Title

Automating the Synthesis and Characterization of Inorganic Materials

Author

Szymanski, Nathan

Publication Date

2024

Peer reviewed|Thesis/dissertation

Automating the Synthesis and Characterization of Inorganic Materials

By

Nathan Szymanski

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Materials Science and Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Gerbrand Ceder, Chair

Professor Bryan D. McCloskey

Professor Mark Asta

Spring 2024

Abstract

Automating the Synthesis and Characterization of Inorganic Materials

by

Nathan Szymanski

Doctor of Philosophy in Engineering – Materials Science and Engineering

University of California, Berkeley

Professor Gerbrand Ceder, Chair

The development of new inorganic materials largely depends on manual experiments that are costly and time intensive. While automation has greatly advanced the computational discovery of promising new materials, the rate at which they are experimentally synthesized has fallen behind. Bridging this gap requires an improved approach to materials synthesis and characterization, whereby automation is used to streamline the experimental realization of predicted compounds. In this dissertation, I will summarize my contributions to this area. These include automating the collection and analysis of X-ray diffraction patterns, developing theory-driven decision-making algorithms to guide experimental solid-state synthesis trials, and implementing these methods in a fully autonomous, robotic platform known as the A-Lab.

X-ray diffraction (XRD) is a cornerstone of materials research that is widely used to identify and characterize the structures of distinct crystalline phases. Traditional interpretation of XRD patterns requires manual analysis, which becomes challenging when dealing with multi-phase samples that are often complicated by experimental artifacts such as lattice strain and texture. In Chapter 2, I will describe the development and validation of a machine learning (ML) framework that can automate the identification of crystalline materials from XRD patterns. This framework leverages an ensemble of convolutional neural networks, uniquely trained with physics-informed data augmentation to ensure they are robust against common experimental artifacts. A distribution of predicted phases is generated for each pattern given to these trained models, from which a measure of prediction confidence is evaluated. This method outperforms traditional peak search-match algorithms on a variety of experimental samples without requiring manual intervention, making autonomous phase identification possible.

Because ML models are fast once trained, they can be integrated with experimental measurements to perform analysis in real time. This provides the opportunity to use any information gained from preliminary analysis to control the subsequent measurements, improving the efficiency of data collection. Such an approach can benefit XRD measurements, which typically require 20-30 min of scan time per sample to obtain results that have sufficient quality for *post hoc* analysis. As outlined in Chapter 3, a much shorter scan time of 5-10 min per sample can be achieved by using in-line ML analysis to steer the diffractometer toward parts of the XRD pattern that matter most for phase identification. This approach is shown to provide more precise detection of impurities and short-lived reaction intermediates that are critical to the study of solid-state synthesis.

In early attempts to synthesize a new compound, XRD often reveals the formation of unwanted byproducts instead of the desired target. Avoiding these byproducts and achieving the target requires careful redesign of the experimental procedure. In solid-state synthesis, the most common approach used to make bulk inorganic materials, redesigning the experiments generally involves choosing alternative precursors or reaction conditions. While conditions like temperature and partial pressures are numerical and can therefore be optimizing using well established methods like Bayesian optimization, precursor selection requires a different approach. In Chapter 4, I will describe an algorithm we developed to optimize the selection of precursors used in solid-state synthesis by actively learning from experimental outcomes. It does so by identifying unfavorable reactions that lead to unwanted byproducts, and then choosing precursors that it expects to avoid these reactions and instead favor the target's formation. The effectiveness of this approach is showcased on three separate targets, for which optimal synthesis recipes are identified while requiring few experimental iterations.

The automation of data analysis and decision making, combined with robotics that can perform solid-state synthesis experiments, have made autonomous materials development possible. The integration of these tools into a platform known as the A-Lab is discussed in Chapter 5. Given a set of targeted materials screened using ab-initio computations, this lab can devise initial synthesis recipes based on historical data mined from the literature. It tests these recipes using robotics for automated powder handling and high-temperature annealing, followed by characterization with XRD. The resulting patterns are analyzed by ML models, which then feed into automated decision making to improve upon the initial recipes and achieve higher target yield. We demonstrate the capabilities of the A-Lab by using it to synthesize 41 materials in just 17 days of closed-loop experimentation.

The work reported herein demonstrates the feasibility of autonomous materials development while also highlighting areas that require further improvement. Several promising directions for future work are highlighted in Chapter 6. These include the development and integration of automated characterization to new techniques beyond XRD, the extension of robotic platforms to deal with air-sensitive samples and to measure device performance, and the generalization of decision-making algorithms to deal with experimental issues like melting and volatility.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Professor Gerbrand Ceder, for providing unparalleled leadership, expertise, and kindness. His clear and rigorous commitment to science, coupled with his genuine concern for the well-being of his students, has been a constant source of inspiration throughout my graduate studies. Professor Ceder's steadfast support in the highs and lows of my PhD has been invaluable, and for that, I am forever grateful.

I am also grateful to all the members of my qualifying exam and dissertation committees – Professor Mark Asta, Professor Bryan McCloskey, Professor Andrew Minor, and Professor Mary Scott – who have each spent considerable time and effort providing feedback that has benefited this dissertation.

It has been a great privilege to work with so many brilliant minds in the CEDER group, where the friendly and collaborative atmosphere has led to countless fond memories. I owe special thanks to Professors Chris Bartel and Yan Zeng, whose mentorship was instrumental in facilitating my academic research and personal growth. I am also thankful to all members of the A-Lab team, especially Bernardus Rendy and Yuxing Fei, for their exceptionally hard work that has made such an exciting project possible. To Lauren Walters, I am indebted for the guidance and advice she has provided from my undergraduate studies to the present. I would also like to thank Pragnay Nevatia for being such an excellent mentee and collaborator, as well as Shilong Wang for his dedicated work on our shared projects.

Last, I would like to thank my family for their unwavering support over the past five years. The love and encouragement from my parents have been the bedrock of my perseverance and success. I also appreciate the time spent with many of my friends at Berkeley – including Rüstem Özgür, Kevin Cruse, Ann Rutt, and Sanghoon Lee – who often provided me with a much-needed respite from work. I am especially thankful to Lily Rahnama for her devoted companionship, which has been a source of strength and joy throughout the past few years, consistently lifting my spirits whenever needed.

CHAPTER 1 Introduction

The discussion presented in this chapter is based, often verbatim, on the following publication: ***N. J. Szymanski****, Y. Zeng, C. J. Bartel, H. Huo, H. Kim, and G. Ceder, "Toward autonomous design and synthesis of novel inorganic materials." Materials Horizons 2021, 8, 2169-2198.*

1.1 Background and motivation

Historically, great innovations in technologies have been driven by the discovery of novel materials. Current materials development largely relies on three key steps: (i) identification of a new composition and structure of interest, (ii) targeted and scalable synthesis of that compound, and (iii) post-processing of the product to carefully optimize its properties[1]. To accelerate this procedure, it is necessary to not only improve the efficacy of each step, but also to integrate all three into a closed loop so that they can occur in rapid succession and benefit from optimal feedback between them. While the initial identification step has been assisted by large-scale *ab initio* simulations[2,3], the latter two generally remain difficult and time-consuming owing to the iterative trial-and-error experimental approach required for both synthesis and property optimization. A breakthrough to overcome these challenges may be found in autonomous experimentation enabled by self-driving laboratories, which aim to aid the human researcher with robotic platforms guided by artificial intelligence (AI).

The automation of experiments has long been a topic of interest, with early examples of widespread utilization demonstrated in the pharmaceutical industry[4]. There, high-throughput chemistry platforms have been developed to accelerate drug discovery using combinatorial sampling of possible molecules and synthesis conditions, which can be performed in an automated and highly parallelized manner to save considerable time and costs[5–7]. More recently, the advent of AI has created a symbiosis between hardware and software, with active learning techniques guiding the exploration of design spaces and leading to increased efficiencies relative to combinatorial techniques[8–10]. This has opened the door to more sophisticated applications ranging from systematic inspection of retrosynthetic routes in small molecule manufacturing[11] to performance optimization in organic photovoltaics[12]. Furthermore, by automating the role of the experimenter as opposed to individual instruments, modern systems are flexible and can rapidly incorporate improvements in the underlying technology[13].

In contrast to organic chemistry, the development of autonomous experimentation for inorganic materials remains in its early stages. Given the challenges associated with handling solid powders, the limited availability of methods that can reliably characterize bulk samples, and the lack of a rigorous theoretical framework describing the factors influencing synthesizability, most of the existing work has demonstrated only partial automation of the experimental process. Within the thin film community, for example, HT automation of synthesis and characterization is routinely carried out to probe the effects of composition and processing conditions on the properties of resulting samples[14–16]. Similar methods have also been used to study bulk powders but are generally more limited with respect to the scope of compounds that can be dealt with[17–19]. Existing workflows are restricted to materials with readily available synthesis recipes, which precludes the discovery of novel systems with new and interesting properties. More recently, AI has been

incorporated into the automation pipeline to achieve closed-loop optimization of synthetic routes for nanoparticles formed in continuous flow reactors[20] and nanotubes grown via chemical vapor deposition (CVD)[21,22]. While these platforms can be used to maximize the yield and purity of a target phase, they rely on a reasonable initial guess for the choice of precursors and synthesis conditions so that a measurable amount of the product is consistently obtained and used to guide the optimization. In novel compound synthesis and discovery, however, there is typically insufficient information available regarding successful reaction pathways, and consequently, most synthesis trials fail to produce any amount of the target phase. A new approach to optimization is therefore needed to fully automate the synthesis process.

Three major aspects that must be automated to reach this goal of autonomy in materials synthesis. First, experimental procedures should be carried out by modular, robotic platforms with the capability of synthesizing and characterizing the materials of interest. Second, the data obtained from characterization should be interpreted by the machine and converted into simple, physically meaningful quantities providing insight into the experimental outcome. Last, but certainly not least, this information should then be passed to a decision-making algorithm that actively learns from previously tabulated data and/or scientific principles to suggest new experimental parameters for subsequent tests. Successful design and integration of all three aspects is essential to complete the closed-loop workflow illustrated in **Figure 1-1**.

Before describing our own contributions to the automation of inorganic materials synthesis and characterization, I will outline in the next three sections variety of relevant techniques that are currently available. These serve as the foundation from which we have developed methods for the robotic execution of solid-state synthesis experiments, machined-learning-enabled interpretation of XRD patterns, and physics-informed decision-making algorithms to optimize experimental synthesis procedures.

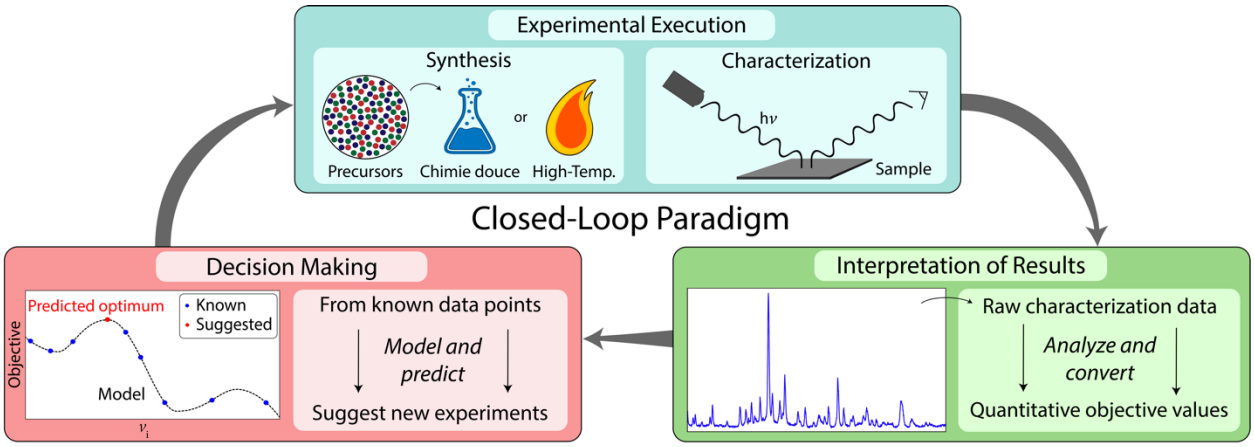

Figure 1-1. A schematic of closed-loop experimentation for inorganic materials synthesis. The three components include automated execution of procedures for synthesis and characterization, interpretation of the resulting data, and decision making to select new experiments based on this data. The integration of these components is necessary to achieve complete autonomy.

1.2 Previous efforts to automate solid-state synthesis

Solid-state synthesis, carried out by mixing powder precursors and firing at high temperatures, is a widely used and scalable approach to produce inorganic materials. Automating this process for high-throughput or closed-loop experimentation, however, remains challenging due to the increased difficulty associated with handling solid powders as opposed to liquids or thin films. Working at high temperatures for long periods of time also poses potential problems caused by the melting of samples and the degradation of containers. Several recent efforts have made steps to automate some key aspects of solid-state synthesis for several classes of materials including PbTe-based thermoelectrics[23], yttrium-doped zirconia[24], and Zr-Ti-C-B ceramics[25]. These existing methods increase the rate at which solid-state syntheses are carried out by decomposing the entire procedure into modular components, each of which is either automated via robotic systems or designed to be conducted in a highly parallelized manner, thereby reducing the time spent by the human researcher per synthesized sample.

Automated weighing and dispensing of powder precursors have been demonstrated with several commercial systems[26,27]. These instruments use gravity to pass samples through a hollow glass or plastic tip and into a container, which is placed on a balance and continuously weighed to control the rate of dispensing and produce the targeted precursor amount. When too much powder is dispensed, small amounts of the sample can be removed using a glass plunger, allowing the automated system to reach a precision on the order of micrograms. Once the precursors have been dispensed, mixing is typically carried out using a ball mill, which can be designed to accommodate many samples at once such that parallelization is possible[25]. If mechanochemical synthesis is desired, high-energy ball milling is often used to encourage the reaction. If, instead, the goal of ball milling is to obtain a well-mixed sample while avoiding any reactions, then low-energy milling can instead be used. From the mixed powders, compacting and densification can be parallelized by stacking samples on top of one another, separated by an inert material, and loading them altogether into a press. Firing of samples is readily parallelized, limited only by the size of the reaction vessel. However, unless separate furnaces are employed, all materials must be synthesized under the same conditions, which prohibits an efficient exploration of all synthesis parameters simultaneously. Ensuing characterization (*e.g.*, by XRD) is usually conducted serially, though their operation periods are often short in comparison to the time required for synthesis and are therefore unlikely to represent the time-limiting step.

Using high-throughput methods, solid-state syntheses can be performed at a rate of more than 200 reactions per day[17]. We stress, however, that previous methods automate individual components of the synthesis process while still requiring a substantial amount of manual intervention between each step. As a result, human efforts constitute a large fraction of the total time allocated for the synthesis and characterization and solid powders. This shortcoming is illustrated by the analysis performed in recent work[23], which shows that of the total 328 minutes necessary to complete a full experimental iteration per sample, 105 minutes are consumed by human efforts. Much of this time is spent performing preparative tasks such as sample loading, cleaning, and extraction. These processes are difficult to automate for solid powders given that their physical properties can vary substantially between different samples, and powders can sometimes adhere to container walls. After synthesis, further manual intervention is required to transfer samples and prepare them for characterization and subsequent analysis. The automation of these aspects and their integration to form a closed-loop workflow for inorganic materials synthesis is addressed in Chapter 5.

1.3 Interpreting synthesis outcomes with X-ray diffraction

Following an experimental synthesis procedure, phase identification is needed to determine what products were made. For crystalline inorganic materials, this entails the application of XRD and comparing the sample's spectrum (also referred to as a pattern) with reference data from sources such as the International Centre for Diffraction Data (ICDD)[28]. However, this comparison is very often complicated by variations that occur between measured and reference patterns due to defects, strain, off-stoichiometry, texture, and poor crystallinity. As a result, interpreting XRD patterns can be an arduous process that requires domain expertise. Even with state-of-the-art computational tools, reliably automating the identification of multi-phase XRD patterns remains a longstanding challenge. The most popular techniques used to complete this task are summarized in **Figure 1-2** and discussed below.

Historically, the analysis of diffraction data has been conducted by decomposing continuous spectra into discrete lists of peak positions (*d*) and intensities (*I*) which are compared with reference data[29]. Peak search-match algorithms rely on a Figure of Merit (FoM) to quantify the degree of similarity between pairs of *d*-*I* lists. A widely used metric is the de Wolff FoM, which is inversely related to the average discrepancy between observed and calculated *d*-spacings[30]. By calculating the FoM for all suspected reference phases, the compound with the highest value may be chosen for a given XRD pattern. However, the reliability of this method hinges on the ability to extract diffraction peaks from the measured pattern, which becomes difficult in the presence of peak overlap, low peak intensity, or strong background signal[31]. These problems are exacerbated when a pattern contains many peaks (*e.g.*, in low-symmetry structures or multi-phase mixtures), and therefore the peak search-match approach generally produces limited accuracy. In a study conducted by Le Meins *et al.*[32], XRD patterns obtained from ten distinct compounds were provided to the broader research community with the task of performing phase identification using peak search-match algorithms. Based on results collected from 25 participants, only 80% of phases were correctly identified on average, even with manual guidance by an expert, thus suggesting the need for improved methods if automation is to be attained.

An alternative to the discrete peak search-match approach is full-profile matching, where *entire spectra* are compared with those of reference phases using a measure of correlation such as cosine similarity, Pearson or Spearman coefficients, or dynamic time warping (DTW)[33–35]. By removing the need to explicitly deconvolute individual peaks, analyzing the full profile provides a more robust treatment of complex and low-symmetry XRD patterns. Furthermore, this method can be combined with non-negative matrix factorization to identify the combination of compounds that best matches a measured pattern, enabling classification of multi-phase mixtures[36]. However, the reliability of full-profile matching remains limited when experimental artifacts cause large changes in peak positions, widths, and intensities. In a study by Iwasaki *et al.*, an accuracy of 70% was achieved with DTW for the classification of multi-phase mixtures comprised of alloys spanning the Fe-Co-Ni chemical space[37]. Misclassifications were largely attributed to variations in XRD patterns induced by off-stoichiometry of the samples. To improve upon existing methods based on full-profile matching, it is necessary to design an approach that can account for the possibility of experimental artifacts.

Deep learning has more recently been used to automate the interpretation of XRD patterns. In the initial study by Park *et al.*, a convolutional neural network (CNN) was trained to categorize the crystal symmetry of simulated patterns from 150,000 phases reported in the Inorganic Crystal Structure Database (ICSD)[38]. With 20% of these patterns reserved for testing, accuracies of 81% and 95% were achieved for the classification of space groups and Bravais lattices respectively.

Nevertheless, characterization of experimental XRD data is complicated by the prevalence of artifacts that cause deviations from simulated reference patterns. In later work by Vecsei *et al.*, a neural network was trained on simulated XRD patterns to classify symmetry as described previously[39]. The model was then applied to experimentally measured patterns extracted from the RRUFF database, producing a lower accuracy of 54% for space group classification. To resolve these shortcomings, simulated patterns in the training set can be augmented to include perturbations associated with experimental artifacts. For example, Oviedo *et al.* demonstrated that by stochastically varying peak positions and intensities in simulated data using for training, the resulting CNN correctly classified the space group for 80% of patterns measured from metal halide thin films[40].

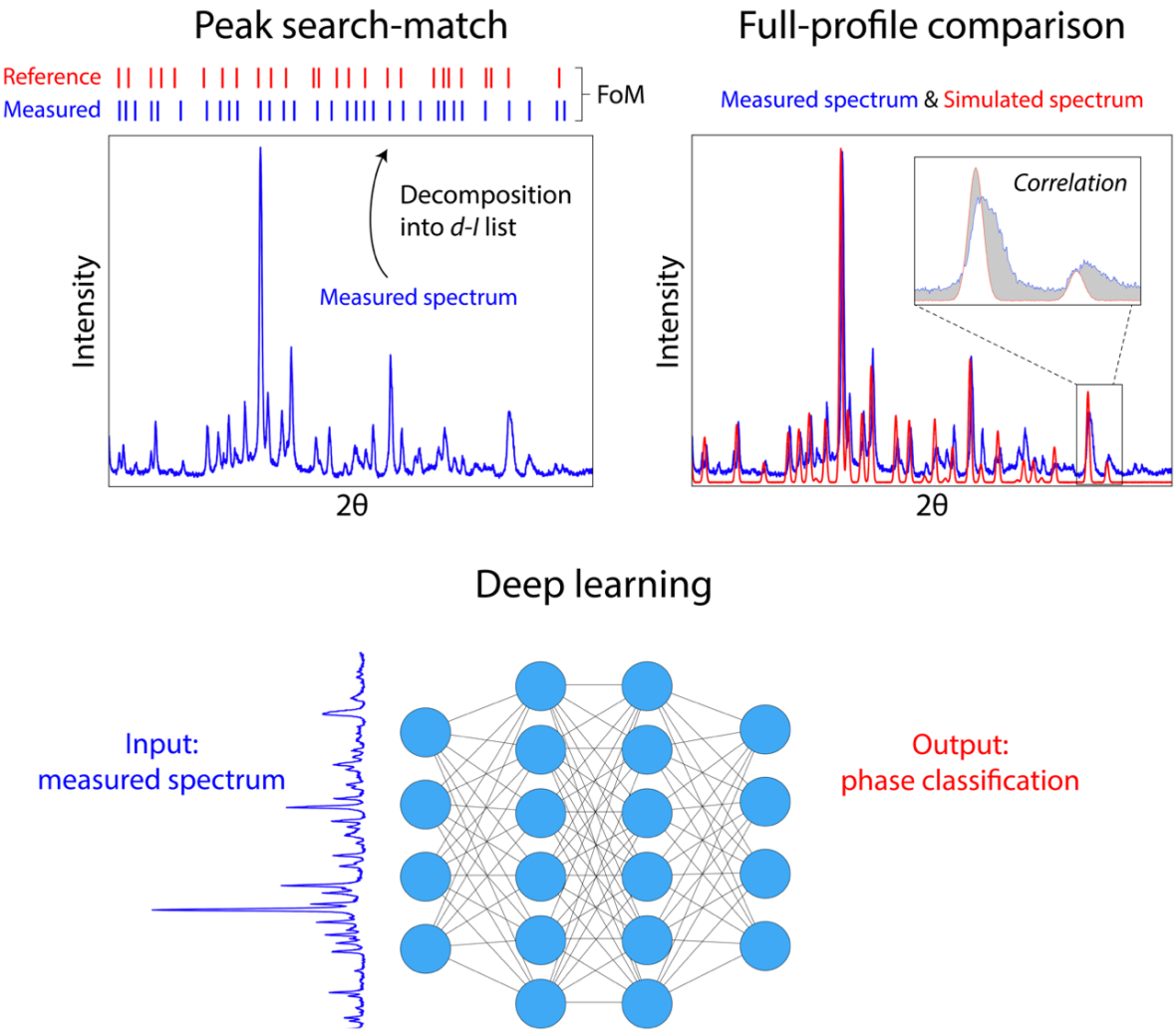

Figure 1-2. Available techniques for automating the interpretation of XRD patterns. (Top left) Peak search-match algorithms rely on the identification of peaks and comparison with reference data using a Figure of Merit. (Top left) Full-profile methods compare entire spectra measured experimentally with reference data, typically simulated, using a correlation metric. (Bottom) Deep learning employs neural networks trained on reference spectra to classify measured patterns into constituent phases.

In addition to symmetry classification, similar techniques based on deep learning and data augmentation have also been used to perform phase identification from experimentally obtained XRD patterns. For example, Maffettone *et al.* trained an ensemble CNN using simulated patterns augmented with changes to peak widths, intensities, and background signals[41]. Their model was tested on patterns measured from samples in the Ni-Co-Al space, with 76% correctly identified. To handle multi-phase mixtures, Lee *et al.* trained a CNN using multi-phase patterns simulated from linear combinations of single-phase patterns for 170 compounds in the Sr-Li-Al-O space[42]. Their model achieved a high accuracy of 98% when classifying experimentally measured patterns obtained from mixtures of high-purity powders including $SrAl_2O_4$, SrO, Li_2O, and Al_2O_3. However, because the training procedure requires many linear combinations of phases with varied weight fractions to be sampled (1,785,405 in total), it restricts the inclusion of experimental artifacts owing to combinatorial explosion. Therefore, the model may fail when applied to characterize arbitrary samples obtained from a synthesis trial, which often contain substantial perturbations in their XRD patterns. With these limitations in mind, we introduce our own method to automate the classification of multi-phase XRD patterns in Chapter 2 of this dissertation.

1.4 Automated decision making to overcome failed syntheses

Upon conducting a batch of experiments and interpreting the resulting data (*e.g.*, by performing phase identification), the next step in the closed-loop automation process is to use this information to make decisions regarding the subsequent experiments to be performed. These decisions are usually made with the goal of optimizing some quantity; for example, maximizing the yield of a product by modifying its synthetic procedure[43] or tuning the properties of a material with respect to its structure, composition, or processing conditions[44]. Alternatively, decisions can be made to formulate experimental tests that reveal information regarding a specific process[45]. In synthesis, for example, this may entail exploring various combinations of reactants and conditions followed by observation of their products to construct a network of possible reaction pathways in a system of interest. Regardless of the underlying motivations, the desired outcome of decision making remains the same: reaching a pre-defined optimum while minimizing time and costs. To this end, a variety of active learning techniques exist to iteratively learn from and query data in the design space. Before reviewing available active learning algorithms, it is first important to understand why active learning is necessary by considering the alternative methods listed in **Figure 1-3** and highlighting their shortcomings.

A simple and widely used optimization strategy is to perform a brute-force search of the design space, thus avoiding decision making altogether. Such is the concept underlying high-throughpout workflows, where a grid of data points is generated from combinatorial sampling of experimental parameters[46]. From this dataset, analysis may then be conducted *ex post facto* to identify relationships among variables and estimate any optima in the objective of interest. As the reliability of these conclusions depends on how well the design space has been sampled, a large number of experiments are typically necessary to obtain satisfactory results. Consequently, successful applications of high-throughput platforms have been limited to problems for which (a) the appropriate experiments are inexpensive, quick, and easily parallelized, or (b) the design space of interest is relatively narrow. A sufficiently dense sampling of compositions on thin films spanning ternary spaces, for example, can typically be achieved using several hundred samples[47]. In contrast, generating a grid of equal density for quaternary systems requires several thousand samples. Additionally, process variables may add extra dimensions to the design space. As the number of necessary experiments scales exponentially with the dimension of the design space,

combinatorial techniques quickly become intractable when many variables are introduced. These problems are sometimes simplified by partitioning the design space and focusing on a much smaller subset of interest[48]; however, this solution is not generalizable because the most interesting region of the design space is generally unknown. Therefore, to efficiently explore the entire design space, active learning is required.

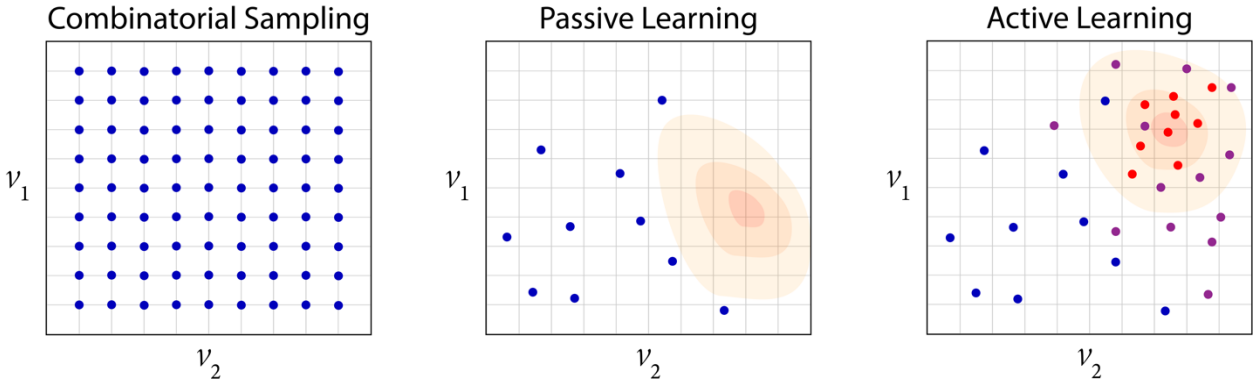

Figure 1-3. An illustration of the three general approaches to optimization. (Left) Combinatorial approaches sample many possible combinations of design variables (v_i), sometimes chosen uniformly across the design space. (Center) Passive learning employs existing data points (blue dots) to form a model of the objective and make predictions regarding the location of its optimum (shaded region). (Right) Active learning builds upon this approach by suggesting new points at which to evaluate the objective (purple dots), from which the information is used to update the initial predictions and once again suggest new points to be queried (red dots), forming an iterative loop which is traversed until convergence to the true optimum is reached.

Contrary to high-throughput experimentation, existing data can be used to learn trends and predict optima in the objective function without performing any new experiments. As the learner simply observes the environment without interacting with it to query new information, this technique is sometimes called *passive* learning to distinguish it from its *active* counterpart. Enabled by the development of ML models and a growing amount of available data, passive learning has found widespread use throughout materials science. For a detailed overview of common ML algorithms and their application in materials science, we refer the reader to several recent reviews on the topic[49,50]. Here, we narrow our discussion to focus on two key limitations of passive learning as applied to optimization. First, the accuracy of the model is heavily reliant on both the volume and diversity of training data. In many situations, the design space of interest is sparsely populated. For example, applying ML to inorganic synthesis remains difficult because there are often few procedures reported to make a given compound, and that information must be extracted from the literature as relevant synthesis databases are limited[51]. Moreover, even in cases where more data is available, it tends to be biased toward specific regions of the design space. This bias commonly originates from a tendency for researchers to only publish positive results while leaving negative results unreported[52], or because many studies pursue minor modifications of an already successful material/procedure. Such data bias will limit the diversity of the training set and negatively affect the performance or applicability of the corresponding model. In addition to the limitations imposed by the sparsity of training data, passive learning models are inherently inept at predicting outliers, instead relying on the recognition of general trends in the data. While this capability is sufficient for many studies, it becomes problematic for optimization problems where the global extrema are

of interest. To overcome these limitations, it is necessary to acquire new data so that the model can continuously learn and improve its accuracy, thus ensuring correct identification of optima.

Active learning techniques are gaining traction throughout materials science, with multiple applications recently demonstrated in automated experimental workflows[53]. In situations where little is known about the system at hand, the objective function being optimized by active learning can be treated as a "black box," *i.e.*, an opaque function that must be queried at individual points through experimentation or simulation. Performing black-box optimization, a topic that has been studied extensively and applied throughout many areas of science and technology, requires the consideration of two key constraints. First, as no analytical form of the objective function is available, optimization must be carried out without the use of exact derivatives. Second, the objective landscape may generally be non-convex, requiring global instead of local optimization. These properties exclude the application of explicit gradient-based and pure local search methods respectively. Moreover, extending black-box optimization specifically to experimentation presents an additional challenge: evaluation of the objective is usually expensive and time-consuming, stressing the importance of reaching convergence in a minimal number of steps. The high cost of data acquisition further excludes algorithms that approximate derivatives via finite differences owing to their inefficiency with respect to the number of evaluations required. Instead, a variety of efficient and derivative-free techniques have been developed to perform global optimization on black-box functions.

Many of the earliest optimization techniques used in experimental workflows were based on genetic algorithms[46]. In this approach, an initial batch (or generation) of experiments is conducted to evaluate the objective function(s). From the results, a new batch of experiments is suggested based on three processes: (i) *selection* dictates which samples are chosen to contribute to the next generation of experiments, (ii) *crossover* determines how the properties of selected samples are merged to suggest new experiments, and (iii) *mutation* applies random variations to the properties of suggested experiments. Each process is controlled by a set of hyperparameters (*e.g.*, the rate at which mutation is applied) defined by the user. From the corresponding modifications, new generations of experiments are iteratively produced until convergence of the objective function(s) is reached. Because genetic algorithms impose a bias toward promising regions of the design space by selectively sampling experiments where the objective function is expected to be optimal, they generally provide increased efficiency relative to combinatorial techniques. Genetic algorithms are also well-suited to handle a large number of variables, both qualitative and quantitative, and can perform well in multi-modal design spaces assuming that a sufficiently high mutation rate is used to escape local optima.

Another widely used optimization technique is the SNOBFIT (Stable Noisy Optimization by Branch and FIT) algorithm[54], which combines aspects from local and global search strategies to efficiently optimize an objective function. From a given dataset, SNOBFIT employs a branching algorithm to partition the design space into unique sub-regions, each containing a single known datapoint. Within each region, a local model of the objective function is constructed via least-squares quadratic fitting of the contained datapoint and its nearest neighbors. These models represent the objective function *locally* but do not necessarily describe it *globally* – each quadratic fitting is performed independently using a subset of known datapoints. The resulting models are then used to predict and suggest sampling of new datapoints in regions where the objective function is expected to be optimal. At the same time, sampling is also suggested in sparsely populated regions to ensure the global optimum is not missed. Although SNOBFIT often improves efficiency relative to combinatorial methods and genetic algorithms, it still displays several

shortcomings that limit its applicability to certain optimization problems. First, its performance deteriorates when applied to problems with high-dimensional parameter spaces[55]. Second, it is not capable of directly handling multi-objective optimization procedures without combining those objectives into one composite score[56]. Last, SNOBFIT operates by establishing a set of individual models (*e.g.*, quadratic functions) fit to approximate local regions of the objective function without providing a global model for the entire system. This approach limits interpretability and makes it difficult to draw conclusions regarding general relationships between the variables and objectives.

In cases where evaluating the black-box objective function is particularly costly, and therefore minimizing the total number of iterations required for to find an optimum is critical, Bayesian optimization is often used. This makes it the method of choice for most autonomous experimental workflows[57]. Bayesian optimization is performed by considering a known and differentiable surrogate model rather than on the objective function itself. This surrogate model approximates the objective function using all available data points (*e.g.*, from previously conducted experiments). This approximation is given by a probabilistic distribution of functions known as the *prior*, which is actively updated as new datapoints are sampled to form a *posterior* distribution that more closely resembles the true objective function. Calculating the prior and posterior are essentially regression problems that can be solved using several techniques, making Bayesian optimization versatile with respect to the types of data it can handle. Two models that are most commonly used for regression are Gaussian processes and random forests, which typically work well with continuous and discrete search spaces, respectively[8].

While Bayesian optimization provides an effective approach to maximize costly objective functions, it still can become inefficient in particularly large search spaces. Such cases warrant the use of *physics-informed* methods that simplify the search space or bias the optimization algorithm toward promising regions of it. For example, *data fusion* has been used to map existing information from multiple sources onto an ensemble model, in which all the knowledge is represented using a single composite function. This method is commonly used to obtain an optimal balance between theory and experiment – the former is cheap to evaluate but prone to inaccuracy, whereas the latter is accurate but expensive to carry out. One application of this method is in the optimization of materials' thermodynamic stability by combining density functional theory (DFT) calculations with experimental observations of decomposition[58]. On the other hand, when multiple data sources are not available, prior knowledge (*e.g.*, chemical intuition) can instead be leveraged to formulate rule-based optimization algorithms – sometimes referred to as hypothetico-deductive modeling. King *et al.* pioneered the automation of this approach in their design and application of a robotic scientist named Adam[59], which systematically probed metabolic networks for gene identification. Though, one notable limitation of the hypothetico-deductive approach to optimization is that it generally requires careful redesign to handle each specific problem. In Chapter 4 of this dissertation, I will describe our own development of a rule-based algorithms that also incorporates elements of data fusion to optimize solid-state synthesis procedures targeting inorganic materials.

CHAPTER 2 Automating the interpretation of multi-phase XRD patterns with deep learning

The work presented in this chapter is based, often verbatim, on the following publication:
***N. J. Szymanski**, C. J. Bartel., Y. Zeng, Q. Tu, and G. Ceder, "Probabilistic Deep Learning Approach to Automate the Interpretation of Multi-phase Diffraction Spectra." Chemistry of Materials 2021, 33 (11), 4204-4215.*

Autonomous synthesis and characterization of inorganic materials requires the automatic and accurate analysis of XRD patterns. For this task, we designed a probabilistic deep learning algorithm to identify complex multi-phase mixtures. At the core of this algorithm lies an ensemble of CNNs trained on simulated diffraction patterns, which are systematically augmented with physics-informed perturbations to account for artifacts that can arise during experimental sample preparation and synthesis. Larger perturbations associated with off-stoichiometry are also captured by supplementing the training set with hypothetical solid solutions. Pattern containing mixtures of materials are analyzed with a newly developed branching algorithm that utilizes the probabilistic nature of the neural network to explore suspected mixtures and identify the set of phases that maximize confidence in the prediction. Our model is benchmarked on simulated and experimentally measured diffraction patterns, showing exceptional performance with accuracies exceeding those given by previously reported methods based on profile matching and deep learning. We envision that the algorithm presented here may be integrated in experimental workflows to facilitate the high-throughput and autonomous discovery of inorganic materials.

2.1 Introduction

The development of high-throughput and automated experimentation has ignited rapid growth in the amount of data available for materials science and chemistry[15,60]. Unlocking the physical implications of resulting datasets, however, requires detailed analyses that are traditionally conducted by human experts. In the synthesis of inorganic materials, this often entails the manual interpretation of XRD patterns to identify the phases present in each sample. Past attempts to automate this procedure using peak indexing[29,32] and full profile matching[33,37] algorithms have been limited by modest accuracy, in large part because measured patterns usually deviate from their ideal reference patterns (*e.g.*, due to defects or impurities). Consequently, the analysis of XRD patterns widely remains a manual task, impeding rapid materials discovery and design. To alleviate this bottleneck, deep learning based on the CNN architecture has recently emerged as a potential tool for automating the interpretation of diffraction patterns with improved speed and accuracy[42,61].

Previous work has demonstrated that CNNs can be used to perform symmetry classification[38] and phase identification[41] from XRD patterns of single-phase samples. Given the lack of well-curated diffraction data obtained experimentally, training is most commonly

performed on labeled sets of *simulated* patterns derived from known crystalline materials, *e.g.*, in the ICSD[62]. However, because many factors can cause differences between observed and simulated diffraction peaks, this approach can be problematic for extension to experimentally measured data. Vecsei *et al.* demonstrated that a neural network trained on simulated data produced an accuracy of only 54% for the classification of experimentally measured diffraction patterns extracted from the RRUFF database[39]. To overcome this limitation, simulated patterns can be augmented with perturbations designed to emulate possible artifacts. For example, Oviedo *et al.* trained a CNN using simulated patterns augmented with random changes in their peak positions and intensities, which were chosen to account for texture and epitaxial strain in the thin films being studied[61]. The resulting model correctly classified the space group for 84% of diffraction patterns measured from 115 metal halide samples. Based on past work, we propose that generalization of deep learning to handle complex XRD patterns requires a more complete data augmentation procedure that properly accounts for all the artifacts and complexities that frequently arise during sample preparation and synthesis.

To extend the application of CNNs to mixtures of materials, Lee *et al.* constructed a training set of multi-phase samples that were simulated using linear combinations of single-phase diffraction patterns from 38 phases in the quaternary Sr-Li-Al-O space[42]. Their model performed well in the identification of high-purity samples, with 98% of all phases correctly labeled based on 100 three-phase patterns. However, the combinatorial nature of their technique requires an exceptionally high number of training samples (nearly two million patterns from 38 phases), which restricts the inclusion of experimental artifacts *via* data augmentation. Moreover, because the number of training samples increases exponentially with the number of reference phases, the breadth of the composition space that can be efficiently considered is limited. Proposing an alternative approach, Maffettone *et al.* designed an ensemble model trained on simulated single-phase patterns to yield a probability distribution of suspected phases for a given sample[41]. From this distribution, the authors infer that high probabilities suggest that the corresponding phases are present in the mixture. While this method avoids combinatorial explosion and thus allows many experimental artifacts to be included during training, it sometimes leads to confusion as obtaining comparable probabilities for two phases does not necessarily imply that both are present. Rather, it may simply mean that the algorithm has difficulty distinguishing between the two phases. An improved treatment of multi-phase patterns therefore necessitates an approach that (i) allows artifacts to be incorporated across many phases and (ii) distinguishes between probabilities associated with mixtures of phases as opposed to similarities between single-phase reference patterns.

In this work, we introduce a novel deep learning technique to automate the identification of inorganic materials from XRD patterns of single- and multi-phase samples. In our approach, training patterns are generated with physics-informed data augmentation whereby experimental artifacts (strain, texture, and domain size) are used to perturb diffraction peaks. The training set is built not only from experimentally reported stoichiometric phases, but also from hypothetical solid solutions that account for potential off-stoichiometries. An ensemble CNN is trained to yield a distribution of probabilities associated with suspected phases, which is shown to be a surrogate for prediction confidence. We extend this probabilistic model to the analysis of multi-phase mixtures by developing an intelligent branching algorithm that iterates between phase identification and profile subtraction to maximize the probability over all phases in the predicted mixture. As a representative example to assess the efficacy of our approach, we trained and tested a model on diffraction patterns derived from materials in the broad Li-Mn-Ti-O-F composition space given

their structural diversity and technological relevance. By also systematically testing on a dataset of experimentally measured XRD patterns designed to sample complexities that often arise during synthesis, we show that our model achieves considerably higher accuracy than state-of-the-art profile matching techniques as well as previously developed deep learning-based methods. The improved performance demonstrated here should be generalizable to any alternative chemical space (beyond Li-Mn-Ti-O-F) through application of the same data augmentation and training procedures to any given set of phases from the space of interest.

2.2 Methods

2.2.1 Stoichiometric reference phases

The identification of inorganic materials from their XRD patterns relies on the availability of suitable reference phases that can be compared to samples of interest. In this work, we focus on the Li-Mn-Ti-O-F chemical space (and subspaces) and retrieved all 1,216 corresponding entries from the ICSD. For the identification of stoichiometric materials, we excluded 386 entries with partial occupancies from this set. To remove duplicate structures from the remaining 830 entries, all unique structural frameworks were identified using the pymatgen structure matcher[63]. For each set of duplicates, the entry measured most recently at conditions nearest ambient (20 °C and 1 atm) were retained. Based on these selection criteria, 140 unique stoichiometric materials were tabulated and used as reference phases. The code used to apply these selection criteria and create a set of unique reference phases from ICSD entries in any given composition space is available at https://github.com/njszym/XRD-AutoAnalyzer.

2.2.2 Non-stoichiometric reference phases

Although many solid solutions are available in the ICSD, they generally cover a narrow composition range while leaving others sparse. We therefore designed an algorithm to extend the space of non-stoichiometric reference phases by using empirical rules to construct hypothetical solid solutions between the available stoichiometric materials. To determine which phases may be soluble with one another, all combinations of the 140 stoichiometric references phases in the Li-Mn-Ti-O-F space were enumerated and two criteria were considered for each pair. First, solubility requires that the two phases adopt similar structural frameworks, which was verified using the pymatgen structure matcher. Second, based on the Hume-Rothery rules, the size mismatch between any ions being substituted with one another should be $\leq 15\%$. To estimate the ionic radii of all species comprising each phase, oxidation states were assigned using the composition-based oxidation state prediction tool in pymatgen[63]. In cases where mixed oxidation states are present (*e.g.*, $Mn^{3+/4+}$), we chose to focus on the state(s) that minimizes the difference between the radii of the ions being substituted and therefore increases the likelihood for solubility. As will be shown by our test results, including more reference phases does not lead to a substantial decrease in accuracy; hence, it is preferable to overestimate solubility such that more structures are created as potential references.

Based on the 140 stoichiometric reference phases in the Li-Mn-Ti-O-F space, 43 pairs of phases were found to satisfy both solubility criteria described above. The phases in each pair were treated as end-members, from which interpolation was used to generate a uniform grid of three intermediate solid solution compositions. For example, between spinel $LiMn_2O_4$ and $LiTi_2O_4$, intermediate compositions take the form $LiMn_{2-x}Ti_xO_4$ with $x \in \{0.5, 1.0, 1.5\}$. The lattice parameters of hypothetical solid solutions were linearly interpolated between those of the

corresponding end-members in accordance with Vegard's law. Atomic positions and site occupancies were similarly obtained by interpolating between equivalent sites in the end-members. This procedure gave a total of 129 hypothetical solid solution states from the 43 pairs of soluble phases. Excluding 14 duplicates resulted in 115 distinct solid solutions. The code for generating hypothetical solid solutions for an arbitrary group of stoichiometric reference phases is available at https://github.com/njszym/XRD-AutoAnalyzer.

2.2.3 Data augmentation

From the reference phases in the Li-Mn-Ti-O-F space, we built an augmented dataset of simulated XRD patterns with the goal of accurately representing experimentally measured diffraction data. Physics-informed data augmentation was applied to produce XRD patterns that sample various changes in peak positions, intensities, and widths. Shifts in peak positions (2θ) were derived by creating modified unit cells with up to $\pm 4\%$ strain in each lattice parameter. This was done by applying strain tensors to the lattice parameter matrix $\left(\vec{a},\ \vec{b},\ \vec{c}\ \right)$ that preserve the space group of each structure. Internal cell coordinates were left unchanged so that only peak positions were affected. Peak widths were broadened by simulating domain sizes ranging from 1 nm (broad) to 100 nm (narrow) through the Scherrer equation. Peak intensities were varied to mimic preferred orientation along preferred crystallographic planes (hereafter referred to as texture). This was done by performing scalar products between the peak indices and randomly selected Miller indices (hkl), followed by a normalization that scaled peak intensities by as much as $\pm 50\%$ of their initial values.

The bounds used for each artifact are chosen such that perturbations to the simulated data are large enough to capture possible experimental complexities, but not so large that they produce patterns that are unlikely to ever arise in experiment. Although it is difficult to rigorously define the range of artifacts that may occur, we used our prior experience and physics-based intuition to determine the extent of strain, texture, and domain size described in the previous paragraph. We note that larger variations may arise when substantial off-stoichiometry is present; however, this situation was treated separately by the addition of non-stoichiometric solid solutions as reference phases. In **Figure 2-1a**, we illustrate the effect of each of the three experimental artifacts on the XRD pattern of spinel Mn_3O_4 as an example. Each artifact was applied separately to the simulated pattern by taking 50 random samples from a normal distribution (*e.g.*, between -5% and $+5\%$), resulting in 150 augmented XRD patterns per reference phase (50 samples for each of the three artifacts). Applying this procedure to all 255 references phases, including both experimentally reported stoichiometric materials and hypothetical solid solutions, resulted in 38,250 simulated diffraction patterns. The code to perform data augmentation for an arbitrary group of reference phases is available at https://github.com/njszym/XRD-AutoAnalyzer. Although all patterns used here were derived using Cu K_α radiation, any wavelength can be specified by the user. Therefore, our model can be applied to diffraction data measured using a variety of in-lab diffractometers or synchrotron light sources.

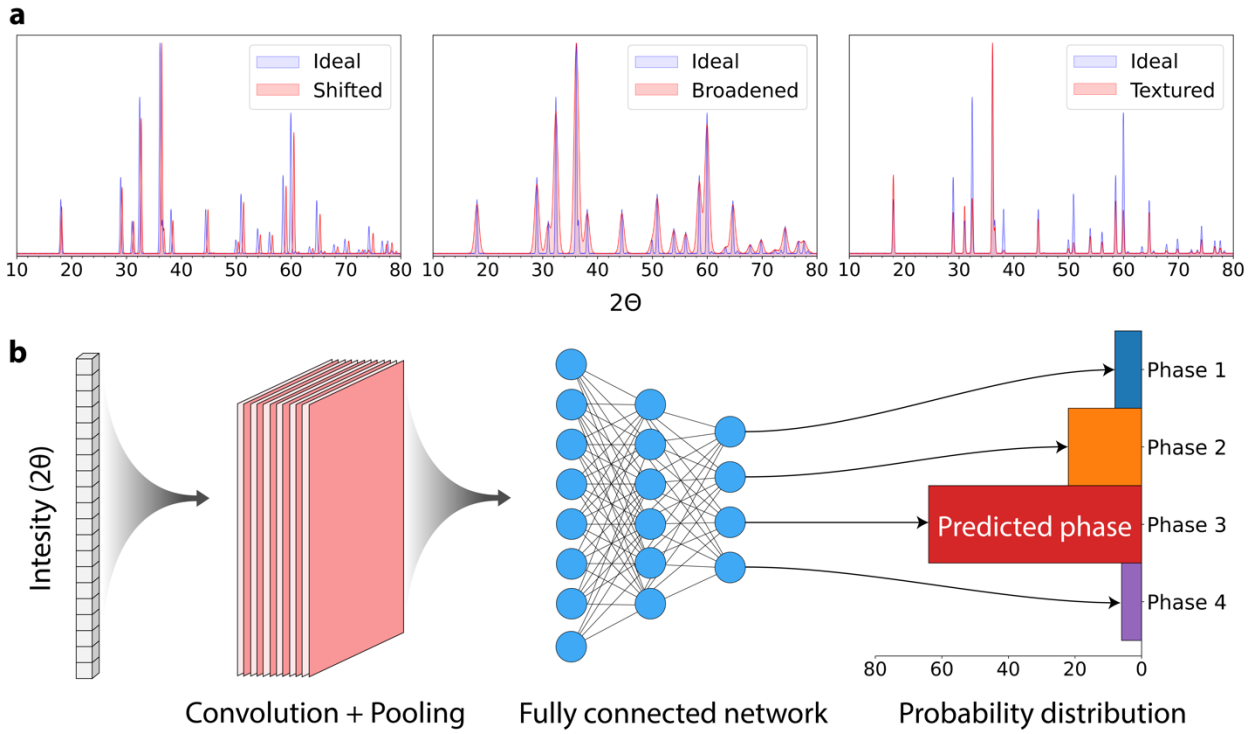

Figure 2-1. An overview of the method used to automate phase identification from XRD patterns. (a) Three examples of the data augmentation procedure designed to sample possible experimental artifacts including peak shift associated with cell strain, peak broadening related to small domain size, and peak intensity variation caused by texture. (b) A schematic of the deep learning pipeline used to map XRD patterns onto a probability distribution of suspected phases.

2.2.4 Convolutional neural network

The workflow used to classify a given XRD pattern is displayed in **Figure 2-1b**. Diffraction patterns are treated as one-dimensional vectors that contain 4,501 values for intensity as a function of 2θ. The range of 2θ is set from 10° to 80°, which is commonly used for scans with Cu K_α radiation ($\lambda = 1.5406$ Å). The intensities (represented as 4,501-valued vectors) serve as input to a CNN that consists of six convolutional layers, six pooling layers, and three fully connected layers. Training was carried out with five-fold cross-validation using 80% of the simulated diffraction patterns, with the remaining 20% reserved for testing (*i.e.*, excluded from training and validation). The code used for training CNNs on new reference phases in arbitrary chemical spaces is also available at https://github.com/njszym/XRD-AutoAnalyzer. To classify patterns outside of the training set, an ensemble approach was used whereby 1,000 individual predictions are made with 60% of connections between the fully connected layers randomly excluded (using dropout) during each iteration. The probability that a given phase represents the pattern is then defined as the fraction of the 1,000 iterations where it is predicted by the CNN. The resulting distribution may be treated as a ranking of suspected phases in the sample, with corresponding probabilities providing measures of confidence.

2.2.5 Intelligent branching algorithm

Given that the CNN was trained only on single-phase XRD patterns, additional methods were developed to automate the identification of materials in multi-phase mixtures. In our workflow, we use an iterative procedure where phase identification is followed by profile fitting and subtraction. Once a phase is identified by the CNN, its diffraction peaks are simulated and fit to the pattern in question using DTW, a well-known technique for correlating features in time series. In contrast to Rietveld refinement, which is typically conducted manually using expert intuition regarding the structure and composition of each phase, DTW is readily automated as it requires no physical input other than a user-specified window in which features can be correlated. For this work, we use a window of $\Delta 2\theta = 1.5°$ since larger peak shifts are typically not expected. After DTW has been applied to fit the simulated pattern along 2θ, its diffraction peaks are scaled to minimize the average difference between the simulated and measured intensities. By using an average difference rather than focusing only on the largest peaks, we aim to avoid scaling errors caused by overlapping peaks between different phases. Following this scaling process, the profile of the identified phase is subtracted to produce a modified pattern that is representative of the mixture minus the phase that has already been identified. In other words, all known peaks are iteratively removed from the pattern. This process is repeated until all significant peaks are attributed to a reference phase, *i.e.*, the cycle is halted once all intensities fall below 5% of the initially measured maximum intensity.

Following the iterative procedure outlined above, one could identify a multi-phase mixture by using the collection of most probable phases given by the model at each step. However, because the pattern is affected by all prior phases that have been identified, such a method over-prioritizes the first iteration of phase identification. In cases where the first phase predicted by the CNN is incorrect, the pattern resulting from profile fitting and subtraction will contain diffraction peaks that do not accurately represent the remaining phases in the sample. All subsequent analyses will therefore be less likely to identify these phases. To improve upon this approach, we developed an intelligent branching algorithm that gives equal importance to each iteration of phase identification. In **Figure 2-2**, we illustrate how the algorithm evaluates several possible sets of phases to classify a diffraction pattern derived from a mixture of Li_2TiO_3, Mn_3O_4, and Li_2O. At each step, the CNN generates a list of suspected phases along with their associated probabilities. As opposed to considering only the most probable phase at each iteration, the branching algorithm investigates all phases with non-trivial probabilities ($\geq 10\%$). By following the pattern associated with the subtraction of each suspected phase, a "tree" is constructed to describe all combinations of phases predicted by the model. Once each route has been fully exhausted, the branch with the highest average probability is chosen as the final set of predicted phases (*e.g.*, the green phases highlighted in **Figure 2-2**). In this way, the algorithm maximizes the likelihood that predictions are representative of *all* phases contained in the actual mixture, as opposed to over-prioritizing the first iteration of phase identification. We found that this is an essential feature to predict multi-phase samples correctly.

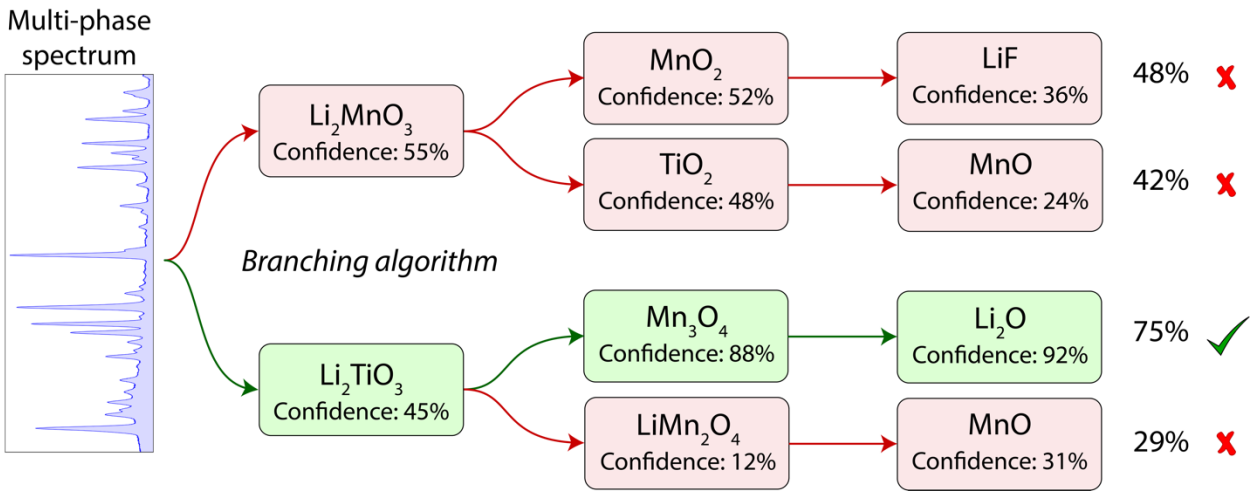

Figure 2-2. A schematic illustrating various pathways enumerated by the branching algorithm for multi-phase identification. This method iteratively performs single-phase predictions followed by profile-stripping, at each step tabulating the probability associated with each phase. This process is repeated until all intensities fall below 5% of the original maximum value. From all branches developed, the one with the highest average probability (highlighted green above) across all levels is chosen as the most likely set of phases present in the mixture.

2.2.6 Experimental measurements

To further validate our model, we built an experimental dataset from a series of measurements designed to sample complexities that often arise during synthesis. Ten materials were chosen to span a range of structures and compositions in the Li-Mn-Ti-O-F space. For a benchmark on pristine single-phase XRD patterns with no intended artifacts, we conducted precise diffraction measurements on each of the ten materials using carefully prepared, high-purity samples. The following modifications were then separately introduced such that each batch of samples contained one anticipated artifact: (i) samples were overlaid with Kapton tape during characterization to produce a diffuse background signal with a magnitude as large as 200% of the highest diffraction peak intensity; (ii) rapid scan rates (30°/minute) were used to generate noisy baseline signals with magnitudes reaching 5% of the maximum diffraction peak intensity; (iii) peak shifts as large as 0.4° were imposed by preparing thick pellets such that specimens were leveled slightly above the sample holder; (iv) broad peaks with full widths at half maxima as large as 1.5° were obtained by ball milling. Several additional materials were also made to sample changes in composition and site occupancy. Six samples of spinel $LiMnTiO_4$ were synthesized at temperatures of 900 °C, 950 °C, and 1000 °C followed by quenching or slow cooling based on previous reports[64]. These samples were intended to contain differences in relative diffraction peak intensities owing to varied distributions of cation site occupancies. Non-stoichiometry was studied using four disordered rocksalt phases, each with a different composition made *via* solid-state synthesis. For the classification of multi-phase XRD patterns, ten two- and three-phase mixtures were prepared from combinations of materials in the Li-Mn-Ti-O-F space that were chosen to include patterns with a substantial amount of peak overlap. The mixtures contained equal weight fractions of all constituent phases. To isolate the effects of multiple phases, these measurements were conducted on samples for which no experimental artifacts were purposefully incorporated.

2.3 Results

2.3.1 Identification of stoichiometric phases

As a first test case, we evaluated the performance of our model on simulated single-phase XRD patterns derived from the 140 stoichiometric reference phases in the Li-Mn-Ti-O-F space. Accordingly, the CNN was trained on 80% of the 21,000 generated patterns (140 materials × 150 augmentations) that were augmented to include physics-informed perturbations to their diffraction peak positions, widths, and intensities. The remaining 4,200 patterns were reserved for testing. To assess the ability of the CNN to handle artifacts not considered during training, the test set was also supplemented with patterns that have diffuse and noisy background signals. A diffuse background was simulated by adding an XRD pattern measured from amorphous silica to the diffraction peaks of the stoichiometric materials. Ten patterns were created for each phase (1,400 in total), with the maximum intensity produced by silica ranging from 100-300% of the maximum peak intensity of the reference phase. Another 1,400 patterns were simulated by adding Gaussian noise with magnitudes ranging from 1-5% of the maximum diffraction peak intensity. Before being passed to the CNN, these 2,800 patterns were pre-processed using the baseline correction and noise filtering algorithms. This procedure is designed to replicate artifacts formed when imperfect corrections are made during pre-processing, which occasionally leads to the disappearance of minor peaks or leaves behind residual intensities related to amorphous impurities. Previous work has dealt with diffuse and noisy background signals by training on patterns with added baseline functions (*e.g.*, polynomials)[41]. However, because these functions are randomly selected rather than derived from possible impurities or defects, they are unlikely to accurately represent experimental measurements. With this in mind, our current approach relies only on physics-informed data augmentation to improve the match between simulated and experimentally measured data.

The performance of our model is compared to a known standard, the JADE software package from MDI[65]. JADE is a widely used program that can automate phase identification with conventional profile matching techniques. During testing, JADE was employed without any manual intervention to ensure a consistent comparison with the CNN, as we are assessing the capability of our approach to perform phase identification as part of an autonomous platform. We emphasize that our model is not designed to replace manual techniques such as Rietveld refinement, but rather to provide more rapid and reliable predictions regarding phase identities. For this task, we applied both the trained CNN and JADE to the test set of simulated diffraction patterns that sample possible experimental artifacts *separately* as discussed in the Methods. In **Figure 2-3a**, we compare the resulting accuracy of each method quantified as the fraction of phases correctly identified. Across the simulated test patterns, the CNN achieves a high accuracy of 94%. In contrast, JADE correctly identifies only 78% of phases when applied to the same set of data. To further verify the effectiveness of the CNN, an additional 1,400 patterns were simulated with *mixed* artifacts such that each pattern contains all aforementioned perturbations to its diffraction peaks (shifting, broadening, and texture) as well as a diffuse and noisy background signal. This incorporates an additional level of complexity not included in the training set, where each pattern contained just one type of perturbation. When applied to the new test set with mixed artifacts, the accuracy of the CNN decreases only 2% (from 94% to 92%), whereas the accuracy of JADE decreases 10% (from 78% to 68%).

The tests show promising results for the CNN, though its performance is not without error. We look to the underlying causes of the occasional misclassifications that occur by dividing the simulated test patterns into four major categories: those augmented *via* the individual application

of peak shifts, peak broadening, peak intensity change, and background effects (including diffuse and noisy baselines). The training set remains unchanged from the previous paragraph. In **Figure 2-3b**, we show the fraction of misclassifications that arise from each perturbation category. Of the 7,000 total test patterns, 418 are misclassified by the CNN. The largest portion (48%) of misclassifications occur for samples containing peak shifts, which we attribute to the overlapping of diffraction peaks between similar phases. This most commonly occurs between isomorphic phases and, as a result, the CNN gives a higher accuracy for the identification of structure (96%) as opposed to composition (92%). We investigated the effects of increasing the bounds on strain that were used during training (beyond ±4%); however, a decrease in accuracy was observed as larger strains were incorporated. For example, training on patterns derived from structures with strain as large as ±6% led to a lower accuracy of 86% when applied to the test set containing samples with as much as ±4% strain. Relative to peak shifts caused by strain, patterns with broad peaks lead to fewer misclassifications, comprising 27% of errors. For this effect, misclassification occurs more frequently in low-symmetry structures as they contain many diffraction peaks that tend to overlap with one another upon broadening. Of the 113 samples that are incorrectly classified by the CNN due to peak broadening, 82 are from phases with monoclinic or triclinic symmetry. The remaining artifacts, including texture and background effects, show a relatively weak influence on the accuracy of the CNN. Because both of these artifacts cause changes in relative peak intensities, the distribution of misclassifications suggest that peak intensities have a more subtle role in single-phase identification.

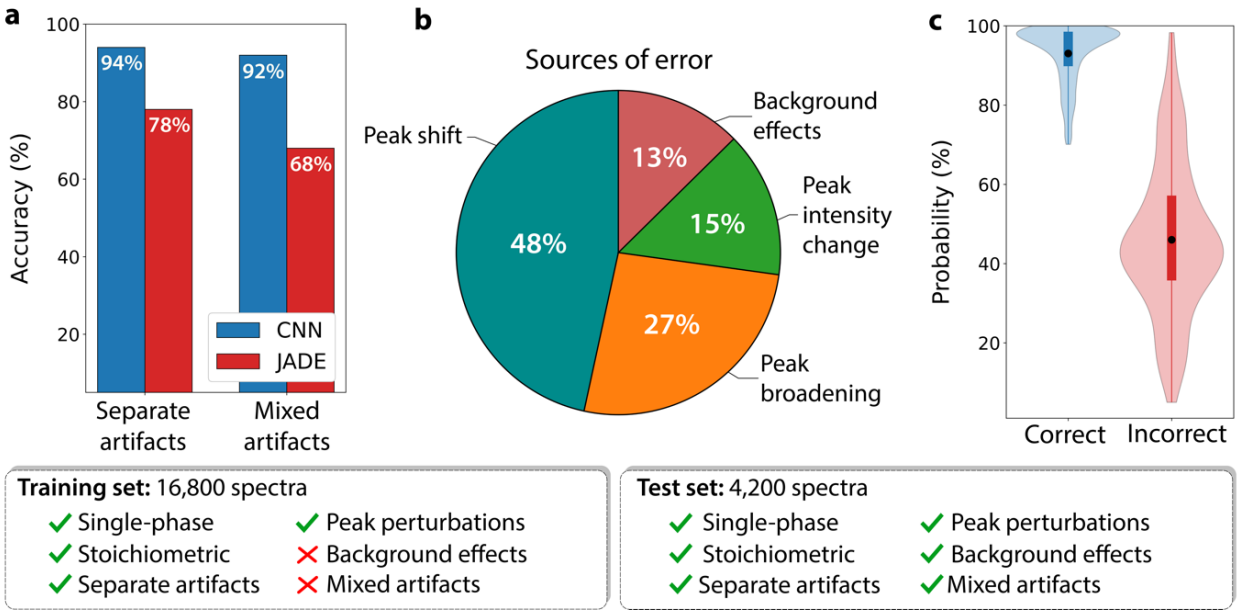

Figure 2-3. Test results from the automated identification of single-phase XRD patterns. (a) Accuracies of the CNN and JADE when applied to simulated patterns with individual or mixed artifacts. (b) Sources of error in the CNN are illustrated by the fraction of misclassifications that occur for patterns with each separate artifact. (c) Distributions of probabilities given by the CNN when correct and incorrect classification are made during testing on data containing mixed artifacts. Violins plots illustrate the density of probabilities, whereas embedded boxes extend from the lower to upper quartiles. Black dots are used to denote the average probability in each case.

To assess the reliability of predictions made by our model, we examined the probability distributions given by the ensemble CNN. In **Figure 2-3c**, we compare the probabilities of correct and incorrect classifications made when the CNN is applied to simulated patterns containing mixed artifacts. All correct classifications are accompanied by a probability greater than 70%, with an average of 93%, whereas incorrect classifications show a wide range of probabilities with a much lower average of 46%. This dichotomy suggests that probabilities are akin to confidence in the prediction and may be used as a reliable metric to gauge the likelihood that a classification is correct. If, for example, predictions are constrained to those with a probability above 70% (which comprise 84% of all patterns in the test set), then the accuracy increases from 92% to 96%. On the other hand, when the probability is lower than 70%, we propose that the model should raise a "red flag," signifying that manual intervention is needed to clarify the identity of the underlying phase. Interestingly, even when an incorrect classification is made regarding the most probable phase, the correct phase is present within the top three suspected phases for 99% of all test data. Therefore, though manual intervention may occasionally be required to handle complex patterns, the problem is greatly simplified by allowing the user to choose from a small set of probable phases.

2.3.2 Incorporating non-stoichiometry

To determine whether the accuracy of our model extends to non-stoichiometric materials, we built a test set of XRD patterns simulated from 20 experimentally reported solid solutions in the Li-Mn-Ti-O-F chemical space. These materials were manually selected from the ICSD to ensure that their compositions are different (greater than 0.05 mole fraction) than those of the stoichiometric phases already considered in the previous section. To isolate the effects of non-stoichiometry, diffraction patterns were simulated without including any experimental artifacts. We first restricted the training set to include only diffraction patterns derived from stoichiometric materials to illustrate the necessity of including additional reference phases with non-stoichiometry (*i.e.*, from hypothetical solid solutions). Similarly, JADE was applied to the new test set containing solid solutions while restricting its reference database to contain only stoichiometric phases. In doing so, neither method can be used to predict the exact compositions of the solid solutions. Instead, their prediction accuracy can be resolved into two components: (i) Is the predicted structure isomorphic to the true structure? (ii) How similar are the predicted and true compositions? Isomorphism was verified using the pymatgen structure matcher. Differences in compositions were quantified using the mole fraction distance between the barycentric coordinates of each phase in the Li-Mn-Ti-O-F chemical space (*i.e.*, with each constituent element representing a vertex). For example, the compositional difference between $LiMnO_2$ and $LiMn_{0.5}Ti_{0.5}O_2$ is quantified as 0.125 mole fraction since 0.5 out of 4 elements are interchanged in the formula unit.

In **Figure 2-4a**, we show the fraction of non-stoichiometric materials with structures correctly identified by the CNN and JADE when only stoichiometric reference patterns are used for training or profile matching. This case is labeled "Without NS" where NS denotes non-stoichiometry. The CNN correctly classifies the structures of 11/20 patterns, whereas JADE gives only 7/20 correct structural classifications. For the same set of data, we illustrate the differences between true compositions and those predicted by the CNN in **Figure 2-4b**. Errors in the predicted compositions range from 0.05 to 0.82 mole fraction, with an average value of 0.38. Therefore, when only stoichiometric reference phases are used, neither the deep learning algorithm nor conventional profile matching techniques can be utilized to reliably predict the structure or composition of non-stoichiometric materials from their diffraction patterns. This conclusion supports our initial expectations given that substantial off-stoichiometry is known to cause large

changes in the positions and intensities of diffraction peaks. Although data augmentation is useful (and necessary) to account for relatively weak deviations from ideality, it is not capable of extrapolating to larger changes well beyond those included in the training set.

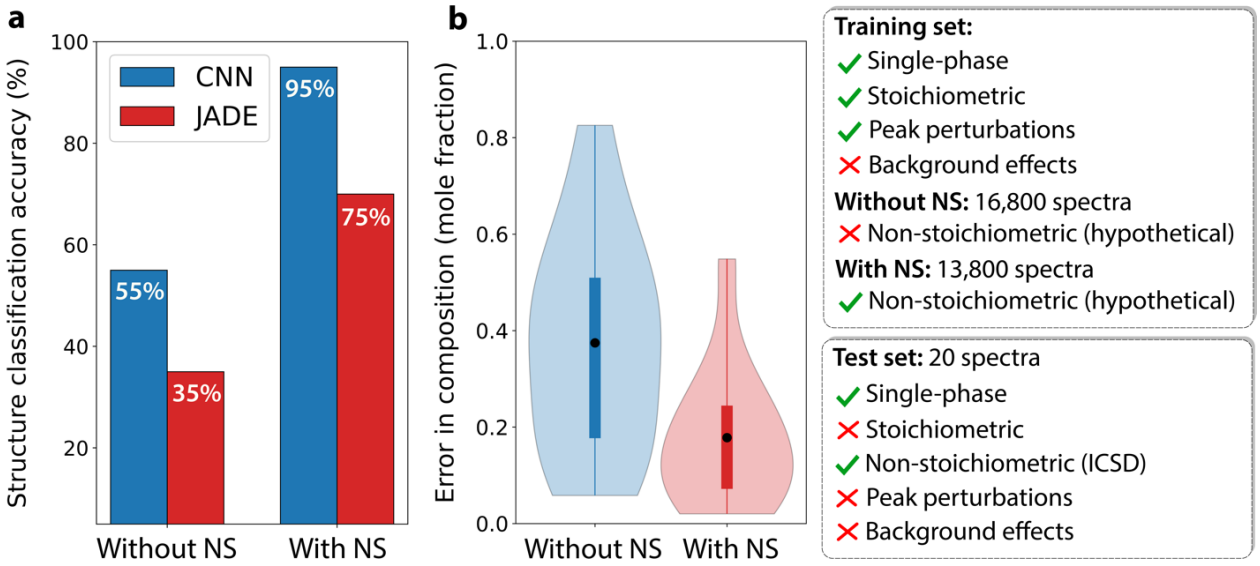

Figure 2-4. Test results from the automated identification of non-stoichiometric phases. (a) For a set of diffraction patterns derived from 20 solid solutions, the fractions of structures correctly identified by the CNN and JADE are shown in two cases: (i) when the training set includes only stoichiometric reference phases (Without NS), and (ii) when the training set is augmented with hypothetical solid solutions (With NS). (b) For the same set of patterns, differences between true compositions and those predicted by the CNN are quantified by their mole fraction difference. Violin plots illustrate the full distribution of errors, whereas embedded boxes range from lower to upper quartiles. Black dots are used to denote the average probability given in each case.

A proper treatment of non-stoichiometry necessitates additional reference phases with compositions that more closely match experimentally observed solid solutions. To this end, we introduced XRD patterns simulated from hypothetical solid solutions spanning the Li-Mn-Ti-O-F space into the training set. In addition to the 21,000 patterns obtained from the 140 stoichiometric materials, 17,250 new patterns were derived from 115 hypothetical solid solutions (115 materials × 150 augmentations). Perturbations were applied *via* the data augmentation procedure described in the Methods, and 80% of the resulting data was used to re-train the CNN. For comparison, the same set of hypothetical solid solutions were also added to the reference database used by JADE. Both updated models were then applied to the test set containing 20 diffraction patterns simulated from the experimentally reported non-stoichiometric materials. The fraction of structures correctly identified by each method is displayed in **Figure 2-4a**, labeled "With NS". In contrast to earlier results, the CNN and JADE achieve much higher accuracies of 95% and 70%, respectively. These improvements in performance are realized without sacrificing much accuracy in the classification of stoichiometric materials – our updated model correctly identifies 89% of phases across the test set containing simulated diffraction patterns with mixed artifacts, a decrease of only 3% compared to the CNN trained only on stoichiometric phases. In **Figure 2-4b**, we present the updated distribution of errors in compositions given by the CNN trained with non-stoichiometric phases. Differences between the predicted and true compositions now range from 0.02 to 0.54 mole

fraction, with an average value of 0.18. Hence, these results highlight the advantages of including non-stoichiometric reference phases, which nearly doubles the number of correctly identified structures and reduces compositional errors by ~50% when classifying experimentally reported solid solutions.

2.3.3 Multi-phase classification

Extending the CNN to characterize mixtures of materials, we constructed three new test sets, each containing 1,000 simulated multi-phase diffraction patterns. These tests were designed to mimic samples with multiple phases by creating linear combinations of single-phase diffraction peaks derived from 140 stoichiometric reference phases in the Li-Mn-Ti-O-F chemical space. The first two sets consider mixtures generated from randomly selected two- and three-phase combinations with equal weight fractions of the reference phases. In the last set, we probe the effects of impurity phases by simulating two-phase patterns where the weight fractions of the majority and minority phases are randomly set to constitute 70-90% and 10-30% of the mixture, respectively. In all three test cases, data augmentation is applied using mixed artifacts (peak shifting, broadening, and texture as well as a diffuse and noisy background signal) so that the resulting patterns provide an realistic representation of experimental measurements.

In addition to our newly developed branching algorithm (denoted B-CNN hereafter), multi-phase identification was performed using three other techniques for comparison: (i) based on the work of Maffettone *et al.*[41], a "single-shot" approach (S-CNN) was employed such that the two or three materials with the highest probabilities are chosen for each two- or three-phase mixture, respectively; (ii) by training the CNN explicitly on simulated multi-phase patterns (M-CNN) as described in the work of Lee *et al.*[42], entire mixtures of phases are directly predicted as opposed to separately identifying individual phases; (iii) using JADE to obtain a list of suspected phases for each mixture based on profile matching, the two or three highest-ranked materials are chosen for two- and three-phase patterns, respectively. Given that method (ii) requires many possible linear combinations of single-phase patterns to produce sufficient data for training, only ideal diffraction patterns were used without applying any data augmentation.

In **Figure 2-5a**, we show the fraction of phases correctly identified by each of the four methods when tested on two- and three-phase mixtures with equally distributed weight fractions. Among all of the techniques considered here, our newly developed B-CNN algorithm achieves by far the highest accuracy, correctly identifying 87% and 78% of all materials from two- and three-phase patterns, respectively. This outperforms the previous methods based on deep learning, with the S-CNN and M-CNN giving accuracies of 70% (54%) and 65% (58%) in the classification of two-phase (three-phase) mixtures. Despite their similarity in performance, these two approaches highlight separate limitations. Recall that the M-CNN does not utilize data augmentation to expand the diversity of its training set, and therefore often fails when applied to diffraction patterns containing large perturbations arising from experimental artifacts. In contrast, the S-CNN accounts for possible artifacts through physics-informed augmentation (as in our approach) and consequently is more robust against changes in the diffraction patterns. However, since the S-CNN identifies all phases in a "single shot" without subtracting known diffraction peaks, it leads to misclassifications when similar reference phases produce comparable probabilities for a given pattern. The B-CNN improves upon both shortcomings using an iterative process of single-phase identification and profile subtraction to achieve higher accuracy. Furthermore, by maximizing the probability over all phases in the predicted mixture, the B-CNN ensures that the first iteration of phase identification is not over-prioritized. If only the most probable phase is evaluated at each

step without maximizing probability over the entire mixture, lower accuracies of 78% and 69% are given across two- and three-phase mixtures, respectively.

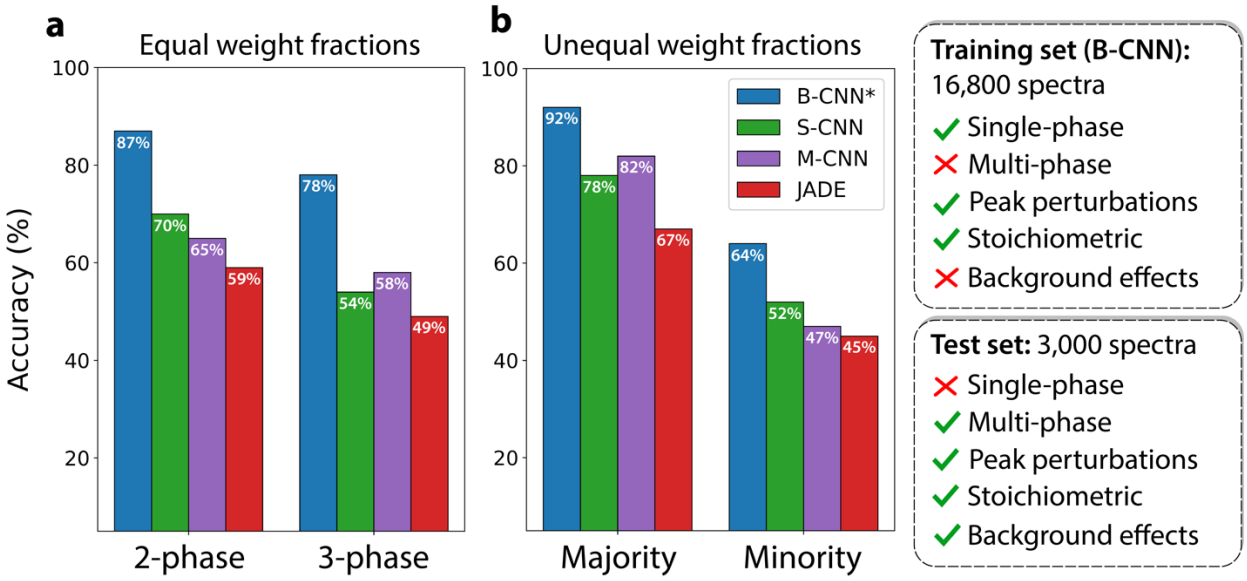

Figure 2-5. Test results from the automate identification of multi-phase XRD patterns. (a) The fractions of phases correctly identified by the B-CNN (*introduced in this work) when applied to simulated patterns of two- and three-phase mixtures with equally distributed weight fractions. For comparison, accuracies obtained using two existing methods (S-CNN[13] and M-CNN[8]) are shown, in addition to results from JADE. (b) These same techniques are applied to diffraction patterns of two-phase mixtures with unequally distributed weight fractions of 10-30% and 70-90%. Accuracies are divided into the identification of majority and minority phases.

In **Figure 2-5b**, we compare the accuracy of each approach for the classification of majority/minority two-phase mixtures. The B-CNN again outperforms all other evaluated approaches. However, the reliability of our model varies substantially in the identification of majority versus minority phases. The B-CNN correctly classifies 92% of all majority phases, matching its performance across single-phase patterns and therefore suggesting the presence of impurity phases has little to no effect on majority phase identification. Identifying minority phases, on the other hand, presents a greater challenge, as reflected by a lower accuracy of 64% given by the B-CNN. We note that most misclassifications occur due to imperfect applications of profile subtraction that occasionally leave behind residual intensities or subtract some diffraction peaks associated with the minority phase of interest. Despite this limitation in the *identification* of minority phases, the model generally performs reliably in their *detection*. Recall that the number of phases in a mixture is determined by halting the B-CNN when all diffraction intensities fall below 5% of the initially measured maximum intensity. With this cutoff, the B-CNN correctly reports the presence of a second phase in 93% of the two-phase mixtures with unequally distributed weight fractions. For comparison, when the B-CNN is applied to simulated single-phase patterns with mixed artifacts using the same cutoff intensity of 5%, the number of phases is overestimated in only 9% of the samples. The key component enabling a reliable prediction for the number of phases is the approach to profile subtraction. Here, known diffraction peaks are fit to the pattern through DTW so that their subtraction yields a new pattern that accurately represents the mixture

minus the phase(s) that has already been identified. This capability is particularly useful in the optimization of synthesis procedures, where it is of interest to know whether the formation of a targeted product is accompanied by some impurity phase.

2.3.4 Application to experimental patterns

As a final demonstration of the generalizability of our approach, the B-CNN was applied to experimentally measured XRD patterns in the Li-Mn-Ti-O-F chemical space. In **Table 2-1**, we list the fraction of phases correctly identified by the CNN versus JADE, with results categorized by the artifacts and number of phases included for each class of samples. For the classification of pristine diffraction patterns, the CNN correctly identifies all ten phases considered. Interestingly, JADE incorrectly classifies one material (Li_2TiO_3) from this category. Upon further inspection, the error is attributed to large deviations in the relative peak intensities between the measured and ideal patterns of Li_2TiO_3, possibly caused by stacking faults in the sample. In the analysis of data with diffuse and noisy background signals, the CNN correctly identifies all but one material (anatase TiO_2), likely due to the fact that it exhibits significant diffraction peaks at low values of 2θ where the amorphous background is strong. JADE is found to be more sensitive to background effects as it yields five misclassifications across these 20 patterns. These misclassifications occur because JADE fails to index peaks that blend in with the background signal and have low intensities or broad widths after a baseline correction is applied. The CNN is more robust against these perturbations since it is trained on patterns that have diffraction peaks with varied intensities and widths.

For patterns containing peak shifts, the CNN correctly identifies five out of six phases. In contrast, JADE struggles to handle changes in peak positions, identifying only two phases from this category. This highlights a key weakness of profile matching techniques, which fail when there is weak overlap between measured and simulated diffraction peaks owing to a shift in 2θ. Fortunately, because the CNN can handle these changes through data augmentation, its performance remains reliable in the classification of patterns with peak shifts. When diffraction peaks are broadened, the CNN and JADE correctly identify five and four phases, respectively, from the five measured patterns. The single misclassification from JADE occurs for Li_2MnO_3 owing to a strong overlapping of its neighboring diffraction peaks, an effect which is accounted for by the CNN during training. For the six patterns with changes in their peak intensities, the CNN correctly classifies five phases while JADE identifies four. The misclassification made by the CNN occurs because the varied peak intensities closely resemble those of a hypothetical solid solution ($Li_{0.5}Mn_{1.5}TiO_4$) that is isomorphic to the true phase ($LiMnTiO_4$). Across non-stoichiometric materials, the CNN correctly predicts all four materials to adopt the rocksalt structure, whereas JADE finds only three phases to be rocksalt. For both methods, the predictions are facilitated by the introduction of hypothetical solids solutions; without including these additional reference phases, neither the CNN nor JADE predicts any of the four samples to be rocksalt-structured.

For the classification of multi-phase mixtures, JADE provides limited accuracy. Only 7/10 and 9/15 phases are correctly identified from two- and three-phase patterns, respectively. Such limitations in accuracy can be attributed to the inability of profile matching techniques to distinguish between diffraction peaks produced by several phases, which often overlap with one another. The B-CNN adeptly overcomes these limitations and correctly identifies 10/10 and 13/15 phases in the two- and three-phase mixtures, respectively. Hence, the benefits provided by deep learning are highlighted by the noticeable disparity between the performance of the CNN versus JADE, especially when applied to multi-phase samples. This advantage is vital to assist in targeted

synthesis, considering that attempts to produce novel inorganic materials are frequently impeded by the appearance of multiple impurity phases. Our deep learning approach can therefore be used to identify not only desired products, but also impurity phases, which provide insight into why a given synthesis procedure failed and inform future attempts.

The results from testing the CNN on experimentally measured XRD patterns (**Table 2-1**) closely match the performance on simulated data. For example, in cases where we include a single type of artifact, the CNN correctly identifies 94% of the phases from both simulated and experimentally measured single-phase patterns. This lends credence to the simulation-based test cases that are rich in data (*e.g.*, a total of 4,200 single-phase test patterns were derived from stoichiometric materials) and suggests that the simulated data used for training and testing provide a realistic representation of experimental measurements.

Table 2-1. Fractions of materials correctly identified by the CNN and JADE when applied to experimental XRD patterns designed to sample possible artifacts arising during sample preparation and synthesis. For patterns of non-stoichiometric materials, a classification is considered correct if the predicted structure is isomorphic to the true structure.

Experimental procedure	Anticipated artifact	CNN	JADE
Single-phase			
Pristine samples	None	10/10	9/10
Kapton tape overlaid	Diffuse baseline	9/10	8/10
Rapid XRD scan	Noisy baseline	10/10	7/10
Thick samples	Shifts in 2θ	5/6	2/6
Ball milled	Broadening	5/5	4/5
Partially disordered	Intensity variation	5/6	4/6
Solid solutions	Non-stoichiometry	4/4	3/4
Multi-phase			
Two-phase mixtures	None	10/10	7/10
Three-phase mixtures	None	13/15	9/15
	Overall accuracy:	**71/76 (93.4%)**	**53/76 (71.4%)**

2.4 Discussion

In summary, we developed an improved deep learning technique that can reliably automate the identification of inorganic materials from XRD patterns. A key advantage of our approach is the physics-informed data augmentation procedure that accounts for several experimental artifacts commonly observed after sample preparation and synthesis. Conventional profile matching techniques often fail when materials variations cause large differences between observed and simulated diffraction peaks, requiring manual intervention to analyze any irregularities and identify the samples of interest. In contrast, our CNN learns these differences during training, and

therefore can autonomously perform phase identification from complex patterns. These benefits are highlighted by the test results presented in this work, which show that the performance of profile matching quickly deteriorates as larger perturbations are applied to the diffraction patterns, whereas the CNN remains reliable in the presence of such perturbations. Furthermore, even though our model is trained only on patterns that account for three types of artifacts (strain, texture, and domain size), it is demonstrated to successfully generalize to patterns outside of the training set. For example, our algorithm achieves a high accuracy for the identification of samples with diffuse and noisy baseline signals, as well as for samples containing unexpected artifacts (*e.g.*, possible stacking faults in Li_2TiO_3).

Of the artifacts considered in our work, changes in peak positions are shown to be the most challenging to deal with, comprising nearly half of all misclassifications made by the CNN when applied to the simulated diffraction patterns of single-phase stoichiometric materials. Because peak positions are derived from the spacings between crystallographic planes, and therefore the lattice parameters of the material, it is difficult to distinguish between isomorphic phases when their structures have a significant degree of strain. We find that our model provides an optimal treatment of changes in peak positions by including samples with as much as $\pm 4\%$ strain in the training set, which is unlikely to be exceeded in experiment unless the materials contain substantial off-stoichiometry. Indeed, tests involving an increased magnitude of strain in the training set led to decreased accuracy during testing owing to degeneracies between the diffraction patterns of similar phases. In general, the bounds used for data augmentation should reflect the experimental system at hand; for example, larger perturbations may be beneficial in cases where certain artifacts are expected to dominate (*e.g.*, epitaxial strain in thin films). When using the approach supplied in our repository (https://github.com/njszym/XRD-AutoAnalyzer), these bounds can be manually specified for any given set of reference phases. To avoid degeneracy of patterns in the training set, the number of reference phases should be constrained to include only those that are expected to arise in experiment – for synthesis, these can be chosen to reflect the composition space spanned by the precursors used and the possibility of reactions with oxygen, water, or CO_2 in air.

The importance of peak positions is further highlighted by our tests involving non-stoichiometric materials. Varying the composition of a material typically leads to changes in its lattice parameters, which in turn shifts the positions of its diffraction peaks. As a result, when the CNN is trained only with stoichiometric reference phases, it frequently fails to identify the structures of non-stoichiometric materials. Because the model is trained to identify individual phases, rather than their symmetry, it does not necessarily learn the subtle relationships between peak positions imposed by the space group of each structure. Instead, it considers the positions of all peaks and makes a comparison with known phases in the training set. Therefore, when non-stoichiometry causes large shifts in the positions of diffraction peaks, the CNN will struggle if it has no reference phase available with comparable peak positions. With this in mind, we improved the treatment of non-stoichiometric materials by building a library of hypothetical solid solutions following Vegard's law. After adding their diffraction patterns to the training set, the CNN correctly identifies the structures for 95% of the non-stoichiometric materials considered during testing. We note that this approach is successful because the lattice parameters of most solid solutions follow Vegard's law with only minor deviations. When deviations do occur, data augmentation ensures that the match between hypothetical and experimentally observed phases need not be exact for the model to maintain a high level of accuracy for the identification of the material's structure.

Despite the improved prediction of structure enabled by introducing hypothetical solid solutions to the training set, predicting the compositions of non-stoichiometric materials remains challenging. This limitation can be understood by considering the effects of non-stoichiometry on diffraction peak intensities, which are influenced by the structure's internal cell coordinates and site occupancies. Given the similarity of structural frameworks between materials forming solid solutions, changes in cell coordinates are usually small and therefore do not contribute significantly to differences in peak intensities. Changes in site occupancies, however, strongly influence peak intensities owing to the distinct scattering factors of substituted species. As opposed to changes in lattice parameters that can be described by Vegard's law, an automatic prediction of site occupancy is more difficult to achieve because site occupancies can redistribute in solid solutions. For example, partial inversion (*i.e.*, swapping Wyckoff positions) between lithium and transition metal ions has been observed in spinel $LiMn_{2-x}Ti_xO_4$[66]. Such differences give rise to errors in predicted compositions, not structures, because site occupancies control peak intensities while leaving peak positions relatively unaffected. Hence, we reiterate that our approach is not designed to give precise refinements of composition, but rather to provide a reliable prediction of structure and an estimate of composition.

Beyond the scope of this work, future efforts may be conducted to design a more accurate prediction of site occupancies so that refinement can be carried out autonomously. A recent report by Mattei *et al.* has shown some progress toward this end, providing an approach to enumerate many possible distributions of site occupancies with the goal of identifying the best match with experimental measurements[67]. As their approach requires that the structural framework of the suspected phase be known prior to refinement, our model may prove useful in coordination with their algorithm. The results from our CNN may also provide a useful starting point for manual Rietveld refinement as they contain necessary information regarding the composition and structure of each phase identified in a pattern. An estimation of the lattice parameters can be given for these phases based on their corresponding entries in the ICSD. Furthermore, because DTW measures the shift in 2θ between experimental and simulated diffraction peaks, it is possible that our model can provide a more precise estimation of the lattice parameters by relating peak shifts with strain parameters through Bragg's law. Demonstrating this capability is outside the scope of the current report but may be considered in future work.

When samples contain more than one material, new challenges arise as diffraction peaks often overlap and can be difficult to distinguish. To handle multi-phase samples, we designed a branching algorithm that iterates between phase identification and profile subtraction to identify the combination of phases that maximizes the average probability given by the CNN. This approach yields exceptionally high accuracy across simulated and experimentally measured multi-phase XRD patterns, exceeding the performance of profile matching techniques and recently published methods based on deep learning. The advantages of our branching algorithm can be summarized by two main points. First, by training only on single-phase patterns, we avoid the combinatorial explosion of training samples that would arise if multi-phase patterns were instead used. Because the number of pristine reference patterns is kept low, many experimental artifacts can be included through physics-informed data augmentation, which ensures the model is robust against perturbations caused by defects or impurities. Second, our algorithm avoids confusion between phases with similar reference patterns by identifying phases in a one-by-one manner and iteratively subtracting their diffraction peaks from the pattern until all non-negligible intensities have been accounted for. The removal of known peaks prevents the algorithm from overestimating the number of phases in a sample, which would otherwise occur if the probability distribution

given by the CNN was assumed to represent a mixture of phases (*e.g.*, assuming all phases with a probability $\geq 50\%$ exist in a given sample).

2.5 Conclusion

We have demonstrated that a deep learning algorithm based on a CNN can be trained to identify inorganic materials from complex diffraction patterns. Physics-informed data augmentation was shown to accurately account possible experimental artifacts in measured diffraction patterns, therefore improving the generalizability of the CNN. Simulated patterns derived from hypothetical solid solutions were also added to the training set, which improves the performance of the model when dealing with off-stoichiometric samples. For samples containing multiple phases, an iterative process of phase identification and profile subtraction was designed to maximize the probability given by the CNN over all phases in the predicted mixture, which performs well when applied to multi-phase samples. The proposed accuracy of our deep learning approach was validated with respect to simulated and experimentally measured diffraction patterns.

Although our current tests focus on materials in the Li-Mn-Ti-O-F space, the algorithm developed here can be applied to any arbitrary composition space given a set of reference phases, which can be extracted from existing crystallographic databases. Based on the 255 reference phases considered in this work, the entire process of pattern simulation, data augmentation, and model training was completed in 20 hours on a single compute node with 16 CPUs. Because the number of training samples required by our method scales linearly with the number of reference phases, new models can be created on much broader composition spaces without requiring excessive amounts time or computational resources. The compositions included during training should be chosen to reflect anticipated elements in the samples being characterized, and therefore it is generally not necessary to include *all* compositions in a single model. Once a model is trained for a given chemical space, it can be applied rapidly and automatically to each experimental XRD pattern to predict what phases are in the sample. Additionally, new reference phases can be introduced to the model at any time without requiring the regeneration of training data for existing phases. Given the efficiency of our approach and the promising results illustrated throughout this work, we suggest that the algorithm developed here may be used to effectively accelerate materials discovery by incorporating automatic phase identification to support high-throughput and autonomous experimental workflows.

CHAPTER 3 Leveraging in-line phase analysis to adaptively steer XRD measurement in real time

The work presented in this chapter is based, often verbatim, on the following publication:
***N. J. Szymanski**, C. J. Bartel., Y. Zeng, M. Diallo, H. Kim, and G. Ceder, "Adaptively driven X-ray diffraction guided by machine learning for autonomous phase identification." npj Computational Materials 2023, 9, 31.*

ML is increasingly being used to assist and improve materials characterization, enabling automated interpretation of experimental results with techniques such as XRD and electron microscopy. Because ML models are fast once trained, there is a key opportunity to bring interpretation in-line with experiments and make on-the-fly decisions to achieve optimal measurement effectiveness, which creates broad opportunities for rapid learning and information extraction from experiments. Here, we demonstrate such a capability with the development of autonomous and adaptive XRD. By coupling an ML algorithm with a physical diffractometer, this method integrates diffraction and analysis such that early experimental information is leveraged to steer measurements toward features that improve the confidence of a model trained to identify crystalline phases. We validate the effectiveness of an adaptive approach by showing that ML-driven XRD can accurately detect trace amounts of materials in multi-phase mixtures with short measurement times. The improved speed of phase detection also enables *in situ* identification of short-lived intermediate phases formed during solid-state reactions using a standard in-house diffractometer. Our findings showcase the advantages of in-line ML for materials characterization and point to the possibility of more general approaches for adaptive experimentation.

3.1 Introduction

Efficient materials characterization is critical to the design of improved technologies. Microscopic and spectroscopic techniques produce large amounts of data that traditionally require time-consuming analysis by an expert, which limits the rate of materials development and precludes their use in automated workflows[68]. Recently, ML has been applied to interpret characterization data more rapidly[69]. For example, autoencoders have been developed to segment images from electron microscopy and identify distinct atoms[70], defects[71], and microstructures[72]. Deep learning has also found use in spectroscopy, where CNNs can be trained to identify crystalline phases from XRD patterns[41,61] or chemical species from Raman spectra[73]. While such methods effectively automate the analysis step of materials characterization, an opportunity exists to fundamentally rethink the measurement step by leveraging in-line ML to interpret experimental output as it becomes available and using this information to modify measurements within a closed-loop process that we call *adaptive characterization*. As will be demonstrated in this work, adaptive characterization can be applied to steer an experiment along the most efficient path toward precise decision making, circumventing the need for iterative experimentation.

There are several notable examples of adaptive characterization techniques developed in recent years. Bayesian optimization has been applied to raster objects in scattering[74] and electron or probe microscopy[75], leading to reduced measurement time relative to grid-based sampling. Such methods have also been used to guide measurements toward the verification of scientific hypotheses by designing surrogate models with built-in physical constraints[76]. Alternatively, decisions can be made with reinforcement learning, *e.g.*, to regulate the time spent scanning samples at a beamline depending on their scattering strength[77]. While much of the past work has applied adaptive characterization to fixed samples with unchanging properties, we show in this paper that the use of in-line ML analysis can also enable improved monitoring of dynamic processes where rapid measurements are required to capture transient states.

For structural characterization, XRD is a prime example of a technique that requires fast and precise measurements when applied in-line with experiments. *In situ* XRD is widely used to monitor reactions and detect the formation of short-lived intermediate phases that often influence the final reaction products[78]. Similarly, *operando* XRD is used to track phase transformations in battery materials during cycling, thus providing mechanistic insight into their performance[79]. For either application, XRD scans must be performed quickly enough to capture short-lived intermediate states while also producing high-resolution data that can be analyzed reliably *post hoc*. These two requirements compete with one another as short scans typically lead to noisy XRD patterns, complicating phase identification. High-brilliance radiation from a synchrotron light source may be used to generate high-resolution patterns with short measurement times, though access to such facilities is limited to a select number of users and experiments each year. Alternatively, we propose that ML can be used to adaptively develop high resolution around features that matter most for phase identification, even on standard in-house diffractometers. Such highly efficient data collection can be accomplished only by closing the loop between experiments and ML-enabled analysis, such that rapid and mathematically optimized decisions are made autonomously and on-the-fly to acquire signal in areas that provide maximal information to confirm the presence (or absence) of certain phases.

Here we formulate an adaptively steered XRD technique for autonomous phase identification, driven by an ML algorithm based on a CNN. Uncertainty quantification is used to decide when additional measurements are needed, while class activation map analysis dictates where those measurements are performed. This approach is validated it on three test cases with increasing complexity based on materials from the Li-La-Zr-O and Li-Ti-P-O chemical spaces. These tests reveal that adaptive XRD consistently outperforms conventional methods on both simulated and experimentally acquired patterns, providing more precise detection of impurity phases while requiring shorter measurement times. We further demonstrate that our ML approach can effectively guide XRD measurements for improved *in situ* characterization of solid-state reactions, with the synthesis of $Li_7La_3Zr_2O_{12}$ (LLZO) considered as an example. The use of adaptive scans to monitor LLZO synthesis led to the successful identification of a short-lived intermediate phase that would otherwise be missed by conventional measurements. These findings provide a clear proof of concept for adaptive characterization of dynamic processes, highlighting the opportunity for autonomous experiments driven by ML.

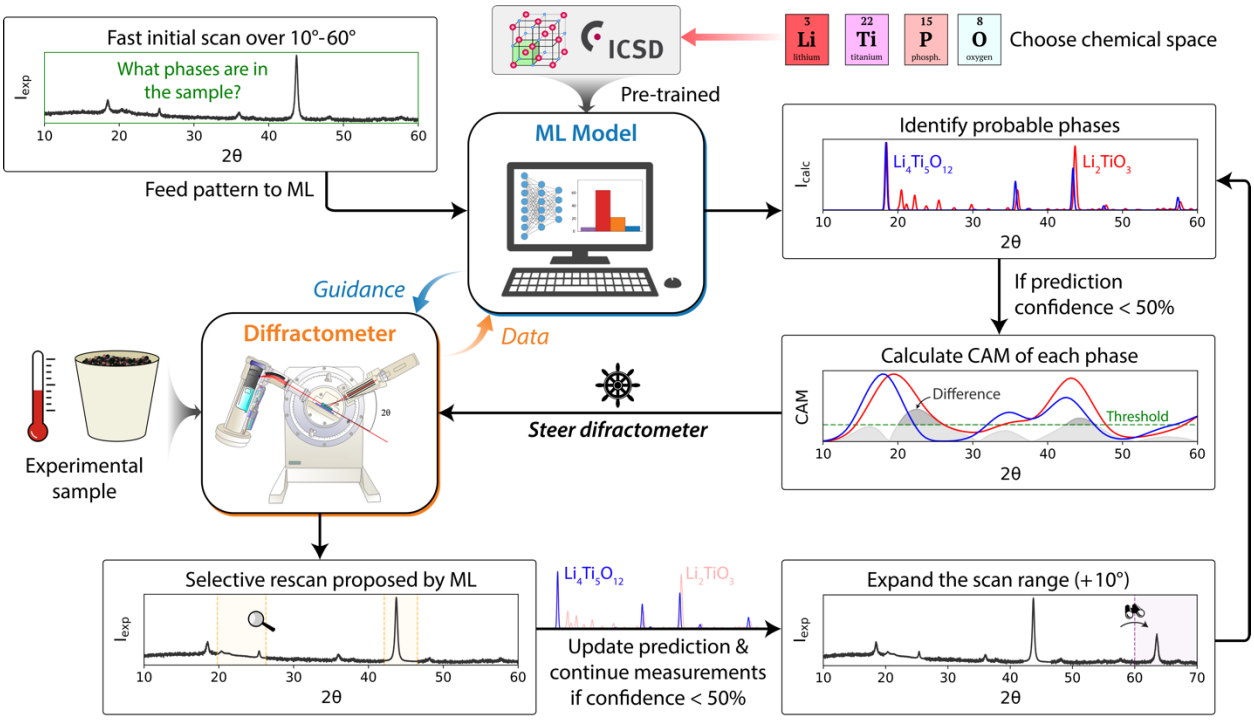

Figure 3-1. A schematic of adaptively driven XRD measurements, outlining the interactions between a physical X-ray diffractometer and a pre-trained ML model that performs automated phase identification.

3.2 Results

3.2.1 Adaptive XRD approach

Figure 3-1 shows the coupling between XRD and the ML algorithm that performs phase identification and controls the diffractometer. Each adaptive measurement begins with a rapid scan over a narrow range of $2\theta = [10°, 60°]$, which was optimized to conserve scan time while still including enough peaks to make a preliminary prediction regarding which phases are most likely present in the sample. Starting from lower angles (10-50°) leads to notably reduced accuracy, while starting from higher angles (10-70°) requires longer scans but does not lead to more accurate phase identification. After performing an initial scan over 10-60°, the pattern is fed to our previously developed deep learning algorithm, XRD-AutoAnalyzer[80]. This algorithm not only predicts a set of phases for a given sample, but also assesses its own level of certainty such that each phase has an associated confidence ranging from 0% to 100%. Because higher confidence is correlated with more reliable predictions, we use it as a metric to decide when a pattern has sufficient resolution to accurately identify all phases in a sample. A confidence cutoff of 50% is found to provide a good balance between measurement speed and prediction accuracy. In cases where the prediction confidence is less than 50%, the ML algorithm can request additional data from the diffractometer in one of two ways:

1) *Resampling* a subset of $2\theta \subseteq [10°, 60°]$ with increased resolution (slower scan rate) to clarify specific peaks that lead to maximal confidence improvement,
2) *Expanding* $2\theta_{\max} > 60°$ with a fast scan rate to detect additional peaks.

To select which 2θ should be scanned with increased resolution, we make use of Class Activation Maps (CAMs) designed to highlight features that contribute most to the classification(s) made by a deep learning model[81]. The CAM for a given XRD pattern is calculated as a function of 2θ and is expected to be large in regions containing important peaks for phase identification[61]. As a result, CAMs tend to be maximal around the most intense peaks in each pattern. However, sampling such features with increased resolution usually reveals little new information as the most prominent peaks can already be detected with low-resolution measurements. Therefore, we instead prioritize resampling in areas of 2θ where the *difference* between the CAMs of the two most probable phases (proposed by XRD-AutoAnalyzer) exceeds a user-defined threshold. This approach ensures that high-resolution scans are used to clarify peaks that distinguish phases with similar XRD patterns.

In cases where there is significant peak overlap between different phases at low 2θ, the scan range can be expanded to reveal additional peaks that assist in disentangling them. However, because measurements carried out at higher 2θ often produce increasingly broad peaks with lower signal-to-noise ratios, they may not always lead to more accurate phase identification. To understand which parts of an XRD pattern provide the most useful information, we use the prediction confidence associated with each phase proposed by XRD-AutoAnalyzer based on $2\theta = [10°,\ 60° + n10°]$ for n between zero ($2\theta_{\max} = 60°$) and eight ($2\theta_{\max} = 140°$). The predicted phases from each subset of 2θ are aggregated into an ensemble (P_{ens}), where the prediction confidence is used to form a weighted average as follows:

$$P_{\text{ens}} = \frac{\sum_{10}^{2\theta_i} c_i P_i}{n+1} \tag{1}$$

In this equation, P_i represents each prediction over $[10, 2\theta_i]$, c_i is the confidence of that prediction, and $n + 1$ gives the total number of 2θ-ranges included in the ensemble. In contrast to the typical analysis technique whereby an individual prediction is made based on a given XRD pattern, the ensemble approach decomposes the pattern into several distinct but overlapping regions from which separate predictions are made and subsequently aggregated using the confidence-weighted sum described in Eqn. 1.

The adaptive XRD approach presented here integrates resampling and expansion of 2θ into one single workflow (**Figure 3-1**). Based on early data obtained from a rapid initial scan over $2\theta = [10°, 60°]$, XRD-AutoAnalyzer makes a preliminary prediction regarding which phases are most likely in the corresponding sample. If the confidence associated with this prediction is less than 50%, a selective rescan is performed over regions of 2θ where the difference between the CAMs of the two most probable phases exceeds a threshold of 25%. An updated prediction is made based on the resampled pattern and the confidence is assessed. If it remains less than 50%, higher angles are scanned (+10° at each step) to detect additional peaks. This iterative process of phase identification, resampling, and expansion is repeated until the prediction confidence exceeds 50% or until a maximum angle of 140° is reached. The requirement of 50% confidence is applied to *all* suspected phases in the mixture, not only the two most probable. In cases where multiple phases have high uncertainty, more than one round of resampling may be performed at each iteration, thus ensuring that the algorithm remains robust on multi-phase samples.

3.2.2 Evaluation of adaptive XRD on simulated patterns

We first evaluated the performance of the adaptive XRD approach in a simulated environment. XRD-AutoAnalyzer was separately trained in two chemical spaces, Li-La-Zr-O and Li-Ti-P-O, which each contain a rich variety of compositions and structures with applications in solid-state batteries. The algorithm requires a list of previously reported phases to be trained on, and as such, all unique materials occupying these chemical spaces were extracted from the ICSD. This included 28 and 45 stoichiometric phases in the Li-La-Zr-O and Li-Ti-P-O spaces, respectively, from which a total of 8000 patterns were simulated including 1400 single-phase, 2400 two-phase, and 4200 three-phase samples. While XRD-AutoAnalyzer is trained only on single-phase patterns, it readily interprets multi-phase samples by iterating between phase identification and peak subtraction following the procedure described in previous work[80]. In the samples containing more than one phase, weight fractions were randomly sampled between 20% and 80%. To mimic the limitations of data acquired experimentally, all simulated patterns were stochastically modified based on artifacts including background noise, strain, texture, and small particle size (Methods section). These are commonly observed in real samples and can alter the positions, intensities, and widths of the corresponding diffraction peaks. The signal-to-noise ratio (s/n) is related to the scan time (t) as follows:

$$s/n = C\sqrt{t} \tag{2}$$

Where C is a scaling constant, which for this work was fit to experimental data obtained from XRD scans XRD scans on a sample of Li_2CO_3 (Sigma Aldrich) using a Bruker D8 Advance diffractometer. The signal-to-noise ratio used in our simulated tests, in addition to the sampling density of 2θ, dictates the total effective scan time of each pattern. A shorter scan time is more efficient but will generally lead to less accurate phase analysis. To probe this relation, we duplicated all 8000 simulated patterns into 10 distinct datasets with varied sampling density (0.02°-0.04°) and signal-to-noise ratio (20-60), corresponding to an effective scan time ranging from 5 to 30 min.

The effectiveness of adaptive XRD in the simulated environment was first tested by limiting the algorithm to only perform resampling in a fixed range of $2\theta = [10°, 60°]$. Starting from patterns with minimal resolution (effective scan time of 5 min), XRD-AutoAnalyzer made initial predictions regarding which phases were present in each sample. When the prediction confidence was less than 50%, high-resolution data (effective scan time of 30 min) was added in regions of 2θ where the CAM differences between suspected phases exceeded 25%, and the total effective scan time was proportionally increased. **Figure 3-2a** shows (in blue) the F_1-score as a function of scan time for phase identification performed by the adaptive algorithm in the Li-La-Zr-O (top panel) and Li-Ti-P-O (bottom panel) spaces. For comparison, we also show the F_1-score associated with phase identification based on conventional scans with uniform resolution and iteratively longer measurement time (in black). With either sampling technique, more phases are accurately identified from patterns with higher resolution (longer effective scan time); however, adaptive XRD reaches convergence more rapidly than its conventional counterpart. The F_1-score achieved with adaptive sampling exceeds 0.88 in 10-15 min of effective scan time for each pattern, whereas conventional sampling requires 25-30 min per pattern to reach the same level of accuracy. The rapid convergence of adaptive XRD demonstrates that it leverages low-resolution measurement data to effectively build a probability density for the likely phases in each sample, from which it identifies the optimal regions of 2θ that should be prioritized to distinguish these phases. The upper limit of the F_1-score observed for both adaptive and conventional sampling can be attributed to the presence of simulated artifacts (*e.g.*, strain and texture) as well as peak overlap

between different phases over the current range of $2\theta \in [10°, 60°]$. Since these issues are not resolved by reducing background noise, the use of longer scan times (> 30 min) leads to only marginal improvement in the F_1-scores of conventional and adaptive analyses.

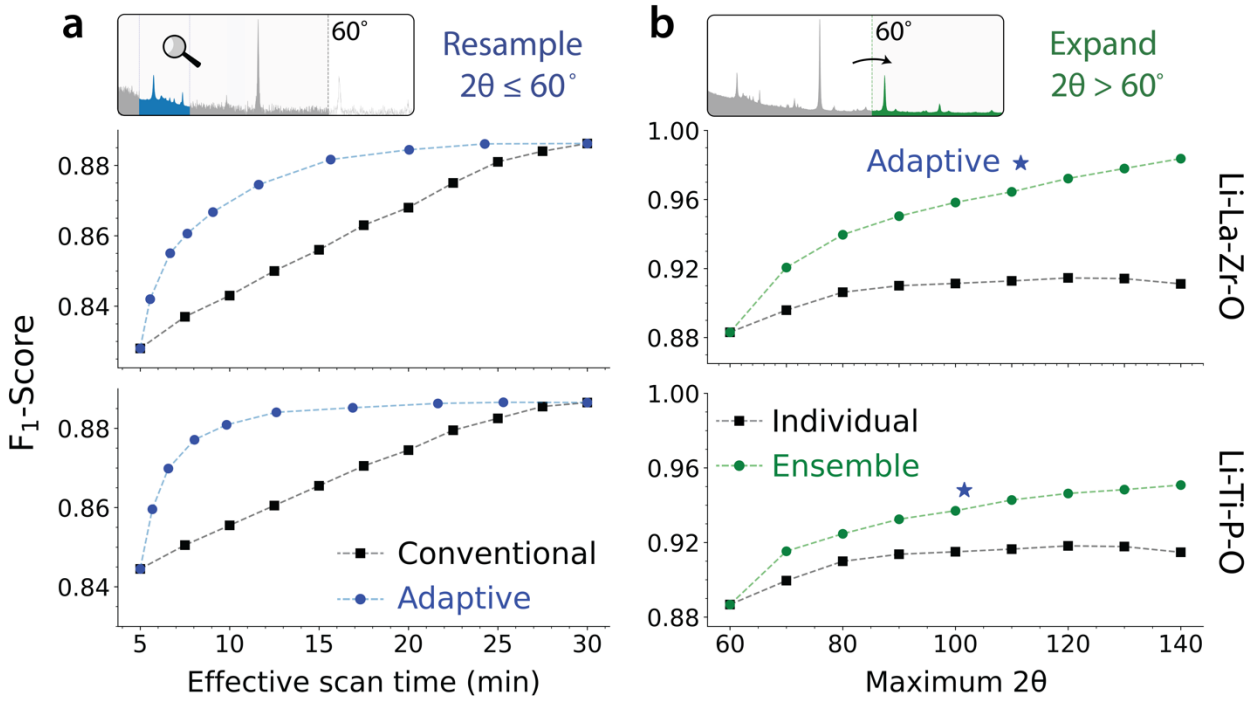

Figure 3-2. F_1-scores achieved by XRD-AutoAnalyzer when applied to XRD patterns generated in a conventional or adaptive manner. (a) Conventional results (black squares) were obtained on patterns with incrementally improved resolution over $2\theta = [10°, 60°]$, whereas adaptive results (blue circles) were found by resampling a subset of $2\theta \subseteq [10°, 60°]$ with high resolution. (b) Individual results (black squares) were calculated by analyzing patterns with distinct maxima ($2\theta_{max}$), which were aggregated in a confidence-weighted sum to form the ensemble predictions (green circles). Adaptive results (blue star) were obtained by halting expansion of $2\theta_{max}$ when the prediction confidence exceeded 50%.

We used the simulated XRD dataset to quantify the extent to which the F_1-score can be improved by including information from higher 2θ in the analysis. Only high-resolution patterns (effective scan time of 30 min) were considered to isolate the effects of the scan range. XRD-AutoAnalyzer was applied to each individual range of $2\theta = [10°,\ 60° + n10°]$ for n between zero ($2\theta_{max} = 60°$) and eight ($2\theta_{max} = 140°$). These are treated separately and referred to as *individual* predictions hereafter. In comparison, *ensemble* predictions were formulated by aggregating the phases identified from all patterns available up to $2\theta_{max}$, as described in the previous section (Eqn. 1).

Figure 3-2b illustrates how the F_1-score varies with increasing scan range for patterns in the Li-La-Zr-O (top panel) and Li-Ti-P-O (bottom panel) spaces. Black (green) datapoints represent individual (ensemble) predictions on patterns with $2\theta_{max}$ denoted by the x-axis. The ensemble predictions show a monotonic increase in the F_1-score as higher angles are included, consistently outperforming the individual predictions. By aggregating all phases identified up to $2\theta_{max} = 140°$, exceptionally high F_1-scores of 0.98 and 0.95 are achieved. In contrast, scanning

higher 2θ does not necessarily lead to better performance for individual predictions, which show a maximal F_1-score of 0.91 at $2\theta_{max} = 120°$, followed by a decreasing score from $120°$ to $140°$. This trend arises from two effects: (1) signal-to-noise ratios decrease at higher 2θ as peaks become less intense, and (2) artifacts related to strain and small particle size cause larger changes to the positions and widths of peaks at higher 2θ. Because the ensemble approach weights each prediction by its associated confidence, it effectively ignores regions where background noise and/or artifacts mask the diffraction peaks, instead giving greater weight to regions where such peaks are more clearly distinguishable.

To keep the total measurement time minimal, higher 2θ should be scanned only when the prediction confidence from XRD-AutoAnalyzer is low. We demonstrate this policy by iteratively expanding $2\theta = [10°,\ 60° + n10°]$ until the prediction confidence exceeds 50% or until $2\theta_{max}$ reaches $140°$. The corresponding results are shown as blue stars in **Figure 3-2b**. With adaptive expansion of 2θ, high F_1-scores of 0.98 and 0.95 were obtained on patterns from the Li-La-Zr-O and Li-Ti-P-O datasets, respectively. These match the best F_1-scores obtained on a full scan range ($2\theta_{max} = 140°$), while also conserving measurement time as only angles up to 106° were sampled on average, therefore showing that the adaptive algorithm can effectively decide when higher angles are needed distinguish suspected phases.

3.2.3 Performance of adaptive XRD on experimental mixtures

As a more challenging test, the effectiveness of adaptive XRD as applied to impurity detection was evaluated on 240 two-phase mixtures prepared using different physical combinations of eight compounds from the Li-Ti-P-O and Li-La-Zr-O chemical spaces. All compounds were purchased in the form of solid powders and manually mixed such that the weight fraction of the minority phase was varied between 2% and 20% (Methods). For each compound, a reference phase from the ICSD was included during the training of XRD-AutoAnalyzer. We compare the effectiveness of determining these minority phases by using: (1) conventional measurements that sampled $2\theta = [10°, 80°]$ in 10 min, followed by automated phase identification with XRD-AutoAnalyzer applied *post hoc*; or (2) adaptive measurements with in-line phase identification and guided sampling, following the workflow outlined in **Figure 3-1**. Both scan techniques were applied with an Aeris X-ray diffractometer from Panalytical. Their relative performance is assessed using the impurity detection rate, defined as the percentage of phases correctly identified at a given weight fraction.

Figure 3-3 displays the detection rates for minority phases in the (a) Li-La-Zr-O and (b) Li-Ti-P-O spaces. When XRD-AutoAnalyzer is applied in-line with adaptive measurements, it successfully identifies $\geq 75\%$ of the minority phases at weight fractions $\geq 6\%$, even at short scan rates. In contrast, a much greater weight fraction of 15% is required to reach a detection rate of 75% using a conventional approach. The increased sensitivity of adaptive XRD holds true for all mixtures tested here, as it consistently detects smaller amounts of the minority phases when compared to conventional scans. Furthermore, it does so while using less scan time. As shown by the distributions of scan times in **Figure 3-3c**, adaptive measurements are completed more rapidly than conventional (10 min) ones, requiring an average scan time of only 6 min per pattern. The improved speed and accuracy of in-line, adaptive XRD is derived from two key advantages: (1) It automatically decides whether additional measurements are needed after a rapid initial scan, and if so, focuses high-resolution scans on regions of 2θ that are most likely to contain peaks associated with the suspected minority phases; and (2) it determines when higher 2θ should be scanned to detect additional peaks that help distinguish phases with similar patterns at $2\theta < 60°$.

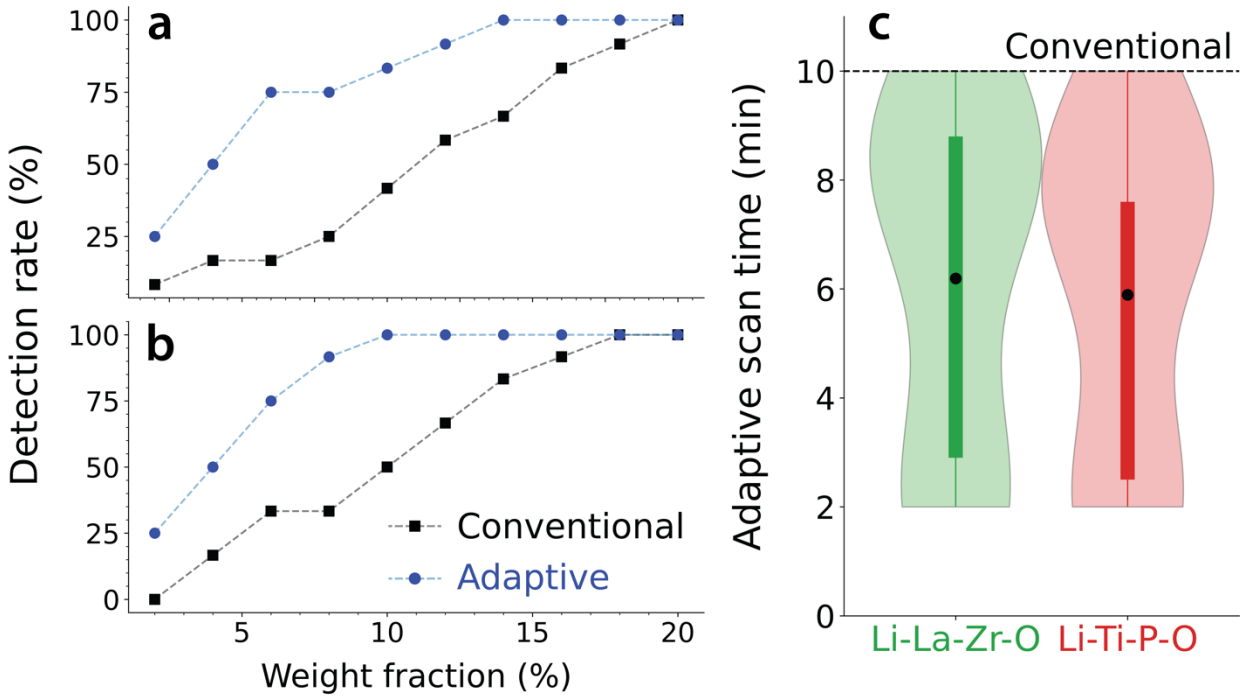

Figure 3-3. The percentage of minority phases correctly identified by XRD-AutoAnalyzer when applied to multi-phase mixtures scanned in a conventional or adaptive manner. These were collected from the (a) Li-La-Zr-O and (b) Li-Ti-P-O chemical spaces. Results are plotted separately for predictions based on conventional and adaptive measurements. (c) The distributions of scan times required by adaptive measurements in each space. For comparison, the conventional scan time (10 min) is represented by the dashed line.

3.2.4 Adaptive XRD for *in situ* characterization

We demonstrate below that the optimized effectiveness by which adaptive XRD collects data can lead to new experimental capabilities. To this end, a solid-state synthesis procedure targeting $Li_7La_3Zr_2O_{12}$ (LLZO) was designed and carried out[82]. During the corresponding synthesis experiments, *in situ* measurements on a Bruker D8 Advance diffractometer were integrated with XRD-AutoAnalyzer to characterize the reaction pathway *via* the identification of precursors, intermediate phases, and final products. Such *in situ* measurements are particularly demanding with respect to the tradeoff between acquisition time and data resolution, as fast reactions and transient intermediate phases can easily be missed when a long acquisition time is used. A precursor powder mixture of $La(OH)_3$, Li_2CO_3, and ZrO_2 was placed in an Anton Paar HTK 1200N oven chamber and heated to 1100 °C at a rate of 20 °C/min, followed by a 1 hour hold at this final temperature. During heating, XRD scans were performed at the onset of a 10 min hold every 100 °C. Three syntheses were separately carried out using distinct measurement techniques (Methods):

1) Fast, non-adaptive scans that sampled 10-80° in 1 min.
2) Slow, non-adaptive scans that sampled 10-80° in 10 min.
3) Adaptive scans with varied $2\theta_{max}$ and measurement times.

Whereas ten patterns can be obtained at each hold when using fast scans (case 1), only one slow scan is performed (case 2). The number of patterns measured with adaptive scans (case 3) varied with temperature, as longer scan times were automatically allocated to samples where phase identification was complicated by a poor signal-to-noise ratio.

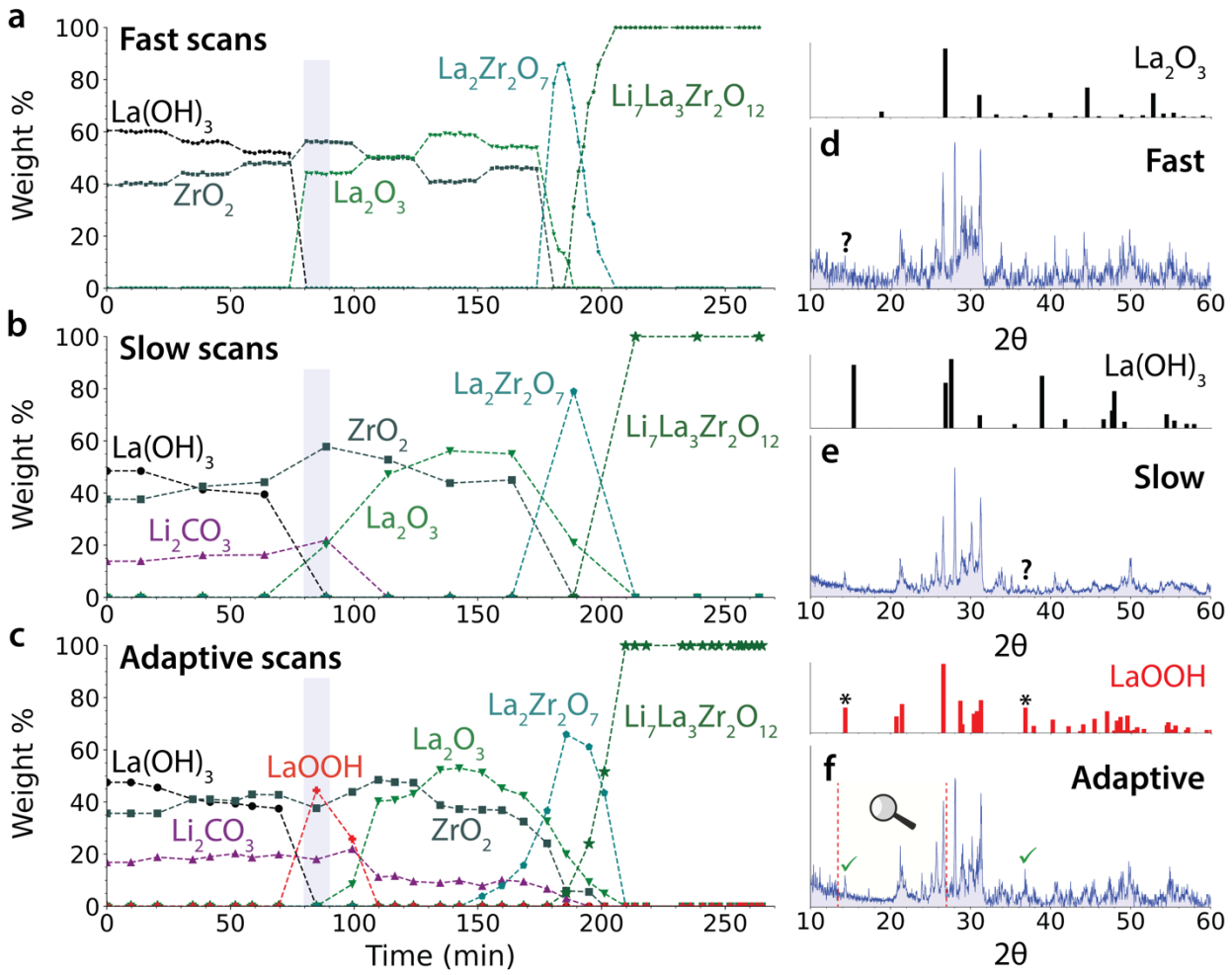

Figure 3-4. The weight fractions for all phases detected during the synthesis of LLZO. Three separate measurement campaigns were performed using (a) fast 1 min scans, (b) slow 10 min scans, and (c) adaptive scans. A short-lived intermediate phase, LaOOH, is detected only with adaptive measurements. Panels d-f illustrate XRD patterns obtained from the blue highlighted regions in panels a-c, which each should contain LaOOH. d, A fast scan misses low-2θ peaks from this phase owing to a poor signal-to-noise ratio. e, A slow scan misses high-2θ peaks as LaOOH transforms before the measurement is complete. f, An adaptive scan successfully detects all peaks from this phase by performing a rapid scan over $[10°, 60°]$, followed by resampling of $[14°, 27°]$ to clarify the smaller peaks.

Figure 3-4 shows the weight fractions for all phases identified by XRD-AutoAnalyzer during the synthesis of LLZO when using different scan techniques. These results show several limitations of conventional XRD. With fast scans, Li_2CO_3 is not detected as its peaks are difficult to resolve from the background noise. The low resolution from fast scanning also precludes the identification of LaOOH, which appears as an intermediate phase between $La(OH)_3$ and La_2O_3[83]. As shown in **Figure 3-4d**, the poor signal-to-noise ratio resulting from a short scan obscures several peaks associated with LaOOH, making it difficult to resolve this phase from others (*e.g.*, La_2O_3). While a longer scan time of 10 min enables the detection of Li_2CO_3 by reducing noise in

the corresponding pattern (**Figure 3-4b**), LaOOH is still missed. Interestingly, the 10 min scans do clarify several low-2θ peaks associated with LaOOH but fail to detect many of its peaks at higher 2θ (**Figure 3-4e**), suggesting that LaOOH transformed before the full range of 2θ was sampled. These findings highlight two competing factors that dictate the accuracy of *in situ* XRD measurements: (1) fast scans yield patterns with low signal-to-noise ratios, complicating the identification of small peaks that blend in with the background noise; (2) slows scans suffer from pattern changes as the measurements are performed, making it difficult to identify short-lived intermediate phases that transform before all their peaks can be detected.

Adaptive XRD overcomes the challenges described in the previous paragraph by achieving an optimal balance between speed and accuracy. As shown in **Figure 3-4c**, Li_2CO_3 is successfully identified with a short scan time of ~3 min as the adaptive algorithm leverages early data to focus high-resolution measurements on a subset of $2\theta = [18°, 32°]$ that contains the major peaks associated with this phase. Note that the algorithm is given no prior information regarding the presence of Li_2CO_3, but it quickly detects some signal above the noise in the relevant area and accordingly requests additional scanning to better resolve that signal. Adaptive XRD also leads to the successful detection of LaOOH, appearing briefly as an intermediate phase at 400 °C. In **Figure 3-4f**, we show how the diffractometer was steered toward $2\theta = [14°, 27°]$ at this temperature, revealing several LaOOH peaks that would otherwise be difficult to resolve from the background noise. Furthermore, because the total measurement time was kept short (~4 min), the full range of $2\theta = [10°, 60°]$ was sampled before LaOOH transformed into La_2O_3. We stress that no human intervention was needed to redirect the diffractometer as the ML algorithm autonomously decides which parts of a pattern are most important for phase identification and accordingly steers measurements toward those regions. In doing so, adaptive XRD enables the identification of short-lived intermediate phases that otherwise would be missed by conventional XRD scans. Knowledge regarding such intermediate phases is often key to understanding and tailoring reaction pathways for inorganic materials synthesis[84].

3.3 Discussion

We believe that the integration of ML-assisted analysis tools can rapidly transform how experimental research is done. In contrast to traditional experimentation, where data is only analyzed after the fact, adaptive methods leverage all available data in real time to make optimal decisions regarding where the experimental measurements should be steered, and as such minimize the time required to obtain all relevant information. These autonomous and adaptive methods require (a) the development of rapid analysis tools that can make predictions, quantify uncertainty, and identify high-value measurement regimes, all on the timescale of the experiment; and (b) the ability to bring this analysis in-line with experiments and control the instruments needed for characterization. Modern ML techniques, while often requiring significant time for training performed off-line, can usually be evaluated within seconds and are therefore ideal decision-making agents to be integrated with experimental hardware.

We demonstrate in this work specifically, that ML-driven adaptive control over XRD measurements enables rapid and autonomous identification of crystalline materials in multi-phase samples, consistently detecting and categorizing phases more quickly and with higher accuracy than conventional XRD scans. By reducing the measurement time while maintaining high precision, adaptive XRD provides an effective method to monitor solid-state reactions *in situ* and identify short-lived intermediate phases using an in-house diffractometer. Although such instrumentation provides reduced intensity relative to a synchrotron light source, adaptive

measurements make efficient use of the available radiation by rationally allocating scan time to resolve peaks with the highest leverage for phase identification. This approach is generalizable and may be extended to alternative diffraction techniques based on neutron or electron scattering. With future developments, we envision that many spectroscopic and microscopic techniques are likely to benefit from ML guidance and interpretation, enabling optimized selection and refinement of critical features in spectra and images. Increased automation of experimental measurements will not only reduce time and labor spent by human researchers, but also give unprecedented access to the characterization of short-lived processes in materials science and chemistry.

3.4 Methods

3.4.1 Automated phase identification

To automatically identify phases from XRD patterns, we used the XRD-AutoAnalyzer algorithm developed in previous work[80] (https://github.com/njszy/XRD-AutoAnalyzer). This approach is based on a CNN with six convolutional layers, six pooling layers, and three fully connected layers. A dropout rate of 60% was used between the fully connected layers. As input to the CNN, each XRD pattern is treated as a one-dimensional vector with 4501 intensities distributed uniformly over 2θ. The output layer contains N neurons, where N is equal to the number of phases in the training set. Here we trained two separate models to analyze data from the Li-La-Zr-O and Li-Ti-P-O chemical spaces, which included 28 and 45 unique phases, respectively. For each phase, 150 XRD patterns with stochastically varied peak positions, widths, and intensities were simulated and used to train the CNN. Training was carried out for 50 epochs. At inference, we divided the trained CNN into an ensemble of 1000 individual models whereby each utilized different connections between its fully connected layers (*i.e.*, with 60% dropout). The result is a probabilistic distribution of predicted phases for a given pattern, where the confidence of each phase is defined as the fraction (%) of models in the ensemble that identify it as the most likely phase. Additional details on the phase identification algorithm are given in our previous work.

3.4.2 Class activation maps

CAMs were originally designed to highlight areas in an image that have the greatest influence on a CNN's output[81]. This can be accomplished by mapping the trained weights of a global average pooling layer, placed after all the convolutional layers in a CNN, onto the final image such that important convolution features have high values in the CAM. Here we use a generalized version known as Grad-CAM[85], which has capabilities similar to the traditional CAM approach but does not require a global average pooling layer in the CNN. Using this technique, we calculated the CAM associated with the classification of an ideal (simulated) XRD pattern for each reference phase in the training sets. All CAMs were normalized between 0 and 100 to ensure consistent comparison between different phases. In cases where XRD-AutoAnalyzer failed to identify a phase with a confidence greater than 50%, the absolute difference between the CAMs of the two most probable phases was calculated and resampling was proposed in areas where the difference exceeds some threshold defined by the user. A threshold of 25% was used during all experimental tests described in the main text, where the % is calculated relative to the maximum value of the CAM (*i.e.,* 100).

3.4.3 Simulated test patterns

A total of 1400 single-phase, 2400 two-phase, and 4200 three-phase patterns were simulated to test the adaptive XRD approach in high-throughput. These patterns were based on 28 and 45 unique crystalline phases from the Li-La-Zr-O and Li-Ti-P-O spaces, respectively. All structures were extracted from the ICSD. Multi-phase patterns were constructed *via* linear combinations of single-phase peaks, where the weight fraction of each individual phase was randomly sampled between 20% and 80%. To mimic data acquired experimentally, all patterns were stochastically augmented with three different artifacts. Peak shifts caused by strain were implemented with up to $\pm 3\%$ changes in the lattice parameters of each phase. Peak intensities were varied by as much as $\pm 50\%$ according to preferred crystallographic orientation (texture) along randomly sampled Miller indices ($[hkl]$ where $0 \leq h, k, l \leq 2$). Different peak widths were sampled using the Scherrer equation based on grain sizes ranging from 5 nm (broad) to 50 nm (narrow). A gaussian shape was assumed for all peaks.

All 8000 simulated patterns were duplicated to form 10 different datasets with varied sampling density and signal-to-noise ratio. The former was set by the number of datapoints contained in each pattern while the former was treated by adding Gaussian noise with a standard deviation reflecting the effective measurement time (see Eqn. 2). The patterns with minimal resolution contained 3250 datapoints spanning 10°-140° ($\Delta 2\theta = 0.04°$), and Gaussian noise was added with a standard deviation of 5% (in terms of the maximum peak intensity). In contrast, the highest resolution patterns contained 6500 datapoints ($\Delta 2\theta = 0.02°$) spanning the same range, in addition to Gaussian noise with a standard deviation of only 2%. To mimic experimental resampling with increased resolution, which would be performed by adaptive XRD, we start from the low-resolution pattern and splice in data from the corresponding high-resolution pattern in regions of 2θ where artificial resampling is performed.

XRD-AutoAnalyzer was used to perform phase identification on the simulated XRD patterns described in the previous paragraph. To quantify the performance of this algorithm when applied autonomously to each dataset, we used the F_1-score:

$$F_1\text{-score} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})} \tag{3}$$

Where TP is the number of true positives (correctly identified phases), FP is the number of false positives (phases incorrectly identified), and FN is the number of false negatives (missed phases). A high F_1-score (close to 1) is desired to successfully identify all phases in a sample without incorrectly identifying phases that are not present.

3.4.4 Two-phase mixture preparation

Mixtures were prepared based on materials in two chemical spaces: Li-La-Zr-O and Li-Ti-P-O. These included Li_2CO_3 (Sigma-Aldrich, 99.9%), LiOH (Sigma-Aldrich, 98%), $La(OH)_3$ (Sigma-Aldrich, 99.9%), ZrO_2 (Sigma-Aldrich, 99.6%), TiO_2 (Alfa Aesar, 99.9%), Li_2TiO_3 (Sigma-Aldrich, 99.9%), and Li_3PO_4 (Sigma-Aldrich, 99.9%). There are 12 possible two-phase majority|minority permutations of the materials in each chemical space (*e.g.*, TiO_2|Li_2CO_3 and Li_2CO_3|TiO_2), where the first phase to appear is the majority phase and the second is the minority phase. For each of these two-phase pairs, 10 mixtures were prepared with iteratively larger amounts of the minority phase. This included weight fractions of 2%, 4%, 6%, 8%, 10%, 12%, 14%, 16%, 18%, and 20% for the minority phase in each permutation. All mixtures were shaker

milled for 10 min with a SPEX 800 mixer, followed by characterization with an Aeris X-ray diffractometer from Panalytical.

3.4.5 *In situ* characterization of $Li_7La_3Zr_2O_{12}$ synthesis

To synthesize $Li_7La_3Zr_2O_{12}$ (LLZO), we used a precursor powder mixture containing Li_2CO_3 (Sigma-Aldrich, 99.9%), $La(OH)_3$ (Sigma-Aldrich, 99.9%), ZrO_2 (Sigma-Aldrich, 99.6%). In addition to the stoichiometric amounts of these precursors needed to make LLZO, 10% excess weight of Li_2CO_3 was included to compensate for suspected volatility at high temperature. These precursors were mixed with ethanol and milled for 10 min using a SPEX 800 mixer, followed by drying at 70 °C in an oven for one hour. The dried sample was loaded into the Anton Paar HTK 1200N oven chamber of a Bruker D8 Advance X-ray diffractometer, which was heated to 1000 °C at a rate of 20 °C/min in air. A hold time of one hour was used at 1000 °C, followed by a natural cool to room temperature. During the heating ramp, a 10 min temperature hold was imposed every 100 °C such that XRD scans could be performed on the sample.

Three different syntheses were carried out, each with a distinct measurement type. First, slow and non-adaptive measurements were employed whereby a single 10 min scan was performed at each 100 °C hold. Second, fast and non-adaptive measurement were applied such that 10 one min scans were performed at each 100 °C hold. Third, adaptive measurements were used with varied scan time and number of scans applied to each 100 °C hold. On average, adaptive scans required ~3 min per pattern. For both types of non-adaptive measurements, the scan range was kept fixed at $2\theta = [10°, 80°]$. In contrast, adaptive scans kept a fixed minimum angle of $10°$, but varied the maximum angle between $60°$ and $140°$ following the workflow described in the main text and illustrated. All patterns were analyzed in an automated fashion using XRD-AutoAnalyzer. The corresponding model was trained on 28 phases in the Li-La-Zr-O chemical space.

CHAPTER 4 Learning to optimize the selection of precursors for solid-state synthesis

The work presented in this chapter is based, often verbatim, on the following publication:
***N. J. Szymanski**, P. Nevatia, C. J. Bartel, Y. Zeng, and G. Ceder, "Autonomous and dynamic precursor selection for solid-state materials synthesis." Nature Communications 2023, 14, 6956.*

Solid-state synthesis plays an important role in the development of new materials and technologies. While *in situ* characterization and ab-initio computations have advanced our understanding of materials synthesis, experiments targeting new compounds often still require many different precursors and conditions to be tested. Here we introduce an algorithm (ARROWS) designed to automate the selection of optimal precursors for solid-state materials synthesis. This algorithm actively learns from experimental outcomes to determine which precursors lead to unfavorable reactions that form highly stable intermediates, preventing the target material's formation. Based on this information, ARROWS proposes new experiments using precursors it predicts to avoid such intermediates, thereby retaining a larger thermodynamic driving force to form the target. We validate this approach on three experimental datasets, containing results from over 200 synthesis procedures. In comparison to black-box optimization, ARROWS identifies effective precursor sets for each target while requiring substantially fewer experimental iterations. These findings highlight the importance of domain knowledge in optimization algorithms for materials synthesis, which are critical for the development of fully autonomous research platforms.

4.1 Introduction

Conventional high temperature synthesis based on solid-state reactions has long been used for the preparation of inorganic materials[86]. This method involves the mixing and subsequent heating of solid powders to facilitate reactions between them. Despite its apparent simplicity, the outcomes of solid-state synthesis experiments are often difficult to predict[78,87]. While DFT calculations can be used to assess thermodynamic stability[88], even materials that are stable can sometimes be difficult to synthesize owing to the formation of inert byproducts that compete with the target and reduce its yield[84,89–91]. Further complicating matters is the prevalence of metastable materials[92] used in countless technologies including photovoltaics[93] and structural alloys[94]. Metastable materials are typically prepared using low-temperature synthesis routes, where kinetic control can be used to avoid the formation of equilibrium phases[95], though recent work has shown that metastable phases can also appear as intermediates during high temperature experiments[96–98]. To optimize the purity of a desired product, whether it be stable or only metastable, requires careful selection of precursors and reaction conditions. This selection process traditionally relies on domain expertise, reference to previously reported procedures for similar targets (if any exist)[99,100], and the use of heuristics such as Tamman's rule[101]. However, there is no clear roadmap to optimize the solid-state synthesis of novel inorganic materials, which can lead to many experimental iterations with no guarantee of success.

A new opportunity exists to accelerate inorganic materials development by leveraging computer-aided optimization to plan solid-state synthesis experiments, learn from their outcomes, and make improved decisions regarding the selection of precursors and conditions that enable the formation of desired phases. Such an approach has found success in organic chemistry, where reactions can often be described by the breaking and formation of individual bonds[102,103]. This enables the use of retrosynthetic methods, which start from the target and work backward through stepwise reactions until a set of available starting materials is reached[104]. As many different reaction paths can lead to a given target, computer-aided optimization techniques based on Monte Carlo tree search and reinforcement learning have been successfully used to rapidly screen for promising synthesis routes[105–107]. In contrast, inorganic materials synthesis has yet to benefit from the widespread use of algorithms that can optimize experimental procedures. Their development is hindered by the difficulty of modeling solid-state reactions, where the corresponding phase transformations involve concerted displacements and interactions among many species over extended distances[78]. Some progress has been made in simplifying the analysis of solid-state reaction pathways by decomposing them into step-by-step transformations that take place between two phases at a time, hereafter referred to as pairwise reactions[84,98]. However, it remains difficult to predict the temperature at which a given pairwise reaction will occur, as well as what phase(s) will form as a result of that reaction.

To determine which reaction outcomes are most plausible for a given set of precursors and conditions, much of the existing work on computer-aided planning for solid-state synthesis has relied on the analysis of thermochemical data based on DFT calculations[68,108]. For example, McDermott et al. introduced a graph-based approach that ranks various reaction pathways by a cost function designed to account for changes in the Gibbs free energy of reaction along each path[109]. A related approach developed by Aykol et al. parameterizes reactions by two axes – one that approximates the nucleation barrier of the targeted phase and another that accounts for its competition with possible byproducts – from which optimal reactants can be identified along the Pareto front[110]. Alternatively, ML models can be trained on synthesis data from the literature and applied to suggest effective precursors and conditions for a given target by considering its similarity with previously reported materials[99,100]. While these methods have been successfully applied in some cases, their use remains limited as they only provide a fixed ranking of synthesis routes for a given material, which is not readily updated should the initial experiments fail.

In the place of fixed ranking schemes, active learning algorithms have also been used for the optimization of synthesis procedures[53,111]. These algorithms can adapt from failed experiments and decide which parameters should be tested in later iterations. Bayesian optimization and genetic algorithms have found success when coupled with synthesis techniques based on flow chemistry[112] and thin film deposition[21]. However, these black-box approaches are often restricted to handling continuous variables such as temperature and time, while categorical variables are more difficult to optimize. For example, choosing which precursors to use for the synthesis of a novel material is particularly challenging as it involves discrete selections from a vast range of chemical compositions and structures, instead of simply fine-tuning parameters on a continuous scale. Recent work has made progress on this front by combining parallel synthesis experiments with tensor decomposition analysis, which can be used to predict the most effective starting materials and processing conditions from just a subset of their possible combinations[113].

In this work, we build upon existing methods to optimize solid-state synthesis procedures by incorporating physical domain knowledge based on thermodynamics and pairwise reaction analysis. This is accomplished using Autonomous Reaction Route Optimization with Solid-State

Synthesis (ARROWS), an algorithm designed to guide the selection of precursors for the targeted synthesis of inorganic materials. Given a desired structure and composition, ARROWS uses existing thermochemical data in the Materials Project to form an initial ranking of precursor sets based on their DFT-calculated reaction energies[2,114]. Highly ranked precursors are suggested for experimental validation throughout a range of temperatures, which are iteratively probed and analyzed using ML algorithms to identify the intermediates that form along each precursor set's reaction pathway. When such experiments fail to produce the desired phase, ARROWS learns from their outcomes and updates its ranking to avoid pairwise reactions that consume much of the available free energy and therefore inhibit formation of the targeted phase. To benchmark the performance of ARROWS, we conducted 188 synthesis experiments targeting $YBa_2Cu_3O_{6.5}$, forming a comprehensive reaction dataset that critically includes both positive and negative results. Our approach is shown to identify all the effective synthesis routes from this dataset while requiring fewer experimental iterations than Bayesian optimization or genetic algorithms. We further demonstrate that ARROWS can be applied in-line with experiments to guide the selection of precursors for two metastable targets, $Na_2Te_3Mo_3O_{16}$ and $LiTiOPO_4$, each of which were successfully prepared with high purity.

4.2 Results

4.2.1 Design of ARROWS

The logical flow of ARROWS is summarized in **Figure 4-1** and detailed in the Methods section. Given a target material specified by the user, ARROWS forms a list of precursor sets that can be stoichiometrically balanced to yield the target's composition. In the absence of previous experimental data, these precursor sets are initially ranked by their calculated thermodynamic driving force (ΔG) to form the target (**Figure 4-1a**). While many factors influence the rates at which solid-state reactions proceed[115], those with the largest (most negative) ΔG tend to occur most rapidly[98,99,116]. However, such reactions may also be slowed by the formation of intermediates that consume much of the initial driving force[90]. To address this, ARROWS proposes that each precursor set be tested at several temperatures, thereby providing snapshots of the corresponding reaction pathway (**Figure 4-1b**). The intermediates formed at each step in the reaction pathway are identified using XRD with machine-learned analysis[80]. ARROWS then determines which pairwise reactions led to the formation of each observed intermediate phase (**Figure 4-1c**), and it leverages this information to predict the intermediates that will form in precursor sets that have not yet been tested (**Figure 4-1d**). In subsequent experiments, ARROWS prioritizes sets of precursors that are expected to maintain a large driving force at the target-forming step ($\Delta G'$), *i.e.*, even after intermediates have formed (**Figure 4-1e**). This process is repeated until the target is successfully obtained with sufficiently high yield, as specified by the user, or until all the available precursor sets have been exhausted.

To validate the effectiveness of ARROWS, new experimental synthesis data is needed. Existing results from the literature tend to be heavily biased toward positive results, which precludes the development of models that can learn from failed experiments[52]. We therefore built a solid-state synthesis dataset for $YBa_2Cu_3O_{6.5}$ (YBCO) by testing 47 different combinations of commonly available precursors in the Y-Ba-Cu-O chemical space, which were mixed and heated at four synthesis temperatures ranging from 600-900 °C. Importantly, this dataset includes both positive and negative outcomes, *i.e.*, reactions that do and do not yield sufficiently pure YBCO. As such, it can be used as a benchmark on which to test ARROWS and compare its efficacy with alternative optimization algorithms. Two additional chemical spaces are also considered, where

we use ARROWS to actively guide the experiments. The first set of experiments targeted $Na_2Te_3Mo_3O_{16}$ (NTMO), which is metastable with respect to decomposition into $Na_2Mo_2O_7$, $MoTe_2O_7$, and TeO_2 according to DFT calculations[117]. The second set of experiments targeted a triclinic polymorph of $LiTiOPO_4$ (*t*-LTOPO), which has a tendency to undergo a phase transition into a lower-energy orthorhombic structure (*o*-LTOPO) with the same composition[118]. The features of each space tested are summarized in **Table 4-1**. Further details regarding the corresponding experiments are provided in the Methods.

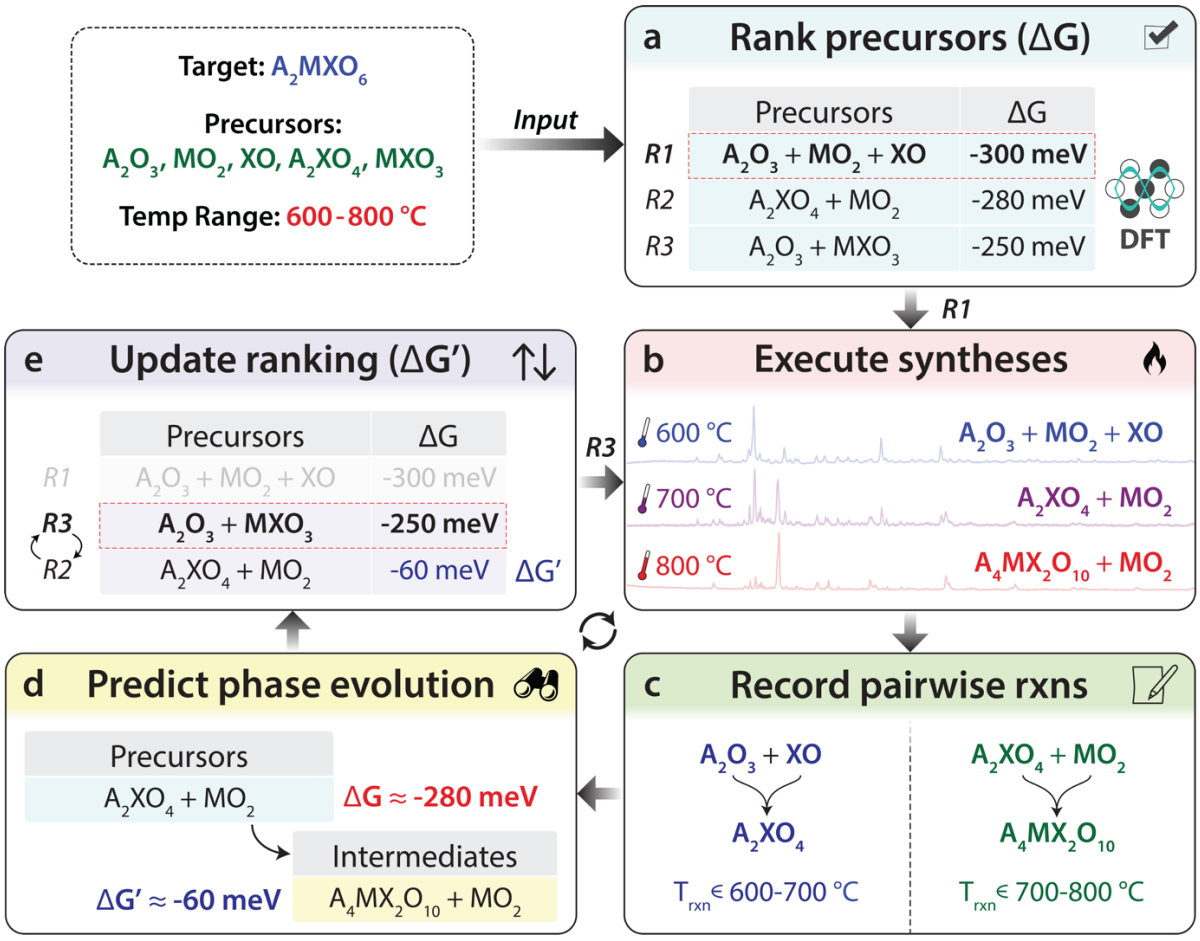

Figure 4-1. A schematic illustrating how ARROWS guides precursor selection. (a) Reactions based on different precursor sets (R1, R2, and R3) are initially ranked by their driving force (ΔG) to form the target, which is obtained from DFT calculations. (b) Experiments are performed at iteratively higher temperatures to identify reaction intermediates. The chemical formulae listed in this panel represent the phases identified from XRD measurements. (c) Pairwise reaction temperatures (T_{rxn}) and products are gleaned from the experimental data. (d) Using the identified pairwise reactions, intermediates are predicted for other precursor sets and their remaining driving forces (ΔG′) are updated accordingly. (e) The precursor ranking is updated based on the newly calculated ΔG′. All chemical formulae shown are placeholders for arbitrary compounds, and in general there is no restriction on the compositions where ARROWS is applicable.

Table 4-1. Information regarding three search spaces on which ARROWS was tested. N_{sets} and N_{exps} represent the number of precursor sets and experiments, respectively.

Target	N_{sets}	Temperatures (°C)	N_{exp}
$YBa_2Cu_3O_{6+x}$	47	600, 700, 800, 900	188
$Na_2Te_3Mo_3O_{16}$	23	300, 400	46
t-$LiTiOPO_4$	30	400, 500, 600, 700	120

4.2.2 YBCO

Before discussing the optimization of YBCO synthesis using ARROWS, we first summarize the outcomes from all 188 experiments to give context regarding the difficulty of obtaining high-purity YBCO while using a hold time of 4 h. Such a short hold time was used specifically to make the optimization task more challenging, as longer heating durations with intermittent regrinding are typically required to form highly pure YBCO samples[119]. Indeed, only 10 of the 188 experiments performed in the current work led to the formation of pure YBCO without any prominent impurity phases that could be detected by XRD-AutoAnalyzer[80]. Another 83 experiments gave partial yield of YBCO, in addition to several unwanted byproducts. **Figure 4-2a** shows the distribution of YBCO yield (wt. %) at each synthesis temperature sampled in this work. Generally, the use of higher temperature leads to increased yield of YBCO, likely due to enhanced reaction kinetics. Precursor selection also has a marked effect on the target's yield. **Figure 4-2b** shows the success rate of each precursor, which we define as the percentage of sets where that compound was included and resulted in the formation of YBCO without any detectable impurities. This plot suggests that the less commonly used binary precursors tend to outperform their standard counterparts. For example, BaO and BaO_2 have moderately high success rates of 46% and 22%, respectively, whereas sets with $BaCO_3$ always produce impure samples (0% success rate). Precursor sets including $Y_2Cu_2O_5$ and $Ba_2Cu_3O_6$ also have comparably high success rates of 33% and 31%, respectively. We will later show that these ternary phases enable the direct formation of YBCO while circumventing inert byproducts such as Y_2BaCuO_5.

Figure 4-2c displays a pie chart containing the four most common impurity phases that coexist with YBCO, or prevent its formation entirely, at 900 °C. Each slice in the pie chart represents the fraction of experiments where the specified impurity phase appears. Most of the impure samples (28/44) contain $BaCuO_2$ or Y_2BaCuO_5, which are known to be relatively inert during the synthesis of YBCO, requiring intermittent grinding to improve the sample's purity[119,120]. CuO is another frequent impurity, though it only ever appears with at least one other byproduct that is Cu deficient. When such phases do not form, CuO contributes to the formation of YBCO, as evidenced by its success rate of 20%. The fourth most common impurity is $BaCO_3$, which is likely slow to react owing to its high decomposition temperature in air (1000 °C)[121,122]. We note that such information could in principle be leveraged when designing the search space, *e.g.*, by removing $BaCO_3$ from the list of precursors since the proposed temperature range lies below its known decomposition temperature. Indeed, doing so reduces the number of experiments required to identify all optimal synthesis routes from 87 to 70.

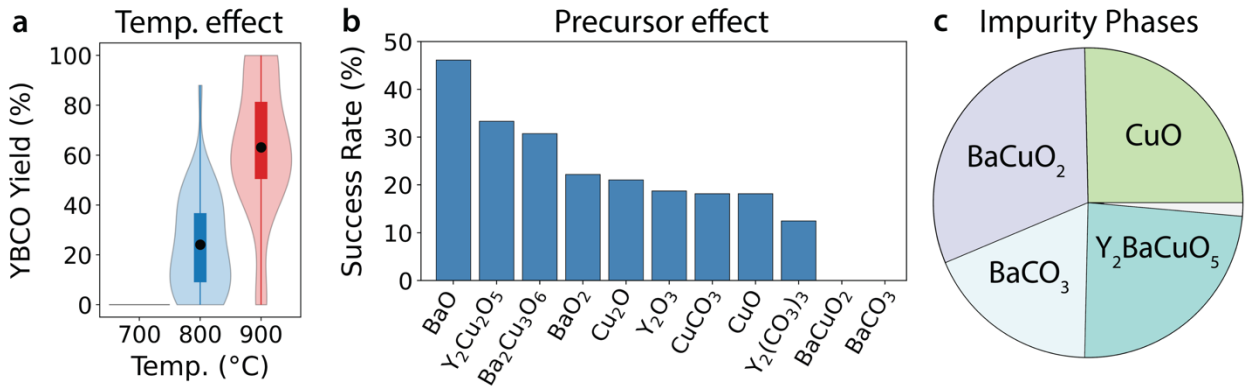

Figure 4-2. A summary of outcomes from the syntheses targeting YBCO. (a) Distributions of YBCO yield (wt. %) at different synthesis temperatures represented using violin and box plots, where each box extends from the lower to upper quartiles. (b) The success rate of each precursor, defined as the percentage of sets where that compound is included and forms YBCO without any impurities. (c) Common impurity phases that prevent YBCO formation are shown by a pie chart, where each slice represents the relative number of occurrences for each compound at 900 °C. The small gray sliver includes two less commonly observed impurities, $YBaCu_3O_7$ and $YBa_4Cu_3O_9$.

To determine whether ARROWS can effectively distinguish between successful and failed synthesis routes, we assessed how many iterations are required to identify all 10 optimal experiments that result in the formation of YBCO without any detectable impurities. While in practice it would be sufficient to identify just one optimal synthesis procedure for a given target, tasking the algorithm with identifying all optimal procedures for YBCO allows us to showcase its ability to learn over many experimental iterations. It also reduces the likelihood that ARROWS discovers an optimal synthesis route by chance, thereby increasing our confidence in the performance of the algorithm.

As a baseline with which to compare the performance of ARROWS on the YBCO dataset, we applied D-optimal design with progressively larger sets of proposed experiments. This approach aims to select the experiments whose outcomes will be maximally informative[123] to a model that maps the input variables (precursors and temperature) onto the output (YBCO yield). Here we assume a linear relationship between the two. We also applied two active learning algorithms, Bayesian optimization (BO) and a genetic algorithm (GA), to the same task by using a one-hot representation of each precursor. These algorithms are known to perform well on numerical inputs such as temperature[124,125]; however, their effectiveness with respect to categorical inputs is less well proven. To specifically probe the latter case, we constrained BO and GA to optimize the selection of precursors while sampling all temperatures for each precursor set. Both black-box algorithms have stochastic elements and were therefore applied to the YBCO dataset 100 times, each with a random starting seed, and their results were averaged. Because ARROWS and D-optimal design are both deterministic algorithms, only a single run was performed to validate each on the YBCO dataset.

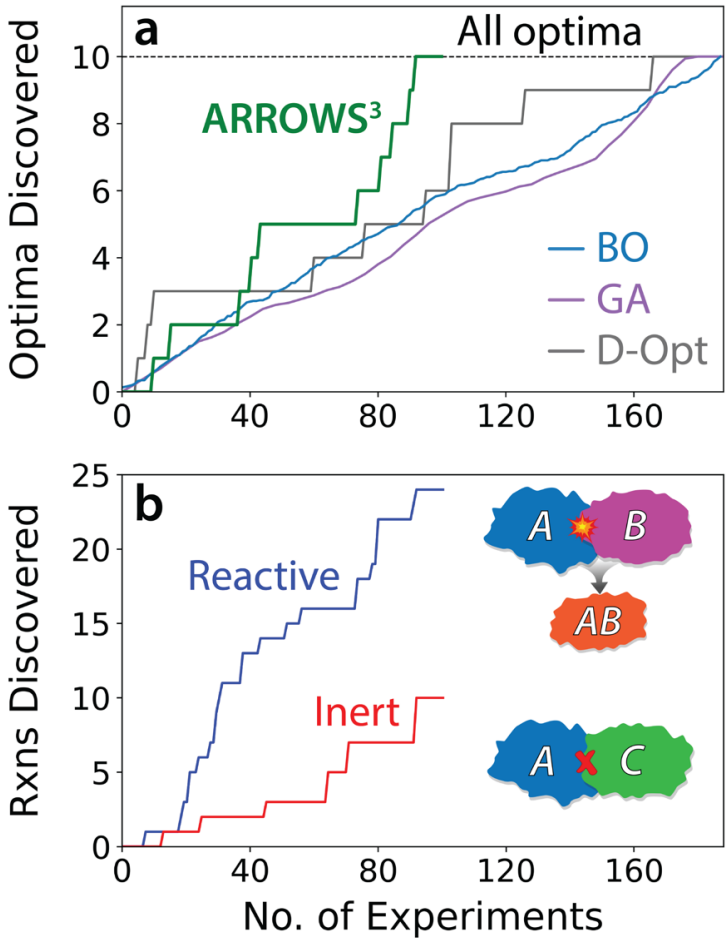

Figure 4-3. Optimization results on the YBCO synthesis dataset. (a) Number of optimal synthesis routes identified as a function of the experimental iterations required by ARROWS, Bayesian Optimization (BO), a Genetic Algorithm (GA), and D-Optimal design (D-Opt). The dashed line represents the total number of optimal synthesis routes in the dataset. (b) Pairwise reactions discovered by ARROWS with respect to the number of experiments queried.

Figure 4-3a shows the number of optimal synthesis routes (those yielding pure YBCO) discovered with respect to the number of experiments queried by each algorithm. ARROWS successfully identified all 10 optimal routes from 87 experiments, which account for just 46% of the entire design space (spanning 188 experiments). D-optimal design, on the other hand, required 165 experiments to accomplish the same task. Though, it is worth noting that D-optimal design was quick to identify three optimal synthesis routes in the first 12 experiments. ARROWS, although slower to identify optimal routes in the early stages of optimization, eventually surpassed D-optimal design once it gathered sufficient information regarding the reactivity of various phases in the Y-Ba-Cu-O chemical space.

Both active learning algorithms performed poorly on the YBCO dataset, with BO and GA requiring on average 164 and 167 experiments to identify all ten optimal synthesis routes. We suspect the ineffectiveness of these algorithms is related to their use of one-hot representations for the precursors, which treat each compound independently and contain no physical information regarding their composition or structure. In contrast, ARROWS encodes compositional and

thermodynamic information in its optimization through its ranking by ΔG. It also learns from failed experiments to avoid pairwise reactions that form inert byproducts such as $BaCuO_2$ and Y_2BaCuO_5, instead prioritizing sets of precursors expected to retain a strong driving force ($\Delta G'$) to form YBCO.

Figure 4-3b displays the number of pairwise reactions learned by ARROWS with respect to the number of experiments that were queried. This plot includes pairs of phases that react within the temperature range considered (≤ 900 °C), denoted reactive pairs, as well as the phases that do not react within that range, denoted inert pairs. From 87 experiments, ARROWS gained information regarding 34 pairwise interactions, including 24 reactive and 10 inert pairs. We find that the identification of new successful synthesis routes is often preceded by the discovery of new pairwise reactions. For example, ARROWS learned from experiments 30-34 that BaO reacts with CuO to form $BaCuO_2$ at 800 °C, which subsequently reacts with Y_2O_3 at 900 °C to form Y_2BaCuO_5. Because these pairwise reactions consume much of the driving force that remains to form YBCO, the algorithm decides to prioritize sets of precursors that do not contain such reactive pairs (BaO|CuO or $BaCuO_2$|Y_2O_3). This decision leads to the successful discovery of three new synthesis routes that produce YBCO without any detectable impurities, as shown by the steep rise of the green curve in **Figure 4-3a** between experiments 38-43. While previous work has shown that $BaCuO_2$ can effectively contribute to YBCO formation when it melts in combination with CuO[84], there was no evidence of melting in our samples owing to the use of low synthesis temperatures (≤ 900 °C) that ensured all products could be easily extracted.

In addition to learning which pairwise reactions should be avoided, ARROWS also learns which reactions are beneficial to achieve high target yield. During the optimization of YBCO synthesis, it learned from experiments 72-80 that BaO_2 reacts with CuO to 700 °C to form $Ba_2Cu_3O_6$, which upon further heating to 900 °C reacts with Y_2O_3 to form YBCO. Accordingly, subsequent experimental iterations are chosen based on precursor sets that either include $Ba_2Cu_3O_6$ or are expected to form it as an intermediate phase. As shown in **Figure 4-3a**, this leads to the rapid identification of all remaining experiments that successfully form YBCO shortly after the 80th experimental iteration.

To showcase the pairwise reactions learned by ARROWS, we present in **Figure 4-4** a heatmap where each square represents a pair of phases. If any information was learned regarding the reactivity of that pair, the square is colored a light shade of blue according to the temperature at which a reaction proceeds. If a pair was instead found to be inert at all temperatures ≤ 900 °C, a dark shade of blue is used. We also denote reactions that produce YBCO (yellow star) or its competing phases, $BaCuO_2$ (orange circle) and Y_2BaCuO_5 (red cross). This heatmap reveals that YBCO forms at 900 °C when $Ba_2Cu_3O_6$ reacts with Y_2O_3 or $Y_2(CO_3)_3$. It is separately observed that $Ba_2Cu_3O_6$ reacts with $Y_2Cu_2O_5$ when both are present at 800 °C, resulting in a mixture of YBCO and CuO. The direct formation of YBCO from $Ba_2Cu_3O_6$ and $Y_2Cu_2O_5$ provides an explanation as to why both phases have high success rates when used as precursors (**Figure 4-2b**). In contrast, the 0% success rates associated with $BaCO_3$ and $BaCuO_2$ can be traced to the limited reactivity of each phase with many of the precursors tested here. This is illustrated in **Figure 4-4** by the dark blue shading that signifies inert reactions pairs in the rows corresponding to $BaCO_3$ and $BaCuO_2$. Even when $BaCO_3$ does react, it often produces $BaCuO_2$ or Y_2BaCuO_5, which are both common impurity phases that preclude the formation of YBCO. The presence of Y_2BaCuO_5 is particularly detrimental to the synthesis of YBCO as it does not react with any precursor in the allotted hold time of 4 h, which ARROWS learns over the course of the 87 experiments we performed.

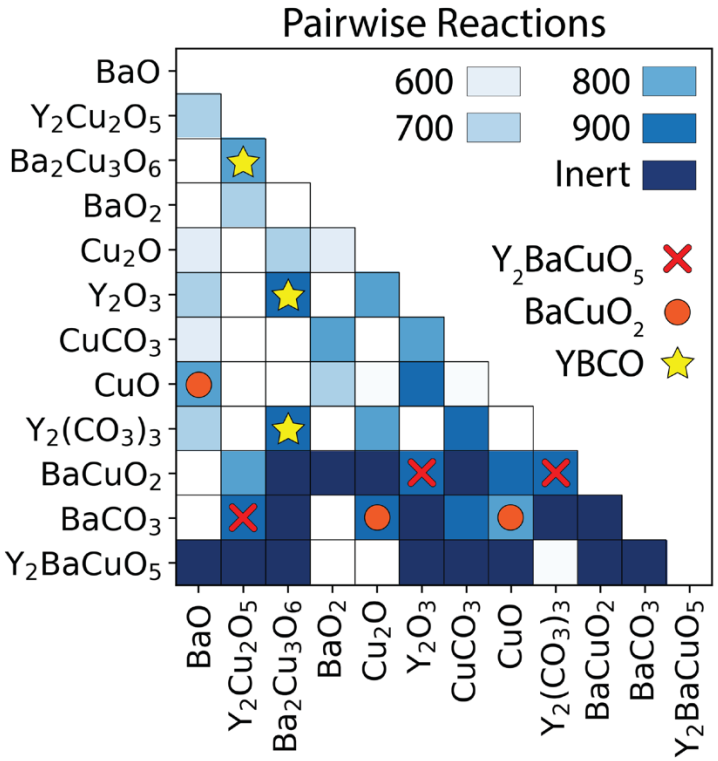

Figure 4-4. A heatmap representing the pairwise reactions learned by ARROWS in the YBCO chemical space. The color of each square represents the temperature (°C) at which a reaction is observed. Inert pairs correspond to phases that do not react within the temperature range considered. White squares (unshaded) represent pairs of phases whose reactivity was not learned by ARROWS. Yellow stars denote pairs that react to produce $YBa_2Cu_3O_{6.5}$ (YBCO). Orange circles and red crosses denote pairs that form impurities, Y_2BaCuO_5 and $BaCuO_2$, respectively.

There also exist some pairs of compounds whose reactivity was not learned by ARROWS during its optimization of YBCO synthesis. These 23 pairs are denoted by the white (unshaded) squares in **Figure 4-4**. We observe two factors that prevent ARROWS from learning pairwise reaction information. First, when two phases (*e.g.*, A|B) react in a three-phase set (A, B, and C), the algorithm is unable to learn how the remaining phase (C) interacts with the already reacted compounds (A and B). Separate experiments based on the individual pairs (A|C and B|C) would be required to determine their reactivity. Second, when multiple pairwise reactions take place within the specified temperature interval (ΔT = 100 °C), the algorithm cannot determine the precise reaction sequence between the lower and upper temperatures (*e.g.*, between 600 and 700 °C). In principle, the second limitation can be overcome by using a smaller temperature interval; however, doing so would also require more experiments.

The results presented in **Figure 4-3** and **Figure 4-4** were obtained by querying experiments in a serial (one-by-one) fashion. This allowed ARROWS to continually learn from each experimental outcome and update its ranking of precursor sets accordingly. However, traditional experiments are often parallelized. For example, multiple sets of precursors with a shared synthesis temperature may be tested simultaneously by loading them into one furnace[113]. Such an approach is also compatible with ARROWS, for which a batch size can be specified to control how many experiments are suggested at each iteration. The use of a larger batch size reduces the number of

iterations (*i.e.*, batches) required to identify all the optimal synthesis routes for YBCO. This also leads to shorter hold times required in the furnace. However, because a larger batch size limits the opportunities where ARROWS can learn and update its ranking, it also leads to a larger number of individual samples that must be queried to identify the optimal routes. The efficiency with which samples are queried becomes particularly affected at later stages in the experiments, where the algorithm has sufficient knowledge of the chemical space to make frequent updates to its ranking of different precursor sets. Hence, there exists a tradeoff between the number of batches and individual samples required to complete the optimization process, and the batch size acts as a hyperparameter to adjust this tradeoff depending on the user's objectives and experimental setup.

4.2.3 NTMO

ARROWS was next tasked with optimizing the yield $Na_2Te_3Mo_3O_{16}$ (NTMO) by choosing from 23 different precursor sets and two synthesis temperatures (300 and 400 °C), which were kept low to avoid melting of the samples[126]. In the top panel of **Figure 4-5a**, we show the weight fraction of NTMO obtained at 400 °C for each precursor set that was tested. The solid dots represent experimentally observed weight fractions, whereas the hollow dots represent predictions made based on the intermediates formed at 300 °C. As detailed in the Methods section, a precursor set occasionally produces identical intermediates phases to a previously explored set. In this case, higher temperatures do not require sampling since their outcomes can already be predicted based on previous synthesis outcomes.

None of the four initial precursor sets, which were selected based on their DFT-calculated reaction energies (ΔG), produced any detectable amount of the target. Their failures can be attributed to the formation of an intermediate phase, $Na_2Mo_2O_7$, that consumes much of the available free energy and precludes the formation of NTMO. This is confirmed by the thermodynamic unfavorable (positive) driving force associated with NTMO formation based on the hypothetical reaction between $Na_2Mo_2O_7$ and two commonly used precursors, TeO_2 and MoO_3:

$$Na_2Mo_2O_7 + 3\ TeO_2 + MoO_3 \rightarrow Na_2Te_3Mo_3O_{16}\ (\Delta G' = +13\ \text{meV/atom})$$

To further illustrate the limiting effect that $Na_2Mo_2O_7$ has on the formation of NTMO, we plot in the bottom panel of **Figure 4-5a** the weight fraction of $Na_2Mo_2O_7$ obtained at 300 °C for each precursor set, revealing a tradeoff between the yield of this phase and that of the target at 400 °C.

From the six initial experiments targeting NTMO, ARROWS acquired knowledge regarding four different pairwise reactions (involving MoO_2 and various Na precursors) that led to the formation of $Na_2Mo_2O_7$ at 300 °C. To maintain a strong thermodynamic driving force to form the target, the algorithm selected all remaining experiments based on precursor sets expected to avoid pairwise reactions that formed $Na_2Mo_2O_7$ and therefore reduced $\Delta G'$. This change in priority from ΔG (based on the precursors) to $\Delta G'$ (based on the predicted intermediates) is highlighted by the vertical dashed line in **Figure 4-5a**. The updated prioritization based on $\Delta G'$ led to a clear increase in the yield of NTMO, as all experiments after the sixth iteration gave $\geq$ 30% yield of NTMO. This improvement is largely attributed to reduced $Na_2Mo_2O_7$ formation when more stable Na precursors such as Na_2CO_3 or Na_2TeO_3 are used.

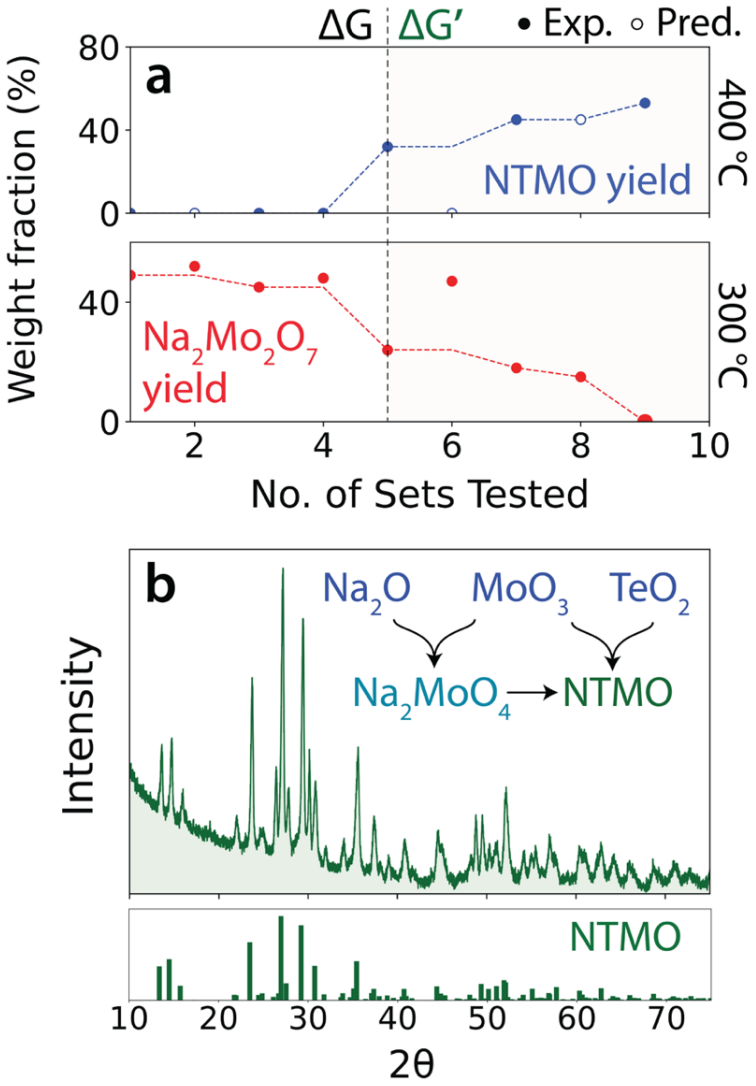

Figure 4-5. Optimization of $Na_2Te_3Mo_3O_{16}$ (NTMO) synthesis using ARROWS. (a) The top panel shows the weight fraction of NTMO obtained from each precursor set when tested at 400 °C. The bottom panel displays the weight fraction of a competing phase, $Na_2Mo_2O_7$, obtained at 300 °C. Solid (hollow) dots represent experimental (predicted) values. (b) XRD pattern measured from the product of the optimized precursor set, Na_2O + TeO_2 + MoO_3 after an 8 h hold at 400 °C. For comparison, a reference pattern is shown for NTMO (ICSD #171758).

ARROWS further discovered from the outcome of the 16th experiment that it is even more effective to use precursors (Na_2O, MoO_3, and TeO_2) that avoid $Na_2Mo_2O_7$ entirely by instead forming Na_2MoO_4. This was the only precursor set for which $Na_2Mo_2O_7$ was not detected at any temperature, and as a result, it successfully produced a sample containing 62% NTMO by weight. It did so by forming Na_2MoO_4, which retains a favorable (negative) driving force to react with the remaining precursors and form the target:

$$Na_2MoO_4 + 3\ TeO_2 + 2\ MoO_3 \rightarrow Na_2Te_3Mo_3O_{16}\ (\Delta G' = -8\ meV/atom)$$

Given that the updated reaction energy is relatively small, we suspect that longer hold times could be used to improve the purity of the synthesis product. To confirm this, we prepared a new sample

containing the same precursors (Na_2O, MoO_3, and TeO_2) and held them at the optimized synthesis temperature 400 °C for a longer hold time of 8 h. The XRD pattern of the resulting product is shown in **Figure 4-5b**, revealing that the use of a longer hold time led to substantially improved purity. The sample contained 94% NTMO by weight, in addition to a 6% TeO_2 impurity. For comparison, we carried out an identical synthesis procedure using a precursor mixture where MoO_3 was replaced with MoO_2, for which the resulting product did not contain any detectable amount of NTMO. This contrasting result highlights the importance of precursor selection and its effect on the reaction pathways that proceed during synthesis. By replacing a single precursor and thus altering which intermediate phase forms first ($Na_2Mo_2O_7$ or Na_2MoO_4), the target yield can vary from 0% to > 90%.

4.2.4 LTOPO

As a final demonstration, ARROWS was used to direct a series of experiments targeting the triclinic polymorph of $LiTiOPO_4$ (*t*-LTOPO) based on a search space consisting of 30 different precursor sets and two synthesis temperatures (400, 500, 600, 700 °C). To achieve this target, the algorithm must learn to avoid the formation of a lower-energy polymorph that exists at the same composition but adopts an orthorhombic structure (*o*-LTOPO)[118]. In the top panel of **Figure 4-6a**, we plot the weight fraction obtained for each polymorph with respect to the number of precursor sets that were sampled by ARROWS during its optimization of the synthesis process. These weight fractions are taken from experimental outcomes at 700 °C, which is the only temperature where either polymorph of LTOPO formed. The solid dots in **Figure 4-6a** represent observed weight fractions, whereas the hollow dots represent predictions made based on the intermediates formed at 400 °C. A total of eight precursor sets were tested before identifying an optimal synthesis route for *t*-LTOPO, though many of these sets produced identical intermediates and therefore did not require sampling of temperatures > 400 °C.

A key distinguishing feature between the precursor sets tested by ARROWS is the amount of $LiTi_2(PO_4)_3$ formed as an intermediate in each case. The weight fraction of this phase contained in each sample made at 400 °C is plotted in the bottom panel of **Figure 4-6a**. Precursor sets 1-2 both formed > 40% wt. of $LiTi_2(PO_4)_3$, consuming much of the driving force left to form the target. This effect is illustrated by the chemical reactions below, representing the phases contained in precursor set 1 before and after annealing at 400 °C:

$$\text{Before: } LiOH + TiO_2 + (NH_4)_2HPO_4 \rightarrow LiTiOPO_4 + 2\ NH_3 + 2\ H_2O\ (\Delta G = -58\ \text{meV/atom})$$

$$\text{After: } Li_4P_2O_7 + 3\ TiO_2 + LiTi_2(PO_4)_3 \rightarrow 5\ LiTiOPO_4\ (\Delta G' = -6\ \text{meV/atom})$$

As outlined in recent work[127], preferential nucleation of *o*-LTOPO tends to occur when preceded by reactions with small changes in the Gibbs free energy. This is confirmed by the synthesis outcome of precursor set 1 annealed at 700 °C, which produces a sample containing 35% *o*-LTOPO and only 17% *t*-LTOPO, in addition to leftover $LiTi_2(PO_4)_3$ and TiO_2 impurities.

To avoid the reactions that form $LiTi_2(PO_4)_3$ and thereby retain larger $\Delta G'$ to form the target, ARROWS suggests precursors where such reactions have not yet been observed. As shown by the data to the right of the dividing line in **Figure 4-6a**, which separates experiments selected using ΔG from those selected using $\Delta G'$, this decision successfully reduced $LiTi_2(PO_4)_3$ formation and led to increased yield of *t*-LTOPO. The plateau in the amount of each phase formed with precursor sets 3-7 is associated with the use of less reactive Li sources – including Li_2CO_3, Li_2TiO_3, and $Li_4Ti_5O_{12}$ – which tend to persist until higher temperature and reduce the amount of $LiTi_2(PO_4)_3$ that forms as an intermediate. While this led to increased yield of the target, *o*-LTOPO

still accompanied its formation at 700 °C. In contrast, the eighth precursor set proposed by ARROWS (Li_2O, TiO_2, and P_2O_5) resulted in 54% target yield and no detectable amount of *o*-LTOPO. Notably, this was also the only precursor set that did not form any $LiTi_2(PO_4)_3$ at 400 °C. It instead formed a set of intermediates that maintained a stronger driving force to form the target as shown by the chemical reaction below:

$$Li_3PO_4 + 2\ TiO_2 + TiP_2O_7 \rightarrow 3\ LiTiOPO_4\ (\Delta G' = -24\ meV/atom)$$

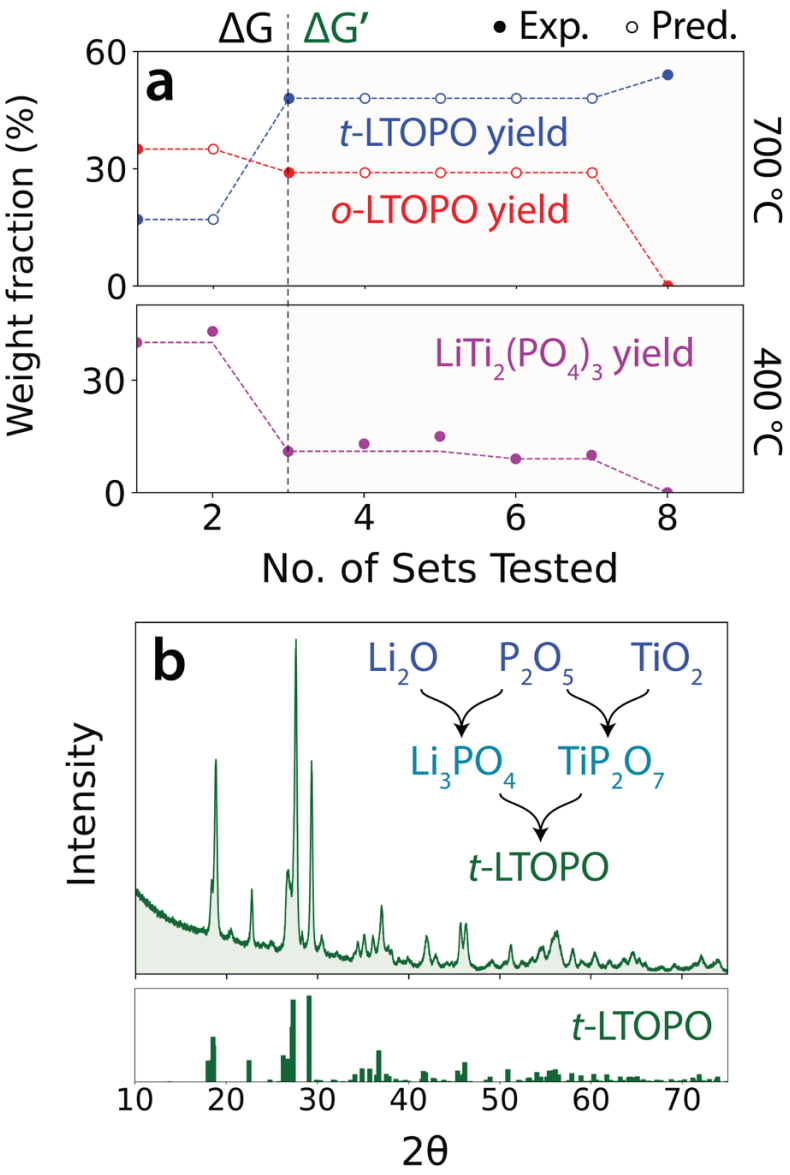

Figure 4-6. Optimization of triclinic $LiTiOPO_4$ (*t*-LTOPO) synthesis using ARROWS. (a) The top panel shows the weight fractions obtained for the target as well as its competing polymorph (*o*-LTOPO) based on each precursor set that was tested at 700 °C. The bottom panel displays the weight fraction of a common impurity phase, $LiTi_2(PO_4)_3$, obtained at 400 °C. Solid (hollow) dots represent experimental (predicted) values. (b) XRD pattern measured from the synthesis product of the optimized precursor set, $Li_2O + TiO_2 + P_2O_5$, which was ball milled and subsequently heated to 700 °C for 4 h. For comparison, a reference pattern for *t*-LTOPO (ICSD #39761) is also shown.

Because ARROWS identified a synthesis route that gave 54% yield for *t*-LTOPO, exceeding our pre-defined objective of 50%, the optimization process was complete. Nevertheless, to verify that the target could made with higher purity using these optimized precursors, we separately performed a synthesis procedure where Li_2O, TiO_2, and P_2O_5 were ball milled prior to heating the mixture at 700 °C for 4 h. The XRD pattern of the resulting product is shown in **Figure 4-6b**, revealing the formation of *t*-LTOPO without any detectable impurity phases. For comparison, the same procedure was also applied to a precursor mixture of LiOH, TiO_2, and P_2O_5. The resulting synthesis product contained $LiTi_2(PO_4)_3$ and *o*-LTOPO impurities, which limited the yield of *t*-LTOPO to 46% when using these non-optimized precursors.

Although *t*-LTOPO was successfully optimized, we advise careful application of ARROWS for synthesizing metastable polymorphs. Our algorithm worked effectively with LTOPO, as its desired (metastable) polymorph is favored at large reaction energies, primarily due to its stable surface energy at small particle size[127]. This makes it well-suited for ARROWS, which learns to prioritize synthesis pathways with large reaction energy at the target-forming step. However, if the stable polymorph instead had low surface energy, its formation would be enhanced by the recommended precursor sets. Therefore, our general recommendation is to use ARROWS for the following cases: 1) targets that are inherently stable; 2) targets that are metastable with respect to phase separation; and 3) targets that are metastable with respect to polymorphic transition but have lower surface energies than the ground states.

4.3 Discussion

Precursor selection often has a marked effect on the outcomes of solid-state synthesis experiments, dictating whether they form desired products or unwanted impurities[84,90]. The importance of choosing optimal precursors is demonstrated by our syntheses targeting YBCO, for which only 10 precursor sets (out of 47 total) are successful in forming YBCO without any detectable impurity phases. Similarly, both NTMO and LTOPO were found to require the use of specific precursor sets that circumvent the formation of competing phases that otherwise limit the yield of the metastable targets. Changing just one precursor can lead to a completely different synthesis outcome, as shown by the 94% wt. increase observed in the yield of NTMO when MoO_2 is replaced by MoO_3. Understanding the origin of such large changes requires a detailed inspection of their associated reaction pathways. While this would typically be accomplished by using *in situ* characterization techniques, we have shown that information regarding the intermediate phases formed during solid-state synthesis can be gathered by probing different annealing temperatures with fixed hold times. For example, the low-temperature (400 °C) synthesis experiments targeting LTOPO reveal whether $LiTi_2(PO_4)_3$ forms as an intermediate, which subsequently controls the yield of the metastable polymorph at higher temperature (700 °C).

ARROWS effectively uses intermediate phase information gleaned from low-temperature experiments to determine where a given reaction pathway goes wrong. It does so by rationalizing each set of experimental outcomes using pairwise reaction analysis, which assumes that a mixture of solid precursors reacts two phases at a time. This assumption is justified by several previous studies[84,98,128], where *in situ* XRD was used to verify that solid-state reactions often proceed in pairs owing to the limited diffusion lengths of species in the solid medium. In the current work, systematic pairwise reaction analysis is used to identify which precursors react to consume much of the available free energy, thereby reducing the driving force ($\Delta G'$) that remains to form the target. Once this information is known, ARROWS prioritizes experiments based on precursor sets

that are expected to avoid such unfavorable pairwise reactions. Our tests on the YBCO dataset showed this to be an effective approach for the rapid identification of optimal synthesis routes, as ARROWS identified all 10 of the best experimental procedures while sampling less than half of the entire search space. Similarly, it identified successful procedures for the synthesis of two metastable phases, NTMO and LTOPO, while requiring only 35% and 14% of their search spaces to be sampled, respectively.

Efficient data collection in vast experimental domains is a longstanding challenge. Traditional approaches based on design of experiments[129,130], including the D-optimal design algorithm tested here, can aid in the selection of experiments that are most informative to model a quantity such as target yield. However, these methods can fall short when dealing with a particularly large search space or when the quantity of interest is sparsely valued. Both challenges exist in solid-state synthesis, where many precursor combinations are often available for a given target, most of which fail to produce that target in any measurable amount. This warrants the use of active learning algorithms that can efficiently navigate the search space by adapting from failed experiments. Here we evaluated the performance of two such methods, Bayesian optimization and genetic algorithms, when applied to optimize the synthesis of YBCO. While each is known to perform well on continuous variables such as time or temperature[131,132], our tests show that they fail on the discrete task of precursor selection. We suspect their ineffectiveness is caused by using one-hot encodings to represent each precursor set, which fails to capture the similarities and differences between various chemical compounds. Recent work on organic synthesis has shown that black-box optimization techniques can perform well in the selection of molecular precursors when they are represented using physical descriptors such as SMILES strings[131]; however, no such universal representation exists for crystalline materials. Further complicating matters is the fact that precursor sets used in solid-state synthesis often have varied lengths – *i.e.*, some sets contain more precursors than others – which make them difficult to represent using a fixed-length input vector for optimization.

ARROWS systematically explores the search space associated with solid-state synthesis by actively learning from failed experiments. To overcome the limitations outlined in the previous paragraph, ARROWS relies on a single metric (the remaining reaction energy) that can be updated is it reconstructs the path a given synthesis procedure takes. Previous work has demonstrated that reaction energies (ΔG) often dictate the selectivity of competing phases in solid-state synthesis[84,98], and reactions with larger ΔG tend to occur more rapidly[99,116]. Initially, when no intermediates are known, the available reaction energy corresponds to the free energy difference between the target and precursors, thus motivating our choice to first prioritize experiments based on precursor sets with the largest reaction energies. Once intermediates become known, ARROWS re-ranks precursor sets based on their updated reaction energies ($\Delta G'$) remaining to form the target. Using this approach, the algorithm can discard reaction pathways that become trapped in metastable states close in energy to the target.

Notably, a unique feature of ARROWS is that it becomes more efficient at identifying optimal experiments as it builds the size of its pairwise reaction database. This was demonstrated by the correlation between the frequency at which optimal synthesis routes were discovered on the YBCO dataset and the number of pairwise reactions that were collected (**Figure 4-3b**). Further improving the utility of the pairwise reactions learned by ARROWS is their transferability across materials in related chemical spaces. For example, our analysis of the YBCO experiments revealed 34 unique pairwise reactions involving common precursors for Y, Ba, and Cu. Should any of these compounds be used for the synthesis of a new material, ARROWS would operate more effectively

by predicting their reaction outcomes a priori. Predictions of this nature will in general become more abundant as the overlap between chemical spaces increases, specifically when considering target materials with two or more shared elements. As the decisions made by ARROWS require minimal human input, the algorithm is well suited to act as the brain behind autonomous platforms that are currently being developed[68]. With years of continuous and autonomous experimentation, such platforms could lead to the development of a standardized pairwise reaction database that covers much of the periodic table, enabling accurate predictions regarding optimal synthesis routes for new materials without requiring additional experiments. Researchers across the field of solid-state chemistry could also contribute to this database and refer to it for their own synthesis design.

There exist several opportunities to further improve the efficiency and interpretability of ARROWS. The algorithm currently relies on thermodynamic arguments to optimize a target's yield, specifically by assuming that synthesis reactions with large driving force will be most effective. Future work may additionally consider the influence of kinetic factors such as diffusion and nucleation rates, though these are currently challenging to assess in a quantitative fashion due to both computational limitations and a lack of clarity on the relevant conditions under which each process should be evaluated. Related efforts have developed approximate models for nucleation rates that consider the structural similarity between precursor and target materials, in addition to their associated reaction energy[110]. Such factors could be incorporated into ARROWS and its precursor ranking scheme by using structural descriptors based on matminer statistics[133] or graph-based representations[134–136]. Descriptors related to particle morphology and sample density could also be included in the optimization process, as both have been reported to affect synthesis outcomes[137,138].

Beyond the selection of optimal precursors, synthesis planning often requires the heating profile to be carefully designed. Previous work has addressed this challenge by using standard optimization techniques[132], which perform well as the heating profile can be described in terms of continuous variables (*e.g.*, temperature and time). However, our findings show that a more physics-informed approach may also be viable. For the synthesis of each target material studied in this work, ARROWS used a short hold time (4 h) to identify the precursors and temperature that give maximal target yield. If necessary, manual decisions were made to increase the hold time if 1) the target yield was lower than desired, and 2) the leftover reactions needed to grow the target were thermodynamically favorable. In doing so, > 90% yield was obtained for all three target materials we considered. Moreover, it was shown that such high yield was possible only for the precursor sets optimized by ARROWS at short hold time, thereby demonstrating that long hold times need not be used when testing various precursors. Decisions regarding when to extend the hold time after identifying an optimal set of precursors could later be incorporated into ARROWS, enabling further progress toward complete autonomy in solid-state synthesis.

While we have shown that ARROWS performs well on three benchmarks, there may still be room for improvement. To aid in the development of new algorithms for decision making in solid-state synthesis, all data reported in this work is made publicly available. In particular, the YBCO dataset contains experimental outcomes from all the available precursor combinations. This critically includes both positive (successful) and negative (failed) synthesis outcomes, and as such, can be used to train and validate algorithms that require both types of data. We anticipate that such algorithms will not only facilitate a more systematic approach to the planning of synthesis experiments performed by human researchers, but also enable the development of fully autonomous platforms for materials development[68]. An additional benefit of ARROWS specifically, when applied in conjunction with automated synthesis platforms, is that multiple

successful synthesis routes can be learned for a given target. Such information on alternate experimental procedures will be valuable when more practical considerations become important, such as the optimization of morphology, synthesis cost, or the ability to industrially scale up the synthesis of a novel compound.

4.4 Methods

4.4.1 Formulation of the search space

Targeted materials synthesis can be framed as an optimization problem for which the objective is to maximize the yield of a desired phase with respect to several experimental variables including the choice of precursors, synthesis temperature, hold time, and atmospheric conditions. Here we assume a fixed hold time and set of atmospheric conditions (*e.g.*, p_{O_2} and p_{CO_2}) which are supplied by the user for a given target, hence constraining the search space to account only for the selection of precursors and synthesis temperature. To define this search space, ARROWS requires that the user provide a list of compounds that are available to be used as precursors. From this list, all unique precursor combinations are enumerated and those that can be stoichiometrically balanced with the target are recorded as possible precursor sets for it. The number of precursors included in each set is limited to the number of elements in each target. For example, only sets containing ≤ 4 precursors will be considered for the synthesis of a quaternary oxide containing three cations. Because carbonates, hydroxides, and high-valent oxides are often used as precursors in solid-state synthesis, ARROWS accounts for the possibility of CO_2, H_2O, and O_2 byproducts when balancing each chemical reaction. Additional byproducts can be specified when necessary. To determine which synthesis temperatures may be tested, ARROWS requires that the user supply bounds ($T_{\min}, T_{\max}$) and a sampling interval (ΔT). In combination with the total number of balanced precursor sets (N_{sets}), this information defines the search space containing N_{exp} points over which optimization is performed for a given target:

$$N_{\text{exp}} = N_{\text{sets}} \left(\frac{T_{\max} - T_{\min}}{\Delta T} + 1 \right) \quad (1)$$

Any prior knowledge regarding the chemical system should be used when designing the search space. For example, the lower temperature bound ($T_{\min}$) may be chosen to exceed the known decomposition temperatures of all carbonates and hydroxides being considered as precursors. Similarly, the upper temperature bound ($T_{\max}$) may be chosen below the melting points of the precursors if the user wishes to retain a product consisting of solid powder. With respect to precursor selection, it may often be beneficial to exclude compounds that are known to be inert in the temperature ranged being considered; however, this can also be learned by ARROWS through experimentation. The algorithm's self-learning capabilities become critical in chemical systems where the precursor properties are largely unknown.

4.4.2 Initial ranking by ΔG

The thermodynamic driving force behind a chemical reaction is set by the change in the Gibbs free energy (ΔG) between its products and reactants. Under constant temperature and pressure, reactions can occur spontaneously only if they reduce the Gibbs free energy ($\Delta G < 0$) of the system. ARROWS initially ranks all the available precursor sets in order of their reaction energies (ΔG) to form the target. Those with the largest (most negative) ΔG are prioritized, whereas those with $\Delta G \geq 0$ are excluded from consideration. For each set, ΔG of the solid compounds is determined using DFT-calculated 0 K formation energies from the Materials Project[2], along with temperature-

dependent free energies approximated using the machine-learned descriptor developed by Bartel et al.[114] In cases where a novel phase (not available in the Materials Project) is considered, we use the DFT-calculated energy of the convex hull at that phase's composition. For gaseous phases, ΔG is obtained from the experimental NIST database[139]. All reaction energies are normalized per atom of the product phase(s) formed to ensure a consistent comparison between different precursor sets.

The initial ranking by ΔG is intended to prioritize sets of precursors that are expected to react under short timescales; however, such precursors are not necessarily the most effective at forming the target. In addition to having a strong thermodynamic driving force to form the target, precursor sets with large ΔG often have similarly large driving forces to form unwanted impurity phases[90]. We have therefore designed ARROWS to learn from the outcomes of failed experiments by determining which reactions led to the formation of such impurity phases. Details on this process are given in the next two sections.

Our consideration of ΔG is a simplification of the factors that dictate solid-state synthesis. In addition to selecting optimal precursors, the particle morphology and heating rate can also have a substantial influence on reaction outcomes[137,138]. Furthermore, certain compounds may react with the atmosphere prior to heating, *e.g.*, to form carbonates or hydroxides. Such factors are currently not accounted for but could in principle be included by studying the evolution of each individual precursor as a function of temperature and time. Because this information is not generally available for all compounds and precursor powders, the current implementation of ARROWS focuses only on ΔG, which is more readily calculated using the methods described in the previous few paragraphs. Future work may consider incorporating additional properties into the algorithm's ranking scheme, and further details on this possibility are provided in the Discussion section of the main text.

4.4.3 Temperature selection for intermediate identification

To pinpoint the origin of any impurity phases that caused a synthesis procedure to fail, it is necessary to identify the intermediate phases that formed while heating. Previous work has demonstrated that precursors used in solid-state synthesis typically do not transform directly to the final products, but instead proceed through a series of pairwise reactions that form transient intermediate phases and incrementally reduce the free energy of the sample[84,98]. Characterizing these intermediates would traditionally require the use of *in situ* XRD; however, we propose that similar information can be obtained by testing a range of synthesis temperatures for a given precursor set. Assuming that a fixed hold time is used at each temperature, the XRD patterns gathered from the resulting samples provide discrete snapshots of the reaction pathway, from which intermediate phases can be identified in a high-throughput and automated fashion using recently developed ML algorithms[80].

By inspecting the temperature-dependent synthesis outcomes for a given precursor set, ARROWS determines which pairwise reactions occurred while heating. To this end, we assume that any phases detected at a specific temperature (T) may act as reactants that lead to the formation of new phases at the next highest temperature ($T + \Delta T$). Accordingly, when XRD measurements reveal a new phase that is not present in the associated precursor set nor identified as an intermediate phase at lower temperature, ARROWS is tasked with identifying the precise combination of phases responsible for its formation. If a new phase is detected at $T_{\min}$, the algorithm evaluates which two-phase combination(s) from the precursor set have the appropriate compositions (*i.e.*, can be stoichiometrically balanced) to produce that phase. In cases where there exists only one such possible combination, it is recorded as an observed pairwise reaction with an

onset temperature less than $T_{\min}$. A similar procedure is followed when new phases are detected at $T > T_{\min}$, except that ARROWS considers the intermediate phases detected at the next lowest temperature ($T - \Delta T$) as possible reactants.

Oftentimes, different sets of precursors can react to form identical sets of intermediates at low temperature, which subsequently result in the same products upon further heating[140]. To avoid testing all temperatures for such redundant synthesis routes, ARROWS suggests that experiments first be performed at $T_{\min}$ for each precursor set. It then checks whether the observed products and their associated weight fractions differ from those obtained using other precursors sets that were previously tested at $T_{\min}$. Differences as large as 10% are allowed between two sets of products while still considering them to be identical as there is often limited precision in the refinements performed using XRD patterns from multi-phase mixtures. If the observed products for a precursor set are indeed unique, the next highest temperature ($T + \Delta T$) is proposed for that set. This process is repeated until the target is successfully obtained with sufficiently high yield, as specified by the user, or until $T_{\max}$ is reached for the specified precursor set.

By default, ARROWS operates under the assumption that a linear heating ramp is used to reach the specified hold temperature (T). In practice, however, a preheating step is occasionally used to decompose certain precursors at a temperature lower than the specified hold. For example, nitrate precursors such as $LiNO_3$ and $NaNO_3$ are often preheated to avoid rapid evolution of gases at higher temperature[141]. To handle such cases, expected decomposition temperatures and products can be incorporated into the pairwise reaction database prior to running ARROWS. Without the user providing this information, the algorithm will still identify the decomposition product except in cases where that product reacts with another phase prior to XRD measurements, which would otherwise preclude its detection.

4.4.4 Updated ranking by $\Delta G'$

ARROWS learns from previously identified pairwise reactions to make informed decisions regarding optimal synthesis routes. It does so by predicting which intermediates will form upon heating precursor sets that have not yet been tested. An example of this process is given below for an arbitrary target (AB_2C):

$$\text{Precursor set not yet tested: } A + 2B + C \ (\Delta G_{\text{initial}})$$
$$\text{Previously identified pairwise reaction: } A + 2B \rightarrow AB_2 \ (\Delta G_{\text{interm}})$$
$$\text{Reaction using anticipated intermediates: } AB_2 + C \ (\Delta G' = \Delta G_{\text{initial}} - \Delta G_{\text{interm}})$$

In this example, the anticipated intermediate phases were determined based on previous synthesis outcomes that involved a reaction between A and B. The updated reaction energy ($\Delta G'$) to form the target (AB_2C) is then calculated based on the intermediates ($AB_2 + C$) that result from this pairwise reaction. Similar analysis is applied to all precursor sets that have not yet been tested and their reaction energies are updated accordingly. In cases where no intermediates can be predicted, the reaction energy remains unchanged ($\Delta G' = \Delta G$). Following these changes, precursor sets are ranked to prioritize reactions with the most negative $\Delta G'$, *i.e.*, those with the largest thermodynamic driving force at the presumed target-forming step. ARROWS uses the updated ranking to continually suggest new precursor sets until an experiment is found that gives sufficiently high yield of the target phase (as specified by the user) or until all precursor sets have been tested.

We acknowledge that it is generally difficult to ascertain whether a given reaction energy is large enough for the associated transformation to occur within a reasonably short timeframe. The reaction rate is determined not only by the energy change, but also by several factors related

to diffusion and nucleation These rates are highly non-trivial to predict and strongly dependent on the specific chemistry being considered. Given these considerations, ARROWS is designed to rank various precursor sets based on their relative reaction energies to form a target phase, whether from the initial precursors (ΔG) or from the intermediates that form during synthesis ($\Delta G'$). While this affects the order in which different precursor sets are tested, none are excluded for having a low reaction energy. Instead, such precursors will be tested at a later stage in the optimization process, if necessary.

4.4.5 YBCO synthesis

The synthesis of YBCO is most commonly performed using Y_2O_3, CuO, and $BaCO_3$[119]. This combination of precursors requires > 12 h of annealing at 950 °C, in addition to intermittent regrinding, to eliminate the unwanted impurity phases that often appear. In contrast, recent work has shown that by replacing $BaCO_3$ with BaO_2, YBCO can be obtained with high purity while using a shorter anneal time of 30 min[84,120]. These findings highlight the importance of precursor selection and its effect on the yield of YBCO under limited hold time, making it a suitable test case for ARROWS. To this end, we considered 11 common precursors from the Y-Ba-Cu-O space: Y_2O_3, $Y_2(CO_3)_3$, BaO, $BaCO_3$, BaO_2, CuO, $CuCO_3$, Cu_2O, $BaCuO_2$, $Ba_2Cu_3O_6$, and $Y_2Cu_2O_5$. These compounds were combined to form 47 different precursor sets that were each tested at four synthesis temperatures (600, 700, 800, and 900 °C) using a fixed hold time of 4 h.

All binary phases listed in the above paragraph (including the carbonates) were purchased from Sigma Aldrich, whereas the ternaries ($Y_2Cu_2O_5$, $BaCuO_2$, and $Ba_2Cu_3O_6$) were synthesized in-house. For each ternary phase, stoichiometric amounts of the starting materials were mixed in ethanol with six 10 mm stainless steel balls using a high-energy SPEX mill (SPEX SamplePrep 8000 M) for 9 min. The resulting slurry was dispensed into a crucible and dried at 80 °C, following by a high temperature anneal at the specified synthesis temperature for each sample. $Y_2Cu_2O_5$ was made from Y_2O_3 and $CuCO_3$ using a 12 h hold at 1050 °C. $BaCuO_2$ was synthesized from $BaCO_3$ and CuO using a 24 h hold at 910 °C. $Ba_2Cu_3O_6$ was prepared from BaO_2 and CuO using a 24 h hold at 600 °C. The corresponding XRD patterns point to successful synthesis outcomes as each sample contains the desired ternary phase with minimal impurities.

To assess the phase purity for each synthesis product, XRD measurements were performed with an Aeris diffractometer from Panalytical. We used XRD-AutoAnalyzer[80] to analyze the resulting XRD patterns and identify any crystalline phases present. This algorithm relies on a CNN to map each pattern onto a set of constituent phases. Here we trained the network on all phases reported in the ICSD within the space of Y-Ba-Cu-O chemistries. After identifying the phases in each pattern using XRD-AutoAnalyzer, their weight fractions were evaluated through analysis of relative peak intensities. A more careful approach based on Rietveld refinement, which accounts for properties such as grain size and texture, would be required to obtain precise weight fractions. However, this work only requires that we compare relative weight fractions between different experiments, enabling ARROWS to identify the most effective synthesis route for a given target. For YBCO specifically, all experiments were performed prior to optimization, and therefore it was used to evaluate the performance of several algorithms including ARROWS, Bayesian optimization, genetic algorithms, and D-optima design.

4.4.6 NTMO synthesis

The initial discovery of NTMO was enabled by the use of a hydrothermal synthesis procedure whereby an aqueous solution of Na_2TeO_3, TeO_2, and MoO_3 was held at 220 °C for 48 h[117]. More

recently, a solid-state synthesis route was also reported: Na_2CO_3, TeO_2, and MoO_3 were mixed and held at 430 °C for 48 h with intermittent regrinding[126]. We suspect that ARROWS can handle the synthesis of phases such as NTMO, which are metastable with respect to decomposition, as it should learn to avoid the formation of any competing phases that result in an unfavorable driving force ($\Delta G > 0$) to form the target. For the experimental campaign targeting NTMO, eleven precursors were purchased from Sigma Aldrich: Na_2O, Na_2CO_3, NaOH, Na_2O_2, MoO_2, MoO_3, TeO_2, Na_2TeO_3, Na_2MoO_4, $Na_2Mo_2O_7$, and $(NH_4)_2MoO_4$. A total of 23 precursor sets were considered, for which synthesis temperatures of 300 and 400 °C were tested at a fixed hold time of 4 h. We avoided the use of higher temperatures as melting is expected to occur near 450 °C, making the product difficult to extract[117]. In contrast to the YBCO campaign, where all possible experiments were performed and ARROWS was only applied *post hoc*, the LTOPO experiments were carried out under the guidance of ARROWS until NTMO was obtained with a weight fraction exceeding 50%. No black-box optimization techniques were used to explore this dataset as only part of the design space was sampled by ARROWS.

4.4.7 LTOPO synthesis

The tendency for LTOPO to crystallize in its triclinic polymorph, as opposed to its orthorhombic ground state, is highly sensitive to the choice of precursors and synthesis temperature[118,142]. Recent work has proposed that the *t*-LTOPO nucleates first owing to its more stable surface energy, which dictates the relative nucleation rate of each polymorph when ΔG is large[127]. Therefore, although ARROWS encodes no structural information and is not designed for the synthesis of metastable polymorphs in general, we believe it is well-suited for *t*-LTOPO (and similarly stabilized metastable polymorphs) since it aims to identify reaction pathways that maintain large ΔG. Ten commercially available phases were purchased from Sigma Aldrich and used as precursors: Li_2O, Li_2CO_3, LiOH, TiO_2, P_2O_5, $NH_4H_2PO_4$, $(NH_4)_2HPO_4$, Li_3PO_4, Li_2TiO_3, and $Li_4Ti_5O_{12}$. A total of 30 precursor sets were considered. Four synthesis temperatures (400, 500, 600, and 700 °C) were sampled for each set at fixed a hold time of 4 h. Synthesis experiments were performed under the guidance of ARROWS until *t*-LTOPO was obtained with a weight fraction exceeding 50%. No black-box optimization techniques were applied.

CHAPTER 5 An autonomous laboratory for the accelerated synthesis of inorganic materials

The work presented in this chapter is based, often verbatim, on the following publication:
***N. J. Szymanski**, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, and G. Ceder, "An autonomous laboratory for the accelerated synthesis of novel materials." Nature 2023, 624, 86-91.*

To close the gap between the rates of computational screening and experimental realization of novel materials[143,144], we introduce the A-Lab, an autonomous laboratory for the solid-state synthesis of inorganic powders. This platform leverages computations, historical data from the literature, ML, and active learning to plan and interpret the outcomes of experiments performed using robotics. Over 17 days of continuous operation, the A-Lab realized 41 compounds from a set of 58 targets including a variety of oxides and phosphates that were identified using large-scale ab-initio phase stability data from the Materials Project and Google Brain. Synthesis recipes were proposed by natural language models trained on the literature and optimized using an active learning approach grounded in thermodynamics. Analysis of the failed syntheses provide direct and actionable suggestions to improve current techniques for materials screening and synthesis design. The high success rate demonstrates the effectiveness of AI-driven platforms for autonomous materials discovery and motivates further integration of computations, historical knowledge, and robotics.

5.1 Introduction

While promising new materials can be identified at scale using high-throughput computations, their experimental realization is often challenging and time-consuming. Accelerating the experimental segment of materials discovery requires not only automation, but *autonomy* – the ability of an experimental agent to interpret data and make decisions based on it. Pioneering efforts have demonstrated autonomy in several aspects of materials research including robotic and Bayesian-driven optimization of carbon nanotube yield[21,22], photovoltaic performance[12], and photocatalysis activity[13]. In contrast to conventional ML algorithms used for optimization, human researchers benefit from a wealth of background knowledge that informs their decision making, and it is increasingly recognized[15,58,145] that autonomy will require a fusion of encoded domain knowledge, access to diverse data sources, and active learning.

Here we present the A-Lab, an autonomous laboratory that integrates robotics with the use of ab-initio databases, ML-driven data interpretation, synthesis heuristics learned from text-mined literature data, and active learning to optimize the synthesis of novel inorganic materials in powder form. While autonomous workflows based on liquid handling have been demonstrated in organic chemistry[146–148], the A-Lab addresses the unique challenges of handling and characterizing solid inorganic powders. These often require milling to ensure good reactivity between precursors,

which can have a wide range of physical properties related to differences in their density, flow behavior, particle size, hardness, and compressibility. The use of solid powders is well suited for manufacturing and technological scale-up, and the A-Lab's approach to synthesis produces multi-gram sample quantities that facilitate device-level testing.

Given a set of air-stable target materials (*i.e.*, desired synthesis products whose yield we aim to maximize) screened using the Materials Project[2], the A-Lab generates synthesis recipes using ML models trained on historical data from the literature, then executes these recipes with robotics. The synthesis products are characterized by XRD, with two ML models working in unison to analyze their patterns. When synthesis recipes fail to produce a high target yield, active learning closes the loop by proposing improved follow-up recipes. Over 17 days of operation, the A-Lab successfully synthesized 41 of 58 target materials which span 33 elements and 41 structural prototypes. Inspection of the 17 unobtained targets revealed synthetic as well as computational failure modes, several of which could be overcome through minor adjustments to the lab's decision making. With its high success rate in validating predicted materials, the A-Lab showcases the collective power of ab-initio computations, ML algorithms, accumulated historical knowledge, and automation in experimental research.

5.2 Results

5.2.1 Autonomous materials discovery platform

The materials discovery pipeline followed by the A-Lab is schematically shown in **Figure 5-1**. All target materials considered in this work are new to the lab, *i.e.*, not present in the training data for the algorithms it uses to propose synthesis recipes. The experiments reported in this study represent the first attempts by the A-Lab to synthesize any of these targets. Each target is predicted to be on or very near (< 10 meV/atom) the convex hull formed by stable phases taken from the Materials Project and cross-referenced with an analogous database from Google Brain. Because the A-Lab handles samples in open air, we only considered targets that are predicted not to react with O_2, CO_2, and H_2O (Methods).

For each compound proposed to the A-Lab, up to five initial synthesis recipes are generated by an ML model that has learned to assess target "similarity" through natural language processing of a large database of syntheses extracted from the literature[100], mimicking a human's approach to base an initial synthesis attempt on analogy to known related materials. A synthesis temperature is then proposed by a second ML model trained on heating data from the literature (Methods)[99]. If these literature-inspired recipes fail to produce > 50% yield for their desired targets, the A-Lab continues to experiment using ARROWS, an active learning algorithm that integrates ab-initio computed reaction energies with observed synthesis outcomes to predict solid-state reaction pathways[149]. Experiments are performed under the guidance of this algorithm until the target is obtained as the majority phase, or all synthesis recipes available to the A-Lab are exhausted.

The A-Lab carries out experiments using three integrated stations for sample preparation, heating, and characterization with robotic arms transferring samples and labware between them. The first station dispenses and mixes precursor powders before transferring them into alumina crucibles. A robotic arm from the second station loads these crucibles into one of four available box furnaces to be heated (Methods). After allowing the samples to cool, another robotic arm transfers them to the third station where they are ground into a fine powder and measured by XRD. The lab's operations are controlled through an Application Programming Interface (API), which enables on-the-fly job submission from human researchers or decision-making agents.

The phase and weight fractions of the synthesis products are extracted from their XRD patterns by probabilistic ML models trained on experimental structures from the ICSD following methodology outlined in previous work[80,150]. Because the target materials considered in this work have no experimental reports, their diffraction patterns are simulated from computed structures available in the Materials Project and corrected to reduce DFT errors. For each sample, the phases identified by ML are confirmed with automated Rietveld refinement (Methods) and the resulting weight fractions are reported to the A-Lab's management server to inform subsequent experimental iterations, if necessary, in search of an optimal recipe with high target yield.

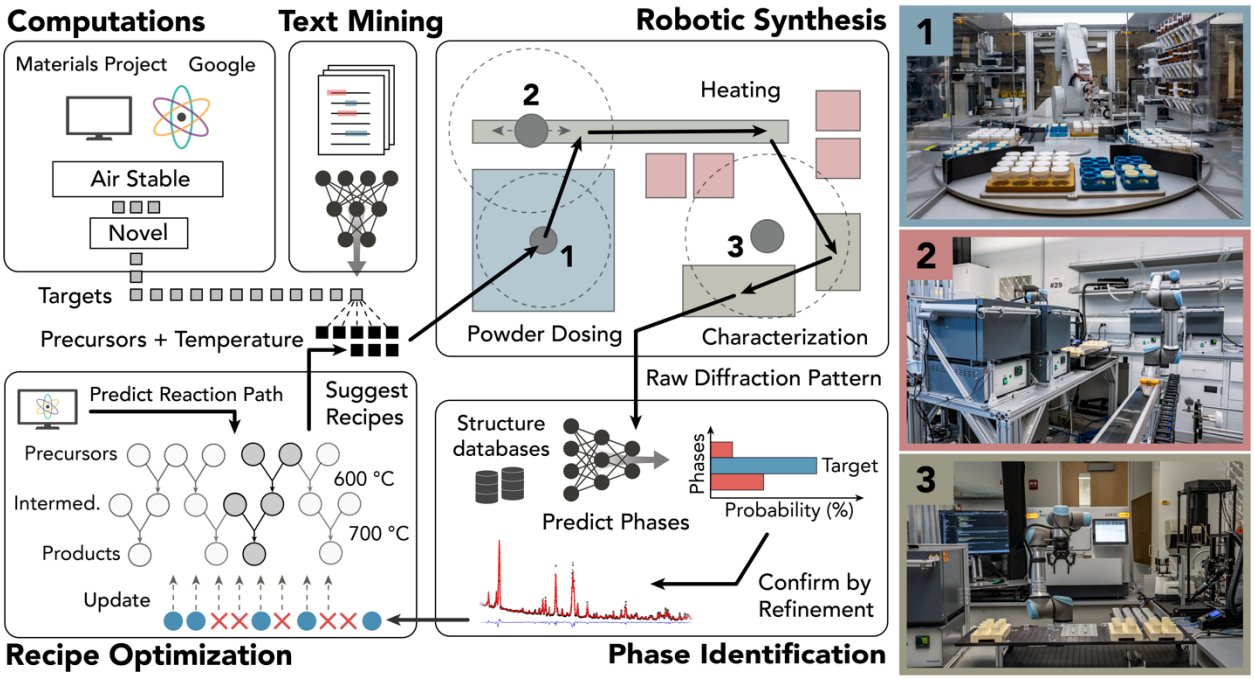

Figure 5-1. The autonomous materials discovery workflow employed by the A-Lab. Air-stable unreported targets are identified using DFT-calculated convex hulls consisting of ground states from the Materials Project and Google Brain. Synthesis recipes for each target are proposed using ML models trained on synthesis data from the literature. These recipes are tested using a robotic laboratory that automates 1) powder dosing, 2) sample heating, and 3) product characterization with XRD. All sample transfer between these stations is performed using robotic arms, forming a fully automated sequence from chemical input to characterization. Phase purity is assessed from XRD, which is analyzed by ML models trained on structures from the Materials Project and the ICSD, and confirmed with automated Rietveld refinement. In cases where high (> 50%) target yield is not obtained, new synthesis recipes are proposed by an active learning algorithm that identifies reaction pathways with maximal driving force to form the target.

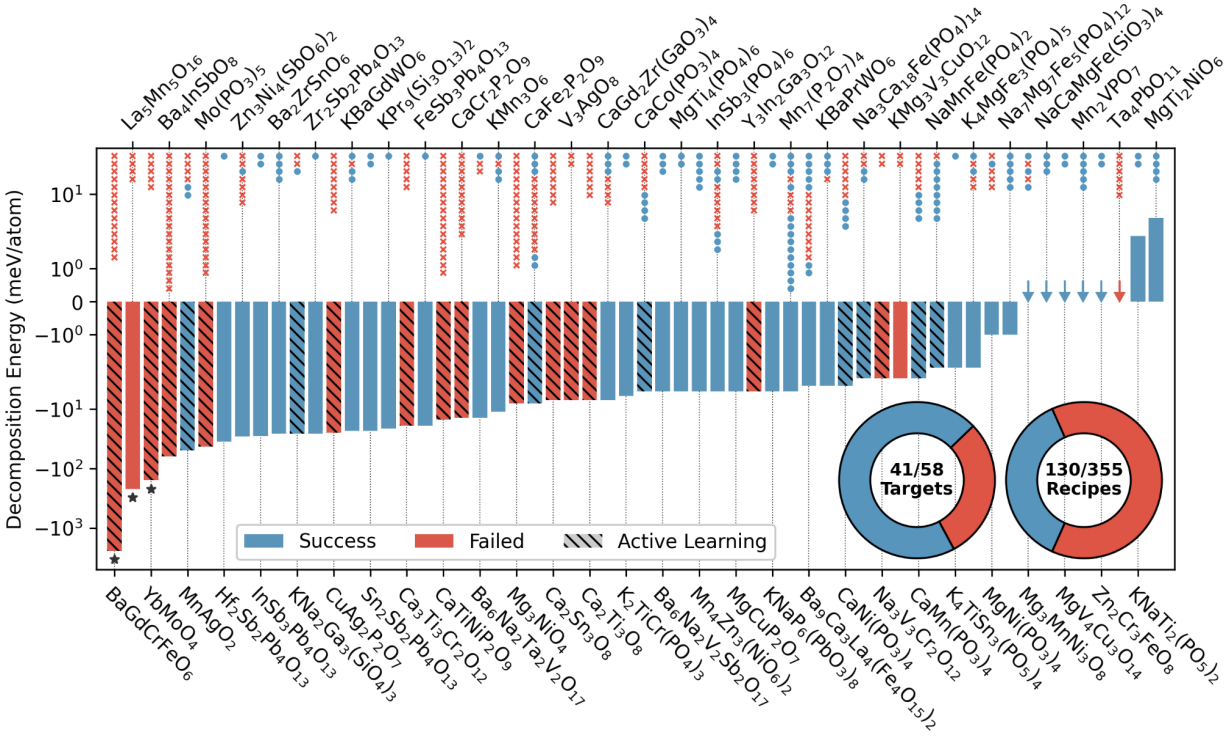

Figure 5-2. Outcomes from the A-Lab's targeted syntheses of DFT-predicted materials. Result summary from the syntheses targeting 58 compounds, plotted against their decomposition energies (log-scale). Arrows indicate values near zero. A total of 41 targets were successfully synthesized (blue bars), while the remaining 17 could not be obtained by the A-Lab (red bars). Targets optimized in the active learning stage of the A-Lab are marked by diagonal lines; all other targets were only attempted using recipes proposed by ML algorithms trained on literature data. The scatter points above each bar represent the outcomes of attempted recipes for each target, ordered from top to bottom in the sequence they were performed. The inset pie charts show the fraction of successful targets (left) and recipes (right). Analyses performed after the fact suggest that the calculated decomposition energies for three targets, marked with stars, may be suspect due to computational errors (see text).

5.2.2 Experimental synthesis outcomes

Using the described workflow, the A-Lab synthesized 41 of the 58 target compounds over 17 days of continuous experimentation, representing a 71% success rate. We show in the next section that this success rate could be improved to 74% with only minor modifications to the lab's decision-making algorithm, and further to 78% if the computational techniques were also improved. The high success rate demonstrates that comprehensive ab-initio calculations can be used to effectively identify new, stable, and synthesizable materials. The outcome for all 58 compounds is plotted in **Figure 5-2** against their decomposition energies (on a log scale), a common thermodynamic metric that describes the driving force to form a compound from its neighbors on the phase diagram[88]. A negative (positive) decomposition energy indicates that a material is stable (metastable) at 0 K. Of the targets considered in this work, 50 are predicted to be stable, while the remaining 8 are metastable but lie near the convex hull. Over the range of decomposition energies considered, we do not observe a clear correlation between decomposition energy and whether a material was successfully synthesized.

In total, 35 of the 41 materials synthesized by the A-Lab were obtained using recipes proposed by ML models trained on synthesis data from the literature. These literature-inspired recipes were more likely to succeed when the reference materials are highly similar to our targets, confirming that target "similarity" is a useful metric to select effective precursors[151]. At the same time, precursor selection remains a highly nontrivial task, even for thermodynamically stable materials. Despite 71% of targets eventually being obtained, only 37% of the 355 synthesis recipes tested by the A-Lab produced their targets. This finding echoes previous work that has established the strong influence of precursor selection on the synthesis path, ultimately deciding whether it forms the target or becomes trapped in a metastable state[98,110].

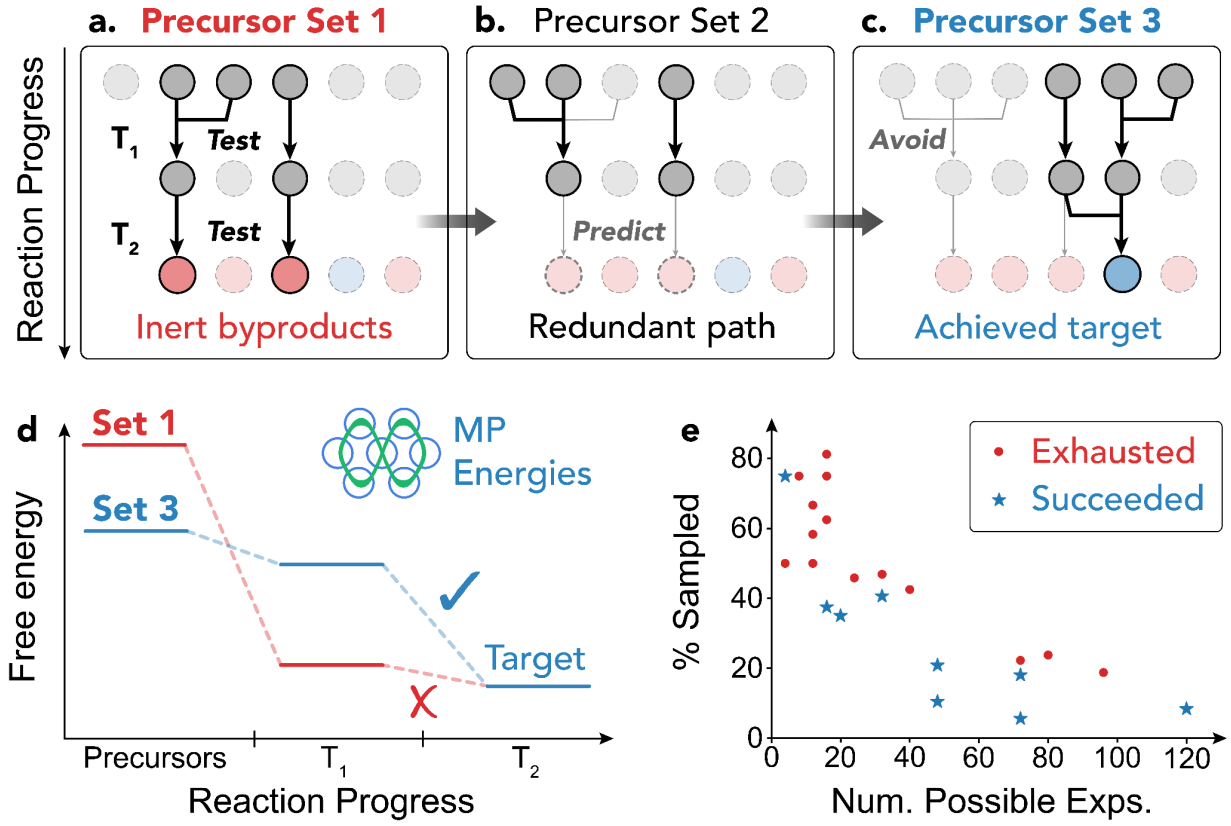

Figure 5-3. An outline of the active learning cycle that drives the A-Lab's targeted syntheses. (a) From a failed synthesis attempt, the A-Lab determines which pairwise reactions occurred. (b) New precursors are recommended by substituting at least one precursor involved in the unfavorable pairwise reaction. In cases where the new precursor set leads to identical intermediates as a previously tested set, it is not explored at any higher temperatures. (c) The successful precursor set avoids all the unfavorable pairwise reactions. (d) The free energy at each step in the reaction pathway, calculated using data from the Materials Project, which shows that the successful pathway maintains a large driving force at the target-forming step. € Following this approach, the scatter plot shows the number of experiments required to exhaust all unique reaction paths for each target (red), or to identify an optimal path with high yield (blue), plotted with respect to the total size of each experimental search space.

The A-Lab's active learning cycle identified synthesis routes with improved yield for nine targets, of which six had zero yield from the initial literature-inspired recipes. Targets optimized with active learning are indicated by the bars containing diagonal lines in **Figure 5-2**. In this framework, improved synthesis routes are designed using two hypotheses: 1) solid-state reactions tend to occur between two phases at a time (*i.e.*, pairwise)[84]; and 2) intermediate phases that leave only a small driving force to form the target material should be avoided, as they often require long reaction time and high temperature[89,90].

The A-Lab continuously builds a database of pairwise reactions observed in its experiments – 88 unique pairwise reactions were identified from the synthesis experiments performed in this work. This database allows the products of some recipes to be inferred (**Figure 5-3**), precluding their testing; a recipe that yields an observed set of intermediates (already present in the lab's database) need not be pursued at higher temperatures, as the remaining reaction pathway is already known. This can reduce the search space of possible synthesis recipes by up to 80% when many precursor sets react to form the same intermediates (**Figure 5-3e**). Furthermore, knowledge of reaction pathways can be used to prioritize intermediates with a large driving force to form the target, computed using formation energies available in the Materials Project. For example, the synthesis of $CaFe_2P_2O_9$ was optimized by avoiding the formation of $FePO_4$ and $Ca_3(PO_4)_2$, which have a small driving force (8 meV/atom) to form the target. This led to the identification of an alternative synthesis route that forms $CaFe_3P_3O_{13}$ as an intermediate, from which there remains a much larger driving force (77 meV/atom) to react with CaO and form $CaFe_2P_2O_9$, causing a ~70% increase in the target's yield.

5.2.3 Barriers to synthesis

Seventeen of the 58 targets evaluated by the A-Lab were not obtained even after its active learning cycle. We identify slow reaction kinetics, precursor volatility, amorphization, and computational inaccuracy as four broad categories of "failure modes" that prevented the synthesis of these targets. The prevalence of each failure mode is shown in **Figure 5-4**, accompanied by their affected targets.

Sluggish reaction kinetics hindered 11 of the 17 failed targets, each containing reaction steps with low driving forces (< 50 meV/atom). In principle, these targets can be made accessible by using a higher synthesis temperature, longer heating time, improved precursor mixing, or intermittent regrinding – standard procedures that are currently outside the domain of the A-Lab's active learning algorithm. As such, we manually reground the original synthesis products generated by the A-Lab and heated them to higher temperatures, which led to the successful formation of two additional targets, $Y_3Ga_3In_2O_{12}$ and Mg_3NiO_4, bringing our total success rate to 74%. One could also use more reactive precursors to provide a greater driving force to form the target; though, our experiments were constrained to air-stable binary precursors that sometimes restricted the A-Lab's choice of synthesis routes to those forming highly stable intermediates. System modifications to enable multi-step heating, intermediate regrinding, and expanded precursor selection should improve the lab's ability to adapt and overcome failed synthesis attempts.

Precursor volatility disrupted all synthesis experiments targeting $CaCr_2P_2O_9$, causing a change in the net stoichiometry of its samples. This can be attributed to the use of ammonium phosphate precursors, $NH_4H_2PO_4$ and $(NH_4)_2HPO_4$, which proceed through a series of decomposition reactions and ultimately evaporate above 450 °C[152]. Still, recipes based on these precursors can succeed if the ammonium phosphate reacts with another precursor prior to its evaporation temperature, effectively locking the phosphate ions in the solid state. For example,

volatility does not appear to be an issue for the Mn-containing phosphates targeted in this work, as each Mn oxides precursor reacts with the ammonium phosphates at low temperature (< 500 °C) to form $Mn_2(PO_4)_3$ as an intermediate. This precursor behavior can in principle be learned when sufficient pairwise reaction data has been collected, after which the A-Lab may favor the selection of precursors that trap in phosphate ions at low temperature and therefore preclude unwanted volatility.

Melting of samples at high temperature inhibited the crystallization of one target, $Mo(PO_3)_5$, whose synthesis attempts produced amorphous samples. While the use of a molten flux can sometimes improve reaction kinetics[153], the formation of an amorphous state that is low in energy may reduce the driving force for crystallization. Indeed, using the workflow outlined by Aykol *et al.*[154], we identified amorphous configurations of $Mo(PO_3)_5$ with energies as low as 61 meV/atom above the crystalline ground state, a finding that is consistent with the widely reported glass-forming ability of phosphate-rich compounds[155,156].

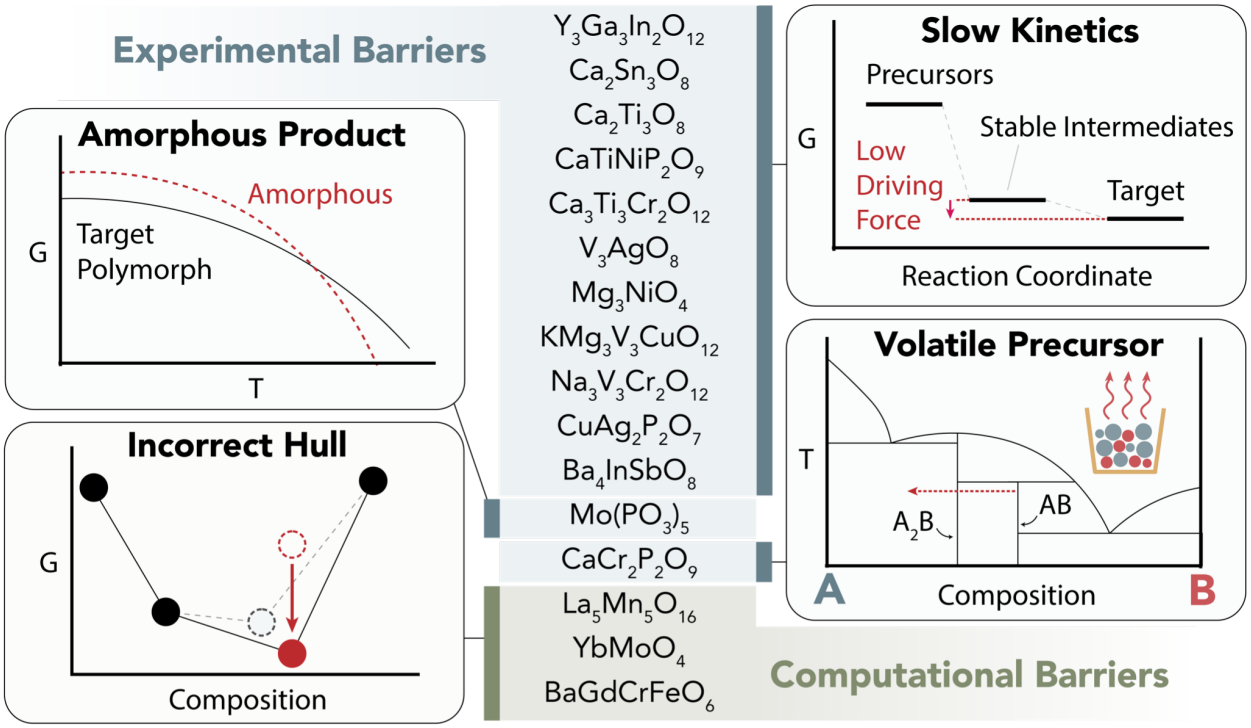

Figure 5-4. Barriers to the synthesis of materials predicted to be stable. The 17 target materials that could not be synthesized by the A-Lab, where each is categorized by the feature that complicates its synthesis. One target (Ta_4PbO_{11}) is excluded from this list, as it is metastable and therefore was predictably unobtained in favor of its stable competitors. The challenges in synthesizing the remaining 16 stable targets fall under two categories: experimental barriers (blue, 13 targets) and computational barriers (green, 3 targets). We distinguish these barriers into four distinct failure modes: slow reaction kinetics, precursor volatility, product amorphization, and limitations associated with DFT calculations performed at 0 K. A schematic explanation for each failure mode is provided.

Some failure modes result from inaccuracies in the computed stability of the target and therefore cannot be addressed by modifications to the experimental procedures. Fundamental electronic structure challenges are likely affecting $La_5Mn_5O_{16}$, as all the attempts to synthesize this phase instead yielded $LaMnO_3$, which DFT unexpectedly predicts to be highly unstable (120 meV/atom above the hull) even though it is widely reported in the literature to be experimentally accessible[157]. If the energy of $LaMnO_3$ were lowered, consistent with its experimental stability, $La_5Mn_5O_{16}$ would be destabilized (above the hull). Errors in the computed energy of $LaMnO_3$ may arise from its strong Jahn-Teller activity[158], compositional off-stoichiometry[159], or the presence of f-states in La – all of which present challenges to conventional DFT. Problems with $YbMoO_4$ were found to be due to a poor pseudopotential choice in the Materials Project which destabilizes the well-known oxide, Yb_2O_3, and it is likely that in more accurate calculations $YbMoO_4$ is not stable. A similar lanthanide-related electronic structure problem may also be responsible for the failure to synthesize $BaGdCrFeO_6$. These examples demonstrate the ability of A-Lab to provide important feedback to high-throughput computed datasets. With improved calculations that exclude the computationally problematic compounds in this work, our total success rate would increase to 78% (43/55 targets).

5.3 Discussion

In 17 days of closed-loop operation, the A-Lab performed 355 experiments and successfully realized 41 of 58 compounds with diverse structures and chemistries. This unexpectedly high success rate (71%) for the synthesis of computationally predicted compounds was achieved by integrating robotics with 1) DFT-computed data to survey the energetic landscape of precursors, reaction intermediates, and final products; 2) heuristic suggestions for synthesis procedures obtained from ML models trained on text-mined synthesis data; 3) ML interpretation of experimental data; and 4) an active learning algorithm that improves upon failed synthesis procedures. The study also revealed several opportunities to enhance the lab's active learning algorithm by addressing failures caused by slow reaction kinetics, which would enable an improved success rate of 74% with in-line solutions.

Our paper demonstrates that autonomous research agents can dramatically accelerate the pace of materials research. Researchers initialized the A-Lab by proposing 58 target materials, which were successfully realized at a rate of > 2 new materials per day with minimal human intervention. Such rapid discovery points to a vast landscape of opportunities in materials synthesis and development. While the current work focused on a limited subset of all possible synthesis targets, many new candidates await evaluation. As the breadth of ab-initio computations continues to grow, so will this list of novel materials.

Advances in simulations, ML, and robotics have intersected to enable "expert systems" that display autonomy as an emergent quality by the sum of its automated components. The A-Lab demonstrates this by combining modern theory- and data-driven ML techniques with a modular workflow that can discover novel materials with minimal human input. Lessons learned from ongoing experiments can inform both the system itself and the greater community through systematic data generation and collection. The systematic nature of the A-Lab provides a unique opportunity to answer fundamental questions regarding the factors that govern the synthesizability of novel materials, serving as an experimental oracle to validate predictions made based on data-rich resources such as the Materials Project. In future iterations of the platform, such an oracle may be expanded to probe factors beyond synthesizability including microstructure and device

performance. Though our current success rate for the synthesis of novel compounds is high, the remaining discrepancies between current predictions and their experimental outcomes is a crucial signal required to improve our understanding of materials synthesis.

5.4 Methods

5.4.1 Materials screening

The 58 targets evaluated by the A-Lab were identified from the Materials Project database (version 2022.10.28). We first obtained all entries from the Materials Project that were marked as "theoretical" (*i.e.*, not represented in the ICSD) and predicted to be thermodynamically stable (at 0 K) or very close to the convex hull (< 10 meV/atom). We did not consider materials with ≤ 2 elements, nor those containing elements that are radioactive (Ac, Th, Pa, U, Np, Pu, Tc), exceedingly rare (Pd, Pt, Rh, Ir, Au, Ru, Os, Re, Tl, Sc, Tm, Pm, Rb, Cs), or toxic (Hg, As). Due to concerns with the experimental handling of certain material systems (*e.g.*, sulfides), we constrained our selection to only include the following types of materials: oxides, carbonates, bicarbonates, hydroxides, sulfates, sulfites, bisulfates, silicates, fluorides, chlorides, bromides, orthoborates, metaborates, tetraborates, phosphates, phosphites, chlorates, chlorites, and hypochlorites. Finally, we removed all compounds predicted to have uncommon and potentially challenging oxidation states (*e.g.*, Co^{4+}), as determined by *pymatgen*[63].

The novelty of each candidate material was verified by cross-checking with several experimental sources. We first removed all compositions that appeared in SynTERRA, a text-mined set of experimental synthesis data extracted from more than 24,000 publications[160]. Additionally, we removed any materials with compositions appearing in the Handbook of Inorganic Substances (de Gruyter)[161]. While these methods are not exhaustive, they provide an automated and high-throughput approach to screen for materials novelty. For the remaining 432 candidates that were labeled as previously unsynthesized using this workflow, we filtered by thermodynamic stability in air. This was done by calculating the formation energy of each compound in a grand potential with respect to oxygen, assuming standard atmospheric conditions (p_{O2} = 21,200 Pa) and temperatures ranging from 600-1100 °C. We further checked for reactivity with CO_2 and H_2O under those same conditions by using the *Interface Reactions* module in *pymatgen*. From the resulting list of 146 new compounds that were stable in air, we selected 58 targets for which precursors were readily available.

The algorithm we used for identifying potential synthesis targets is available on GitHub (github.com/mattmcdermott/novel-materials-screening). It operates autonomously once given the following information: which elements to consider in the target materials, how large an upper limit to impose on each material's energy above the convex hull, the atmospheric conditions under which the materials will be synthesized, and a threshold on the reaction energies that exist between each material and the gaseous species present in the specified atmosphere. The algorithm then scrapes the Materials Project and produces a list of candidate materials that satisfy these criteria. Additional filtering may be considered based on the availability and cost of precursors for each target. While this is done manually in the current version of the algorithm, potential improvements could automate the process by leveraging online data from chemical inventory lists and vendor websites.

5.4.2 Synthesis recipes from text-mined knowledge

We have established a pipeline for recommending synthesis recipes by using a knowledge base of 33,343 solid-state synthesis procedures extracted from 24,304 publications[160]. For a given target, the initial recipe is selected based on the most common precursors in the knowledge base. We then transition to a similarity-based strategy for recipe selection[100,151]. Each target is transformed into a numerical vector by using a synthesis context-based encoding model. The similarity between a given (new) target and each known material in the knowledge base is evaluated using the cosine similarity between their encoded vectors. After identifying the reference material that is most similar to the target, its precursors are included in the new recommendation. When these precursors do not cover all the elements in the target, we use a masked precursor completion model to account for such missing precursors. Subsequent recommendations are implemented by moving down the list of known materials ranked to be most similar to the target.

For each set of recommended precursors, the most effective synthesis temperature is predicted using an XGBoost regressor trained in previous work[99]. The target and its associated precursors are transformed into three sets of features: (1) precursor properties including melting points, standard enthalpies of formation, and standard Gibbs free energies of formation; (2) target compositional features indicating which elements are present; and (3) the calculated thermodynamic driving force associated with pairwise reaction paths from precursors to target. While the proposed synthesis temperature is dependent on the precursors, not just the target, it may vary for each recipe. However, to maximize the efficiency with which the A-Lab operates, we chose to use one fixed temperature for each target. This temperature was calculated by averaging the proposed synthesis temperatures for the top five precursor sets recommended for a given target. This allowed all such precursor sets to be batched in a single furnace.

5.4.3 Robotic synthesis & characterization

The A-Lab performs fully automated solid-state synthesis and characterization. It is a bespoke robotic platform that consists of a precursor preparation station with a central robot arm (Mitsubishi) for powder dispensing and mixing (custom-made with Labman Automation Ltd.), a high-temperature heating station with four box furnaces (based on F48055-60, Thermo Scientific with custom actuators to control its door), a product handling station developed in-house for powder retrieval and sample loading, a characterization station with a powder X-ray diffractometer (Aeris Minerals, Malvern Panalytical), and two collaborative robot arms (UR5e, Universal Robots) that transfer samples and labware between stations.

The synthesis process starts from the precursor preparation station, where the necessary consumables (plastic vials, ZrO_2 balls, and crucibles) and precursor dosing bottles containing between 50 and 100 g of powders are manually loaded prior to starting a new experimental campaign. Prescribed amounts of the precursor powders are dispensed into a plastic vial by an automatic dispenser-balance (Quantos, Mettler Toledo). The precursor powders are then mixed thoroughly with ethanol and ten 5 mm ZrO_2 balls in a Dual Asymmetric Centrifuge (Smart DAC250, Hauschild) for 9 minutes. To ensure proper slurry viscosity, the ethanol amount is automatically calculated based on the amount and density of each powder comprising the mixture. The resulting slurry is transferred with an automated pipettor (rLine LH-710969, Sartorius) into an alumina crucible, which is then dried at 80 °C in a closed evaporation system. A UR5e robot arm on a linear rail (Olympus Controls) removes the dried samples from the precursor preparation station and loads them into one of four box furnaces. Heating is executed in batches, with each furnace containing up to 8 samples on an alumina tray. Each batch is heated to 300 °C with a slow

ramping rate of 2 °C/min to raise the likelihood that any phosphate precursor has time to react before it becomes volatile at higher temperature. The samples are then heated to the specified synthesis temperature with a nominal ramp rate of 15 °C/min, followed by a 4 h dwell. After the dwell is complete, the samples are naturally cooled to 100 °C, at which point a UR5e arm removes the samples from the furnace and waits another 10 minutes to allow the samples to cool to room temperature.

A separate UR5e arm transfers the cooled samples to the next station for powder retrieval and characterization. There, a 10 mm alumina ball is placed in each crucible by an automatic ball dispenser developed in-house, then sent to a vertical shaker that grinds the samples into fine powders. The resulting powders are then poured by the UR5e arm from the crucibles into a clean plastic vial covered using a steel mesh. By inverting the container, the powder is dispensed through the mesh onto an XRD sample holder and subsequently flattened with an acrylic disk. The UR5e arm transfers each flattened sample into the diffractometer for X-ray measurements, which are performed using 8-min scans that range from $2\theta = 10°$ to 100°. The XRD sample holders must be cleaned manually when the lab has depleted its stock. Precursor powders should also be refilled or replaced, when necessary, though this can be performed without halting the lab's workflow.

5.4.4 Phase analysis

Given an XRD pattern obtained from an unknown sample, we apply XRD-AutoAnalyzer to identify the constituent phases and estimate their weight fractions[80]. This algorithm relies on a CNN consisting of six convolutional layers, with max pooling applied between each, followed by three fully connected layers with ReLU activation. Batch normalization and a dropout rate of 50% is applied between the fully connected layers for regularization. At inference, we apply Monte Carlo dropout to create an ensemble of 100 networks with 50% of their connections randomly excluded. The final prediction is taken as the phase that appears most frequently in the ensemble, and its associated confidence is defined as the fraction of models that predict it.

A unique model instance is trained on the chemical space defined by each target. Experimental structure entries with elements shared by the given target are extracted from the ICSD, also including carbonates and hydroxides. For the DFT-calculated target, we apply a machine learned volume correction to its lattice parameters prior to including it in the training set. From each reference phase, 200 diffraction patterns are simulated with stochastic variations derived from experimental artifacts including lattice strain, crystallographic texture, impurity peaks, and poor crystallinity. These augmented patterns are used to train the CNN for 50 epochs, after which they are ready for the analysis of novel patterns.

To confirm the predictions of the CNN, we use an automated approach to multi-phase Rietveld refinement. An agent with two deep neural networks (actor/critic) were trained using reinforcement learning based on a proximal policy optimization algorithm implemented in a custom gym environment that interacts with the GSAS-II software package through a scripting interface[162,163]. The environment is initialized by refining the background, followed by the scale factor and sample displacement. After initialization is performed based on these parameters, the algorithm freely refines the lattice parameters, phase fractions, isotropic microstrains, and particle sizes. For each step in the refinement, our algorithm decides which parameters to refine and/or reset to the initial values with the objective of minimizing the difference between the calculated and the experimentally observed patterns.

When the automated refinement gives a poor fit, manual analysis is performed. For targets where we suspect the poor fit resulted from configurational disorder, we refined the XRD patterns

using cation-disordered versions of the target's structure taken from the Materials Project. The cations allowed to be exchanged (disordered) with one another were selected based on the Hume-Rothery rules. Such cases were still considered successful so long as the disordered version of the target retained the same crystal structure and overall composition as the ordered version.

5.4.5 Active learning algorithm

Active learning is performed using Autonomous Reaction Route Optimization with Solid-State Synthesis (ARROWS), our recently developed algorithm that learns from previous experimental outcomes to identify improved reaction pathways[149]. Given the products obtained from a set of precursors proposed by our natural language models at temperature T_{NLP}, ARROWS first suggests that a lower temperature (T_{NLP} - 300 °C) be tested for the same precursor set. The intent of this approach is to reveal which intermediate phases lead to the formation of each impurity observed at higher temperature. From the low-temperature synthesis outcome, information is extracted regarding the pairwise reactions that occurred, including those between the precursors (to form the observed intermediates) as well as those between the intermediates (to form the high-temperature impurities). New synthesis experiments are then proposed based on sets of precursors expected to avoid such reactions, giving priority to those with a maximal thermodynamic driving force to form the target. The driving force is calculated as the free energy difference between a target and its associated precursors, where all solid energies (at 0 K) are extracted from the Materials Project and corrected using a machine-learned descriptor that accounts for vibrational entropy contributions at the specified temperature[114].

After testing a precursor set at low temperature (T_{NLP} - 300 °C), iteratively higher temperatures (ΔT = 100 °C) are probed until the target is obtained with a yield exceeding 50%, or until the temperature reaches T_{NLP}. At each step, the algorithm determines which pairwise reactions occurred and records them in a database that is referred to throughout all other experiments performed by the A-Lab. In subsequent iterations, ARROWS prioritizes sets of precursors containing pairs of phases that are expected to form the desired target while avoiding pairs that form unwanted impurities. Moreover, to avoid testing redundant synthesis routes for which different precursors form identical products, the algorithm checks whether the low-temperature (T_{NLP} - 300 °C) intermediates obtained from a given precursor set differ from those obtained with previous (unsuccessful) recipes. If not, then no further experiments are proposed for that set of precursors. This process is repeated until the target is successfully obtained, or until all the available precursor sets have been exhausted.

CHAPTER 6 Summary and Outlook

Achieving complete autonomy in the synthesis of novel inorganic materials requires the automation of experimental procedures for powder handling and XRD, interpretation of the resulting data to determine what products were obtained from a set of experiments, and decision making to learn from these outcomes and design new experiments. We have shown that this is possible with the development of the A-Lab, whose initial application (Chapter 5) resulted in the synthesis of 41 DFT-predicted compounds in just 17 days of experimentation. The success of this platform was enabled by the integration of robotics with algorithms that can automatically identify crystalline phases from XRD patterns, and in cases where these phases do not include the desired synthesis product, suggest new experiments with alternative precursors that facilitate increased yield of the desired product.

As described in Chapter 2, automated XRD interpretation was accomplished by training an ensemble of CNNs on simulated patterns which are systematically augmented to account for common experimental artifacts like strain and texture. Because these models are fast once trained, we showed in Chapter 3 how they can be applied in-line with XRD measurements to steer the diffractometer toward areas of the pattern that matter most for phase identification, substantially reducing the scan time required per sample. The information gleaned from these methods can then be fed into an active learning algorithm designed to optimize experimental procedures for the targeted synthesis of new materials (Chapter 4). This algorithm uses existing synthesis data to construct a database of pairwise reactions that have been observed. This database, coupled with the analysis of ab-initio computed reaction energies, enables the selection of optimal precursors whose reaction pathways maintain a large thermodynamic driving force to form the desired synthesis product.

While the methods described in this work have proven sufficiently effective to synthesize a variety of compounds in the A-Lab, there remain some limitations that warrant further attention. The platform's current hardware is designed only to handle materials that are stable in air at room temperature. This precludes the synthesis of any compounds that readily react with gaseous species (O_2, H_2O, CO_2) present in air. Future iterations of the A-Lab may therefore consider housing its robotics within a glovebox (containing N_2 or Ar) to expand the range of materials that it can make. These iterations may also include the addition of new robotics that can take any samples made by the lab and use them to fabricate devices such as Li-ion batteries. Coupling this with methods to measure device performance (*e.g.*, electrochemical cycling) would provide a unique opportunity to perform end-to-end optimization of the manufacturing process.

Another limitation lies in the analysis techniques used by the A-Lab. Identifying all the compounds present in a multi-phase mixture remains a challenging problem, especially when those phases exhibit strong peak overlap in the corresponding XRD pattern. One approach that could help in overcoming this limitation is the incorporation of physical domain knowledge into the pipeline for phase analysis. For instance, one could form a prior distribution of suspected phases in a sample based on its most likely synthesis products, which could be predicted from ab-initio computed reaction energies. This prior would then be updated based on the phases identified from

the XRD pattern of that sample. This reflects the concept of *data fusion* outlined in Chapter 1. Further building on this idea of leveraging multiple information sources, one could also perform phase identification using additional characterization methods beyond XRD. Perhaps the most promising approach is combining Scanning Electron Microscopy (SEM) with Energy Dispersive Spectroscopy (EDS) to map the distribution of particle compositions within a sample. Analysis of the resulting data, with algorithms such as non-negative matrix factorization, can be used to decompose SEM/EDS data into distinct phases. These phases could then be aggregated with the predictions from computations and XRD to bolster the accuracy of phase analysis.

One final area that could benefit from further advances is the decision-making process used by the A-Lab. Its current implementation focuses solely on choosing precursors that maximize the thermodynamic driving force to form the desired synthesis product, avoiding the formation of inert reaction intermediates that are slow to form that product. However, the experiments discussed in Chapter 5 demonstrated the prevalence of several issues that cause a synthesis trial to fail but are not related to low thermodynamic driving force. These include melting and volatility, which often complicated the synthesis of metal phosphates and could not be overcome by choosing alternative precursors. It would therefore be beneficial to expand the decisions available to the A-Lab, *e.g.*, to use lower synthesis temperatures or slower heating/cooling rates to circumvent these issues. We also observed several cases where highly stable reaction intermediates could not be avoided during synthesis, despite using alternative precursors, leading to sluggish formation of the desired target. To deal with this problem, future iterations of the lab's algorithms could detect the occurrence of slow reactions (by observing intermediates with low thermodynamic driving force to react) and overcome them using improved mixing, higher synthesis temperature, or longer hold time. These advances, in combination with the growing database of pairwise reactions known to the A-Lab, promise to bolster its performance and accelerate the discovery of novel inorganic materials.

REFERENCES

1. D. P. Tabor *et al.* Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat Rev Mater* **5**, 5–20 (2018).
2. A. Jain *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater* **1**, 011002 (2013).
3. S. Kirklin *et al.* The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *Npj Comput Mater* **1**, 15010 (2015).
4. M. Shevlin. Practical High-Throughput Experimentation for Chemists. *ACS Med Chem Lett* **8**, 601–607 (2017).
5. C. A. Nicolaou, I. A. Watson, H. Hu, & J. Wang. The Proximal Lilly Collection: Mapping, Exploring and Exploiting Feasible Chemical Space. *J Chem Inf Model* **56**, 1253–1266 (2016).
6. A. Weber, E. von Roedern, & H. U. Stilz. SynCar: An Approach to Automated Synthesis. *J Comb Chem* **7**, 178–184 (2005).
7. P. Szymański, M. Markowicz, & E. Mikiciuk-Olasik. Adaptation of High-Throughput Screening in Drug Discovery - Toxicological Screening Tests. *Int J Mol Sci* **13**, 427–452 (2011).
8. F. Häse, L. M. Roch, C. Kreisbeck, & A. Aspuru-Guzik. Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent Sci* **4**, 1134–1145 (2018).
9. L. M. Roch *et al.* ChemOS: Orchestrating autonomous experimentation. *Sci Robot.* **3**, eaat5559 (2018).
10. N. S. Eyke, B. A. Koscher, & K. F. Jensen. Toward Machine Learning-Enhanced High-Throughput Experimentation. *Trends Chem* **3**, 120–132 (2021).
11. S. Steiner *et al.* Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
12. B. P. MacLeod *et al.* Self-driving laboratory for accelerated discovery of thin-film materials. *Sci Adv* **6**, eaaz8867 (2020).
13. B. Burger *et al.* A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
14. S. Sun *et al.* Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis. *Joule* **3**, 1437–1451 (2019).
15. A. Ludwig. Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *Npj Comput Mater* **5**, 70 (2019).
16. F. Ren *et al.* Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci Adv* **4**, eaaq1566 (2019).
17. B. Li *et al.* Hydrogen Storage Materials Discovery via High Throughput Ball Milling and Gas Sorption. *ACS Comb Sci* **14**, 352–358 (2012).
18. G. H. Carey & J. R. Dahn. Combinatorial Synthesis of Mixed Transition Metal Oxides for Lithium-Ion Batteries. *ACS Comb Sci* **13**, 186–189 (2011).

19. T. Adhikari *et al.* Development of High-Throughput Methods for Sodium-Ion Battery Cathodes. *ACS Comb Sci* **22**, 311–318 (2020).
20. S. Krishnadasan, R. J. C. Brown, A. J. deMello, & J. C. deMello. Intelligent routes to the controlled synthesis of nanoparticles. *Lab Chaip* **7**, 1434–1441 (2007).
21. P. Nikolaev *et al.* Autonomy in materials research: a case study in carbon nanotube growth. *Npj Comput Mater* **2**, 16031 (2016).
22. J. Chang *et al.* Efficient Closed-loop Maximization of Carbon Nanotube Growth Rate using Bayesian Optimization. *Sci Rep* **10**, 9040 (2020).
23. B. R. Ortiz, J. M. Adamczyk, K. Gordiz, T. Braden, & E. S. Toberer. Towards the high-throughput synthesis of bulk materials: thermoelectric PbTe–PbSe–SnTe–SnSe alloys. *Mol Syst Eng* **4**, 407–420 (2019).
24. T. A. Stegk, R. Janssen, & G. A. Schneider. High-Throughput Synthesis and Characterization of Bulk Ceramics from Dry Powders. *J Comb Chem* **10**, 274–249 (2008).
25. S. Shuang *et al.* High-throughput automatic batching equipment for solid state ceramic powders. *Rev Sci Instrum* **90**, 083904 (2019).
26. XPR Automatic Balance from Mettler Toledo, https://www.mt.com/my/en/home/products/ Laboratory_Weighing_Solutions/ Analytical/Excellence/XPR_Automatic_Balance.
27. Flex SWILE from Chemspeed Technologies, https://www. chemspeed.com/flex-swile-nmr/.
28. S. Gates-Rector & T. N. Blanton. The powder diffraction file: A quality materials characterization database. *Powder Diffr* **34**, 352–360 (2019).
29. A. Altomare *et al.* Advances in powder diffraction pattern indexing: N-TREOR09. *J Appl Cryst* **42**, 768–775 (2009).
30. P. M. d. Wolff. A simplified criterion for the reliability of a powder pattern indexing. *J Appl Cryst* **1**, 108–113 (1968).
31. R. Oishi-Tomiyasu. Reversed de Wolff figure of merit and its application to powder indexing solutions. *J Appl Cryst* **46**, 1277–1282 (2013).
32. J.-M. Le Meins, L. M. D. Cranswick, & A. L. Bail. Results and conclusions of the internet based '"Search/match round robin 2002"'. *Powder Diffr* **18**, 106–113 (2003).
33. C. J. Gilmore, G. Barr, & J. Paisley. High-throughput powder diffraction. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles. *J Appl Cryst* **37**, 231–242 (2004).
34. E. Hernández-Rivera, S. P. Coleman, & M. A. Tschopp. Using similarity metrics to quantify differences in high-throughput data sets: Application to X-ray diffraction patterns. *ACS Comb Sci* **19**, 25–36 (2017).
35. L. A. Baumes, M. Moliner, & A. Corma. Design of a full profile-matching solution for high-throughput analysis of multiphase samples through powder X-ray diffraction. *Chem – Eur* **15**, 4258–4260 (2009).
36. V. Stanev *et al.* Unsupervised phase mapping of X-ray diffraction data by nonnegative matrix factorization integrated with custom clustering. *Npj Comput Mater* **4**, 43 (2018).

37. Y. Iwasaki, A. G. Kusne, & I. Takeuchi. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. *Npj Comput Mater* **3**, 4 (2017).
38. W. B. Park *et al.* Classification of crystal structure using a convolutional neural network. *IUCrJ* **4**, 486–494 (2017).
39. P. M. Vecsei, K. Choo, J. Chang, & T. Neupert. Neural network based classification of crystal symmetries from x-ray diffraction patterns. *Phys Rev B* **99**, 245120.
40. F. Oviedo *et al.* Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks. *Npj Comput Mater* **5**, 60 (2019).
41. P. M. Maffettone *et al.* Crystallography companion agent for high-throughput materials discovery. *Nat Comput Sci* **1**, 290–297 (2021).
42. J.-W. Lee, W. B. Park, J. H. Lee, S. P. Singh, & K.-S. Sohn. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD powder patterns. *Nat Commun* **11**, 705 (2020).
43. Z. Zhou, X. Li, & R. N. Zare. Optimizing chemical reactions with deep reinforcement learning. *ACS Cent Sci* **3**, 1337–1344 (2014).
44. D. Xue *et al.* Accelerated search for materials with targeted properties by adaptive design. *Nat Commun* **7**, 12241 (2016).
45. L. Porwol *et al.* An Autonomous Chemical Robot Discovers the Rules of Inorganic Coordination Chemistry without Prior Knowledge. *Angew Chem Int Ed Engl* **59**, 11256–11261 (2020).
46. W. F. Maier, K. Stöwe, & S. Sieg. Combinatorial and High-Throughput Materials Science. *Angew Chem Int Ed Engl* **46**, 6016–6067 (2007).
47. M.-X. Li *et al.* High-temperature bulk metallic glasses developed by combinatorial methods. *Nature* **569**, 99–103 (2019).
48. R. Zarnetta *et al.* Identification of Quaternary Shape Memory Alloys with Near-Zero Thermal Hysteresis and Unprecedented Functional Stability. *Adv Funct Mater* **20**, 1917–1923 (2010).
49. J. Schmidt, M. R. G. Marques, S. Botti, & M. A. L. Marques. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput Mater* **5**, 83 (2019).
50. L. Himanen, A. Geurts, A. S. Foster, & P. Rinke. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv Sci* **6**, 1900808 (2019).
51. E. Kim *et al.* Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem Mater* **29**, 9436–9444 (2017).
52. P. Raccuglia *et al.* Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
53. T. Lookman, P. V. Balachandran, D. Xue, & R. Yuan. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *Npj Comput Mater* **5**, 21 (2019).

54. W. Huyer & A. Neumaier. SNOBFIT — Stable noisy optimization by branch and fit. *ACM Trans Math Softw* **35**, 1–25 (2008).

55. L. M. Rios & N. V. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J Glob Optim* **56**, 1247–1293 (2013).

56. A. M. Gopakumar, P. V. Balachandran, D. Xue, J. E. Gubernatis, & T. Lookman. Multi-objective optimization for materials discovery via adaptive design. *Sci Rep* **8**, 3738 (2018).

57. B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, & N. d. Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proc IEEE* **104**, 148–175 (2015).

58. S. Sun *et al.* A data fusion approach to optimize compositional stability of halide perovskites. *Matter* **4**, 1305–1322 (2021).

59. R. D. King *et al.* The automation of science. *Science* **324**, 85–89 (2009).

60. H. S. Stein & J. M. Gregoire. Progress and prospects for accelerating materials science with automated and autonomous workflows. *Chem Sci* **10**, 9640–9649 (2019).

61. F. Oviedo *et al.* Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks. *Npj Comput Mater* **5**, 60 (2019).

62. A. Belsky, M. Hellenbrandt, V. L. Karen, & P. Luksch. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Cryst* **B58**, 364–369 (2002).

63. S. P. Ong *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput Mater Sci* **68**, 314–319 (2013).

64. D. T. Murphy, S. Schmid, J. R. Hester, P. E. R. Blanchard, & W. Miiller. Coordination Site Disorder in Spinel-Type LiMnTiO4. *Inorg Chem* **54**, 4636–4643 (2015).

65. JADE Pro; Materials Data MDI: Livermore, CA, USA, 2019.

66. N. Krins *et al.* $LiMn_{2-x}Ti_xO_4$ spinel-type compounds ($x \leq 1$): Structural, electrical and magnetic properties. *Solid State Ion* **177**, 1033–1040 (2006).

67. G. S. Mattei *et al.* Enumeration as a Tool for Structure Solution: A Materials Genomic Approach to Solving the Cation-Ordered Structure of $Na_3V_2(PO_4)_2F_3$. *Chem Mater* **32**, 8981–8992 (2020).

68. N. J. Szymanski *et al.* Toward autonomous design and synthesis of novel inorganic materials. *Mater Horiz* **8**, 2169–2198 (2021).

69. Z. Chen *et al.* Machine learning on neutron and x-ray scattering and spectroscopies. *Chem Phys Rev* **2**, 031301 (2021).

70. M. Ziatdinov *et al.* Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. *ACS Nano* **11**, 12742–12752 (2017).

71. A. Maksov *et al.* Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. *Npj Comput Mater* **5**, 12 (2019).

72. S. Akers *et al.* Rapid and flexible segmentation of electron microscopy data using few-shot machine learning. *Npj Comput Mater* **7**, 187 (2021).

73. J. Liu *et al.* Deep convolutional neural networks for Raman spectrum recognition: a unified solution. *Analyst* **142**, 4067–4074 (2017).

74. M. M. Noack *et al.* A Kriging-based approach to autonomous experimentation with applications to X-ray scattering. *Sci Rep* **9**, 11809 (2019).

75. R. K. Vasudevan *et al.* Autonomous Experiments in Scanning Probe Microscopy and Spectroscopy: Choosing Where to Explore Polarization Dynamics in Ferroelectrics. *ACS Nano* **15**, 11253–11262 (2021).

76. M. A. Ziatdinov *et al.* Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries. *Adv Mater* **34**, 2201345 (2022).

77. P. M. Maffettone *et al.* Gaming the beamlines—employing reinforcement learning to maximize scientific outcomes at large-scale user facilities. *Mach Learn Sci Technol* **2**, 025025 (2021).

78. H. Kohlmann. Looking into the Black Box of Solid-State Synthesis. *Eur J Inorg Chem* 4174–4180 (2019).

79. X. Wei, X. Wang, Q. An, C. Han, & L. Mai. Operando X-ray Diffraction Characterization for Understanding the Intrinsic Electrochemical Mechanism in Rechargeable Battery Materials. *Small Methods* **1**, 170083 (2017).

80. N. J. Szymanski, C. J. Bartel, Y. Zeng, Q. Tu, & G. Ceder. Probabilistic Deep Learning Approach to Automate the Interpretation of Multi-phase Diffraction Spectra. *Chem Mater* **33**, 4204–4215 (2021).

81. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, & A. Torralba. Learning Deep Features for Discriminative Localization. Preprint at arXiv:1512.04150. (2015).

82. R. Murugan, V. Thangadurai, & W. Weppner. Fast lithium ion conduction in garnet-type $Li_7La_3Zr_2O_{12}$. *Angew Chem Int Ed Engl* **46**, 7778–7781 (2007).

83. A. Neumann & D. Walter. The thermal transformation from lanthanum hydroxide to lanthanum hydroxide oxide. *Thermochim Acta* **445**, 200–204 (2006).

84. A. Miura *et al.* Observing and Modeling the Sequential Pairwise Reactions that Drive Solid-State Ceramic Synthesis. *Adv Mater* **33**, 2100312 (2021).

85. R. R. Selvaraju *et al.* Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* **128**, 336–359 (2020).

86. U. Schubert & N. Hüsing. *Synthesis of Inorganic Materials*. (Wiley-VCH, 2019).

87. J. R. Chamorro & T. M. McQueen. Progress toward Solid State Synthesis by Design. *Acc Chem Res* **51**, 2918–2925 (2018).

88. C. J. Bartel. Review of computational approaches to predict the thermodynamic stability of inorganic solids. *J Mater Sci* **57**, 10475–10498 (2022).

89. A. Miura *et al.* Selective metathesis synthesis of MgCr2S4 by control of thermodynamic driving forces. *Mater Horiz* **7**, 1310–1316 (2020).

90. P. K. Todd *et al.* Selectivity in Yttrium Manganese Oxide Synthesis via Local Chemical Potentials in Hyperdimensional Phase Space. *J Am Chem Soc* **143**, 15185–15194.

91. A. Narayan *et al.* Computational and experimental investigation for new transition metal selenides and sulfides: The importance of experimental verification for stability. *Phys Rev B* **94**, 045105 (2016).
92. W. Sun *et al.* The thermodynamic scale of inorganic crystalline metastability. *Sci Adv* **2**, e160022 (2016).
93. G. P. Nagabhushana, R. Shivaramaiah, & A. Navrotsky. Direct calorimetric verification of thermodynamic instability of lead halide hybrid perovskites. *Proc Natl Acad Sci USA* **113**, 7717–7721 (2016).
94. Z. Li, K. G. Pradeep, Y. Deng, D. Raabe, & C. C. Tasan. Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off. *Nature* **534**, 227–230 (2106).
95. A. Stein, S. W. Keller, & T. E. Mallouk. Turning Down the Heat: Design and Mechanism in Solid-State Synthesis. *Science* **259**, 1558–1564 (1993).
96. D. P. Shoemaker *et al.* In situ studies of a platform for metastable inorganic crystal growth and materials discovery. *Proc Natl Acad Sci USA* **111**, 30 (2014).
97. A. S. Haynes, C. C. Stoumpos, H. Chen, D. Chica, & M. G. Kanatzidis. Panoramic Synthesis as an Effective Materials Discovery Tool: The System Cs/Sn/P/Se as a Test Case. *J Am Chem Soc* **139**, 10814–10821 (2017).
98. M. Bianchini *et al.* The interplay between thermodynamics and kinetics in the solid-state synthesis of layered oxides. *Nat Mater* **19**, 1088–1095 (2020).
99. H. Huo *et al.* Machine-Learning Rationalization and Prediction of Solid-State Synthesis Conditions. *Chem Mater* **34**, 7323–7336 (2022).
100. T. He *et al.* Inorganic synthesis recommendation by machine learning materials similarity from scientific literature. *Sci Adv* **9**, eadg818 (2023).
101. R. Merkle & J. Maier. On the Tamman-Rule. *Z Anorg Allg Chem* **631**, 1163–1166 (2005).
102. C. W. Coley, W. H. Green, & K. F. Jensen. Machine Learning in Computer-Aided Synthesis Planning. *Acc Chem Res* **51**, 1281–1289 (2018).
103. C. W. Coley, L. Rogers, W. H. Green, & K. F. Jensen. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Cent Sci* **3**, 1237–1245 (2017).
104. E. J. Corey. The Logic of Chemical Synthesis: Multistep Synthesis of Complex Carbogenic Molecules (Nobel Lecture). *Angew Chem Int Ed Engl* **30**, 455–612 (1991).
105. M. H. S. Segler, M. Preuss, & M. P. Waller. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
106. J. S. Schreck, C. W. Coley, & K. J. M. Bishop. Learning Retrosynthetic Planning through Simulated Experience. *ACS Cent Sci* **5**, 970–981.
107. X. Wang *et al.* Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning. *Chem Sci* **11**, 10959.
108. K. Kovnir. Predictive Synthesis. *Chem Mater* **33**, 4835–4841 (2021).
109. M. J. McDermott, S. S. Dwaraknath, & K. A. Persson. A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis. *Nat Commun* **12**, 3097 (2021).

110. M. Aykol, J. H. Montoya, & J. Hummelshøj. Rational Solid-State Synthesis Routes for Inorganic Materials. *J Am Chem Soc* **143**, 9244–9259 (2021).

111. A. G. Kusne *et al.* On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat Commun* **11**, 5966 (2020).

112. A. A. Volk & M. Abolhasani. Autonomous flow reactors for discovery and invention. *Trends Chem* **3**, 519–522 (2021).

113. H. Hayashi, K. Hayashi, K. Kouzai, A. Seko, & I. Tanaka. Recommender System of Successful Processing Conditions for New Compounds Based on a Parallel Experimental Data Set. *Chem Mater* **31**, 9984–9992 (2019).

114. C. J. Bartel *et al.* Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry. *Nat Commun* **9**, 4168 (2018).

115. D. A. Porter, K. E. Easterling, & M. Y. Sherif. *Phase Transformations in Metals and Alloys*. (CRC Press, Taylor & Francis Group, 2009).

116. M. Z. Bazant. Theory of Chemical Kinetics and Charge Transfer based on Nonequilibrium Thermodynamics. *Acc Chem Res* **46**, 1144–1160 (2013).

117. E. O. Chi, K. M. Ok, Y. Porter, & P. S. Halasyamani. $Na_2Te_3Mo_3O_{16}$: A New Molybdenum Tellurite with Second-Harmonic Generating and Pyroelectric Properties. *Chem Mater* **18**, 2070–2074 (2006).

118. H. Morimoto *et al.* Charge/discharge Behavior of Triclinic $LiTiOPO_4$ Anode Materials for Lithium Secondary Batteries. *Electrochem.* **84**, 878–881 (2016).

119. L. C. Pathak & S. K. Mishra. A review on the synthesis of Y–Ba–Cu-oxide powder. *Supercond Sci Technol* **18**, R67–R89 (2005).

120. B. D. Fahlman. Superconductor Synthesis — An Improvement. *J Chem Educ* **78**, 1182 (2001).

121. I. Arvanitidis, Du. Sichen, & S. Seetharaman. A Study of the Thermal Decomposition of BaCO3. *Met. Mater Trans B* **27**, 409–416 (1996).

122. A. Małecki, J. Obłąkowski, & S. Łabuś. The role of $BaCO_3$ in high temperature synthesis of electronic materials. *Mater Res Bull* **30**, 731–737 (1995).

123. P. F. de Aguiar, B. Bourguignon, M. S. Khots, D. L. Massart, & R. Phan-Than-Luu. D-optimal designs. *Chemom Intell Lab Syst* **30**, 199–210 (1995).

124. A. D. Clayton *et al.* Algorithms for the self-optimisation of chemical reactions. *React Chem Eng* **4**, 1545–1554 (2019).

125. H. Takeda *et al.* Process optimisation for NASICON-type solid electrolyte synthesis using a combination of experiments and bayesian optimisation. *Mater Adv* **3**, 8141 (2022).

126. W. Zhang, J. Sun, X. Wang, G. Shen, & D. Shen. Crystal growth and optical properties of a noncentrosymmetric molybdenum tellurite, $Na_2Te_3Mo_3O_{16}$. *CrystEngComm* **14**, 3490 (2012).

127. Y. Zeng *et al.* Selective formation of metastable polymorphs in solid-state synthesis. Preprint available at arXiv:2309.05800v1 (2023).

128. T. F. Malkowski *et al.* Role of Pairwise Reactions on the Synthesis of $Li_{0.3}La_{0.57}TiO_3$ and the Resulting Structure–Property Correlations. *Inorg Chem* **60**, 14831–14843 (2021).

129. G. Franceschini & S. Macchietto. Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci* **63**, 4846–4872 (2008).

130. B. Cao *et al.* How To Optimize Materials and Devices via Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics. *ACS Nano* **12**, 7434–7444 (2018).

131. B. J. Shields *et al.* Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).

132. A. M. K. Nambiar *et al.* Bayesian Optimization of Computer-Proposed Multistep Synthetic Routes on an Automated Robotic Flow Platform. *ACS Cent Sci* **8**, 825–836.

133. L. Ward *et al.* Matminer: An open source toolkit for materials data mining. *Comput Mater Sci* **152**, 60–69 (2018).

134. T. Xie & J. C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys Rev Lett* **120**, 145301 (2018).

135. C. Chen, W. Ye, Y. Xuo, C. Zheng, & S. P. Ong. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chem Mater* **31**, 3564–3572 (2019).

136. K. Choudhary & B. DeCost. Atomistic Line Graph Neural Network for improved materials property predictions. *Npj Comput Mater* **7**, 185 (2021).

137. J. R. González-Velasco, R. Ferret, R. López-Fonseca, & M. A. Gutiérrez-Ortiz. Influence of particle size distribution of precursor oxides on the synthesis of cordierite by solid-state reaction. *Powder Technol* **163**, 34–42 (2005).

138. G. E. Kamm *et al.* Relative Kinetics of Solid-State Reactions: The Role of Architecture in Controlling Reactivity. *J Am Chem Soc* **144**, 11975–11979 (2022).

139. P. J. Linstrom & W. G. Mallard. The NIST Chemistry WebBook: A Chemical Data Resource on the Internet. *J Chem Eng Data* **46**, 1059–1063 (2001).

140. N. J. Szymanski *et al.* Understanding the Fluorination of Disordered Rocksalt Cathodes through Rational Exploration of Synthesis Pathways. *Chem Mater* **34**, 7015–7028 (2022).

141. S. Patoux & M. M. Doeff. Direct synthesis of LiNi1/3Co1/3Mn1/3O2 from nitrate precursors. *Electrochem Commun* **6**, 767–772 (2004).

142. Ilia Geigman, N. G. Furmanova, P. G. Nagornyi, L. D. Yun, & M. V. Rotenfel'D. Crystal structure and V^{4+} EPR of the Li-Ti double oxyorthophosphate α-LiTiOPO4. *Crystallogr Rep* **38**, 759–762 (1993).

143. A. Jain, Y. Shin, & K. A. Persson. Computational predictions of energy materials using density functional theory. *Nat Rev Mater* **1**, 15004 (2016).

144. J. Sun *et al.* Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nat Chem* **8**, 831–836 (2016).

145. Z. Ren *et al.* Embedding physics domain knowledge into a Bayesian network enables layer-by-layer process innovation for photovoltaics. *Npj Comput Mater* **6**, 9 (2020).

146. J. Li *et al.* Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).

147. P. J. Kitson *et al.* Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals. *Science* **359**, 2018.

148. C. W. Coley *et al.* A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).

149. N. J. Szymanski, P. Nevatia, C. J. Bartel, Y. Zeng, & G. Ceder. Autonomous and dynamic precursor selection for solid-state materials synthesis. *Nat Commun* **In print**, https://doi.org/10.1038/s41467-023-42329-9 (2023).

150. N. J. Szymanski *et al.* Adaptively driven X-ray diffraction guided by machine learning for autonomous phase identification. *Npj Comput Mater* **9**, 31 (2023).

151. T. He *et al.* Similarity of Precursors in Solid-State Synthesis as Text-Mined from Scientific Literature. *Chem Mater* **32**, 7861–7873 (2020).

152. A. Pardo, J. Romero, & E. Ortiz. High-temperature behaviour of ammonium dihydrogen phosphate. *J Phys Conf Ser* **935**, 012050 (2017).

153. S. K. Gupta & Y. Mao. Recent Developments on Molten Salt Synthesis of Inorganic Nanomaterials: A Review. *J Phys Chem C* **125**, 6508–6533 (2021).

154. M. Aykol, S. S. Dwaraknath, W. Sun, & K. A. Persson. Thermodynamic limit for synthesis of metastable inorganic materials. *Sci Adv* **4**, eaaq014 (2018).

155. B. Bridge & N. D. Patel. The elastic constants and structure of the vitreous system Mo-P-O. *J Mater Sci* **21**, 1186–1205 (1986).

156. F. Muñoz & L. Sánchez-Muñoz. The glass-forming ability explained from local structural differences by NMR between glasses and crystals in alkali metaphosphates. *J Non Cryst Solids* **503–504**, 94–97 (2019).

157. P. Norby, I. G. Krogh Andersen, E. K. Andersen, & N. H. Andersen. The crystal structure of lanthanum manganate(iii), $LaMnO_3$, at room temperature and at 1273 K under N_2. *J Solid State Chem* **119**, 191–196.

158. Y.-J. Kim, H.-S. Park, & C.-H. Yang. Raman imaging of ferroelastically configurable Jahn–Teller domains in LaMnO3. *Npj Quantum Mater* **6**, 62 (2021).

159. J. A. Alonso *et al.* Non-stoichiometry, structural defects and properties of $LaMnO_{3+\delta}$ with high δ values ($0.11 \leq \delta \leq 0.29$). *J Mater Chem* **7**, 2139–2144 (1997).

160. O. Kononova *et al.* Text-mined dataset of inorganic materials synthesis recipes. *Sci Data* **6**, 203 (2019).

161. P. Villars, K. Cenzual, & R. Gladyshevskii. *Handbook of Inorganic Substances*. (De Gruyter, 2017).

162. B. H. Toby & R. B. von Dreele. GSAS-II: the genesis of a modern open-source all purpose crystallography software package. *J Appl Cryst* **46**, 544–549 (2013).

163. J. H. O'Donnell, R. B. Von Dreele, M. K. Y. Chan, & B. H. Toby. A scripting interface for GSAS-II. *J Appl Cryst* **51**, 1244–1250 (2018).