, B. Ouyang, D.A. Kitchaev, R.J. Clément, J.K. Papp, M. Balasubramanian, Y. Tian, T. Lei, T. Shi, B.D. McCloskey, J. Lee, G. Ceder, Adv. Energy Mater. 9 (2019) 1802959.
[67] B. Ouyang, T. Chakraborty, N. Kim, N.H. Perry, T. Mueller, N.R. Aluru, E. Ertekin, Chem. Mater. 31 (2019) 3144–3153.
[68] J. Huang, B. Ouyang, Y. Zhang, L. Yin, D.-H. Kwon, Z. Cai, Z. Lun, G. Zeng, M. Balasubramanian, G. Ceder, Nat. Mater. 22 (2023) 353–361.
[69] T.M. Brenner, D.A. Egger, L. Kronik, G. Hodes, D. Cahen, Nat. Rev. Mater. 1 (2016) 15007.
[70] H.C. Zhou, J.R. Long, O.M. Yaghi, Chem. Rev. 112 (2012) 673–674.
[71] M. Jansen, Angew. Chem. Int. Ed. Eng. 30 (1991) 1547–1558.
[72] Z. Zhang, P.-N. Roy, H. Li, M. Avdeev, L.F. Nazar, J. Am. Chem. Soc. 141 (2019) 19360–19372.
[73] J.G. Smith, D.J. Siegel, Nat. Commun. 11 (2020) 1483.
[74] Z. Zhang, H. Li, K. Kaup, L. Zhou, P.-N. Roy, L.F. Nazar, Matter 2 (2020) 1667–1684.
[75] Z. Zhang, L.F. Nazar, Nat. Rev. Mater. 7 (2022) 389–405.
[76] H. Fang, P. Jena, Proc. Natl. Acad. Sci. 114 (2017) 11046–11051.
[77] H. Fang, P. Jena, Nat. Commun. 13 (2022) 2078.
[78] Y. Sun, B. Ouyang, Y. Wang, Y. Zhang, S. Sun, Z. Cai, V. Lacivita, Y. Guo, G. Ceder, Matter 5 (2022) 4379–4395.
[79] B. Lee, K. Jun, B. Ouyang, G. Ceder, Chem. Mater. 35 (2023) 891–899.
[80] 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. 1 (2013), 011002.
[81] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, JOM 65 (2013) 1501–1509.
[82] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, Comput. Mater. Sci 58 (2012) 218–226.
[83] C. Draxl, M. Scheffler, J. Phys.: Mater. 2 (2019), 036001.
[84] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky, Comput. Mater. Sci 111 (2016) 218–230.
[85] C.J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain, G. Ceder, NPJ Comput. Mater. 6 (2020) 97.
[86] O. Kononova, H. Huo, T. He, Z. Rong, T. Botari, W. Sun, V. Tshitoyan, G. Ceder, Sci. Data 6 (2019) 203.
[87] T.a.G. Xie, C. Jeffrey, Phys. Rev. Lett. 120 (2018) 145301.
[88] C. Chen, S.P. Ong, Nat. Comput. Sci. 2 (2022) 718–728.
[89] O. Kononova, T. He, H. Huo, A. Trewartha, E.A. Olivetti, G. Ceder, iScience 24 (2021) 102155.
[90] H. Huo, Z. Rong, O. Kononova, W. Sun, T. Botari, T. He, V. Tshitoyan, G. Ceder, NPJ Comput. Mater. 5 (2019) 62.
[91] K. Cruse, A. Trewartha, S. Lee, Z. Wang, H. Huo, T. He, O. Kononova, A. Jain, G. Ceder, Sci. Data 9 (2022) 234.
[92] Z. Wang, K. Cruse, Y. Fei, A. Chia, Y. Zeng, H. Huo, T. He, B. Deng, O. Kononova, G. Ceder, Digital Discovery 1 (2022) 313–324.
[93] Y. Kang, H. Park, B. Smit, J. Kim, Nat. Mach. Intell. 5 (2023) 309–318.
[94] C. Chen, Y. Zuo, W. Ye, X. Li, S.P. Ong, Nat. Comput. Sci. 1 (2021) 46–53.
[95] K. Hippalgaonkar, Q. Li, X. Wang, J.W. Fisher, J. Kirkpatrick, T. Buonassisi, Nat. Rev. Mater. 8 (2023) 241–260.
[96] N.J. Szymanski, Y. Zeng, H. Huo, C.J. Bartel, H. Kim, G. Ceder, Mater. Horiz. 8 (2021) 2169–2198.