PHYSICAL REVIEW MATERIALS **1**, 063802 (2017)

Automated crystal structure solution from powder diffraction data: Validation of the first-principles-assisted structure solution method

Logan Ward,[*] Kyle Michel,[†] and Chris Wolverton[‡]
Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA
(Received 10 July 2017; published 27 November 2017)

High-throughput first-principles computational methods, such as density functional theory (DFT), offer the ability to predict the properties of materials, provided their crystal structures are known. However, there are many compounds for which the structure is unknown and, consequently, many potentially useful materials cannot be assessed by high-throughput DFT searches. Here, we demonstrate an automated tool to solve the structures of materials from powder diffraction patterns based on the first-principles-assisted structure solution (FPASS) method. We validate this tool by using it to solve 95 already-known crystal structures and find that FPASS can determine the correct structure in all cases. We then tuned FPASS to improve its performance on the most-difficult test cases, which include structures with larger numbers of symmetrically unique atoms. We also used FPASS to solve the structures of 10 materials and found, using DFT, several are interesting candidates for semiconductor applications.

DOI: 10.1103/PhysRevMaterials.1.063802

I. INTRODUCTION

Determining the crystal structure of a material is often the first step in being able to understand or predict its properties. In fact, crystal structure is the only requirement to predict the properties of a compound with density functional theory (DFT). Recent applications of high-throughput DFT demonstrate how it can be used to automatically screen promising compounds for many potential applications [1–7]. However, despite significant advances in techniques to solve crystal structures from powder diffraction data, there are many materials for which the structure is currently unknown. For example, there are thousands of diffraction patterns that are not associated with a crystal structure in the powder diffraction file (PDF) and tens of thousands of entries in the inorganic crystal structure database (ICSD) that are incomplete [8,9]. Beyond simply filling in gaps in scientific knowledge, solving these structures would significantly expand the databases of material properties computed using high-throughput DFT [3,5,10,11].

Reconstructing three-dimensional crystal structures from one-dimensional powder diffraction data is a nontrivial problem, even in an ideal case [12]. Conventionally, solving a crystal structure involves first determining the shape, symmetry, and content of the unit cell, and then performing the "structure solution" step to determine atomic positions [13]. A variety of techniques exist for performing structure solution, which all require varying degrees of expert judgement to employ [14]. Given the large number of unsolved structures, the solution by conventional means by experts is intractable. What would enable the solution of these crystals is a set of *automated* tools for crystal structure solution.

One potential tool for automated crystal structure solution is the first-principles-assisted structure solution (FPASS) method [15]. Like other direct-space crystal structure solution methods [14], FPASS uses a global optimization algorithm to find a crystal structure that best explains the experimental data. As FPASS is designed to solve crystal structures given composition, lattice parameters, and symmetry; it is well-suited to solve the many incomplete structures in the PDF for which these are already known. The other key feature of FPASS is that it uses energies calculated from DFT to search for structures with a combination of low DFT energy and an optimal match to experimental data. The combination of symmetry-constrained genetic algorithm and DFT energies employed by FPASS make it possible to easily solve structures that are difficult to determine with conventional techniques [15]. The inclusion of energy in structure solutions, in particular, has been shown to be beneficial in FPASS and other methods as well [13,16]. As DFT is used to compute the energy of candidate crystal structures, no problem-specific selection of an empirical potential is required—making it possible to run large numbers of FPASS calculations without any need to first validate an empirical potential [13,16–18]. All of these features suggest that FPASS would be a suitable tool for automated structure solution. However, to date, FPASS has only been tested in a few case studies [15,19,20] and no automated implementation of this method exists.

In this work, we present a new implementation of the FPASS method and demonstrate that it can be used to automatically solve crystal structures. Our method includes a modified version of the original genetic algorithm and an automated method for matching candidate solutions to raw diffraction patterns. To evaluate the performance of this method, we first validate the ability of the algorithm to solve 95 common crystal structures given the known diffraction pattern, unit cell, and symmetry—and found that FPASS correctly solves all of them. During these validation tests, we determined that crystal structures with large numbers of possible combinations of Wyckoff sites are difficult to solve with FPASS and show how the algorithm that be tuned to perform better on such cases. With this knowledge, we applied FPASS to dozens of unsolved crystal structures from the PDF and were able to solve a significant fraction of them automatically. We

[*]Present address: University of Chicago, Chicago, IL 60637.
[†]Present address: Citrine Informatics, Redwood City, CA 94063.
[‡]Corresponding author: c-wolverton@northwestern.edu

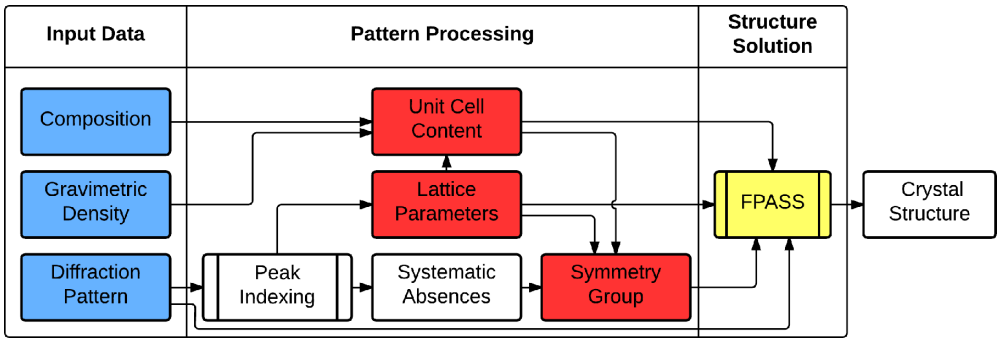

FIG. 1. Flow chart describing the process of solving a crystal structure from powder diffraction data using FPASS. The typical solution process starts by measuring the composition, diffraction pattern, and gravimetric density of a compound, which are then used to determine the unit cell contents, lattice parameters, and symmetry group of a crystal. Once this information is determined, FPASS is used to find the lowest-energy structure within these constraints.

then added these structures to the Open Quantum Materials Database (OQMD) [3,10] and used high-throughput DFT to predict their properties, and found several have band gap energies of approximately 1 eV—making them interesting for semiconductor applications. As our method requires minimal human interaction, we plan to continuously apply it to unsolved structures and then automatically predict the properties of these materials using high-throughput DFT.

II. METHODS

A general outline of how crystal structures are solved from powder diffraction data using FPASS is shown in Fig. 1. FPASS is used to determine atomic positions in a crystal after the lattice parameters, unit cell contents, and space group have been determined with peak indexing techniques. In this section, we will describe the theory and techniques behind FPASS. In particular, we will discuss the method used to compute the energy of a candidate structure, how our software determines how well a structure matches a powder diffraction pattern, and the details of the genetic algorithm used to efficiently solve the structure.

A. FPASS software

The FPASS algorithm is implemented as part of the materials interface (Mint) software [21], which performs all steps in FPASS automatically. Mint itself was designed to perform many common tasks in the atomic-scale simulation of materials, including symmetry determination and generating input files for various simulation packages. Consequently, many of the analyses required by FPASS (e.g., symmetry determination) are present in Mint. Mint is written in C++ and available freely under the LGPL license [21]. A practical example of running FPASS with Mint is given in the Supplemental Material [22].

B. Density functional theory energy of candidate structures

We use DFT to compute the energy of candidate structures, which enables the reliable prediction of energies without the need to fit or select appropriate empirical potentials. DFT only requires the crystal structure as input and is known to be able to reliably predict the ground state structure and formation energy of many inorganic systems [3,10,23,24]. As a result, DFT makes an excellent choice for a high-throughput solution tool—one can be confident in the accuracy of the calculated energies for a broad variety of crystal structures without needing to validate the energy calculation method before each solution.

For this work, we performed all DFT calculations using the Vienna *ab initio* Simulation Package (VASP) [25–28]. Unless otherwise stated, all calculations were performed with projector augmented-wave basis sets [29,30]: the GGA exchange-correlation functional of Perdew, Burke, and Ernzerhof [31], a cutoff energy of 1.3 times the maximum cutoff energy of all of the provided pseudopotentials, and 1000 K-points per reciprocal atom. We employed the DFT settings used by the OQMD, a collection of the structures and DFT-predicted formation energy of hundreds of thousands of crystalline materials, when comparing the stability of a candidate crystal structure against others [10].

C. Matching candidate structures against powder diffraction patterns

Another component of FPASS is methods to determine how well a proposed structure agrees with the experimental diffraction pattern. Comparing a structure to an XRD pattern requires methods for computing XRD patterns for hypothetical structures, processing experimentally-measured patterns, and adjusting structures to better match the diffraction pattern. Each of these techniques are detailed in the following subsections.

1. Diffraction pattern calculation

Techniques to calculate the diffraction pattern of a crystal structure are well-established in the crystallography community [32,33]. At a high level, the powder diffraction pattern is calculated by first finding all reflections that will occur within a certain range of diffraction angles. Next, reflections that would cancel each other out in a powder diffraction pattern due to the symmetry of the crystal are removed, and symmetrically identical peaks are grouped together for computational efficiency. These two steps generate a list of diffraction peaks that should be observed in a powder diffraction pattern and are automatically performed by Mint based on the unit cell and symmetry of

the structure. Once the list of peaks is generated, the intensity of each individual peak is computed based on the following relation [12]:

$$I_{hkl} = K \times m_{hkl} \times LP(\theta) \times T_{hkl} \times |F_{hkl}|, \quad (1)$$

where K is a scaling factor, m_{hkl} is the number of symmetrically equivalent planes corresponding to this peak, and $LP(\theta)$ is the Lorentz and polarization factors at the diffraction angle for a particular reflection, θ. Mint uses the version of this function for patterns that were created without a monochromater. T_{hkl} describes the effect of preferred grain orientations (i.e., texturing). We use the March-Dollase function, which describes texturing using the direction of preferred orientation and a single parameter to describe the magnitude [34]. $|F_{hkl}|^2$ is the structure factor component of the diffracted intensity. We account for thermal vibrations using a single isotropic thermal factor, B, for each symmetrically-distinct group of atoms

2. Raw diffraction pattern processing

Comparing a calculated diffraction pattern to an experimental pattern is easiest when comparing the integrated intensities of each peak. As the experimental data available for use may be the raw diffraction signal (intensity as a function of angle), we needed to implement an automated scheme for detecting the positions and integrating the intensities of each peak. As illustrated in Fig. 2, our implementation uses a version of the processing algorithm described by Pecharsky and Zavalij [12]:

(1) *Noise filter:* The raw x-ray pattern is first passed through a noise filter that smooths the data by averaging the diffracted intensities of points with similar diffraction angles.

(2) *Background removal:* First, the background signal determined by calculating a running average where every point in a 2-degree window is assigned a weight inversely proportional to the 4th power of the intensity at that point. This background signal is then subtracted from the smoothed data from step 1.

(3) *Peak detection:* Once the background has been removed, the locations of diffraction peaks are found by identifying local minima in the second derivative of intensity with respect to diffraction angle. By identifying peaks based on the second derivative, we can easily separate peaks that are slightly overlapped [12].

(4) *Intensity extraction:* A pseudo-Voight function, which is known to describe the peaks in x-ray diffraction patterns well [12], is fit to the intensity values for each peak. We then integrate the area under each function to determine the intensity of each diffraction peak. The fitting functions corresponding to overlapping peaks are fitted concurrently in order to accurately determine the contribution of each individual peak.

3. Matching structures to diffraction pattern

Before measuring how well a structure matches an experimental diffraction pattern, we adjust the candidate structures so that its computed pattern better matches the reference pattern—a technique known as structure refinement. Refinement is accomplished by minimizing the function:

$$R = \frac{\sum_{\text{peaks}} (I_{\text{calc}} - I_{\text{obs}})^2}{\sum_{\text{peaks}} I_{\text{obs}}}, \quad (2)$$

where I_{calc} is the calculated integrated intensity a single diffraction peak, and I_{obs} is the integrated intensity of the same peak in the experimental pattern. Both sums are over all peaks in the calculated and observed patterns. Each peak in the calculated pattern is assigned to the closest peak of any peak in the observed pattern that is within 0.15 degrees. If multiple calculated peaks match a single peak in the observed pattern, their intensities values are added together. Peaks from

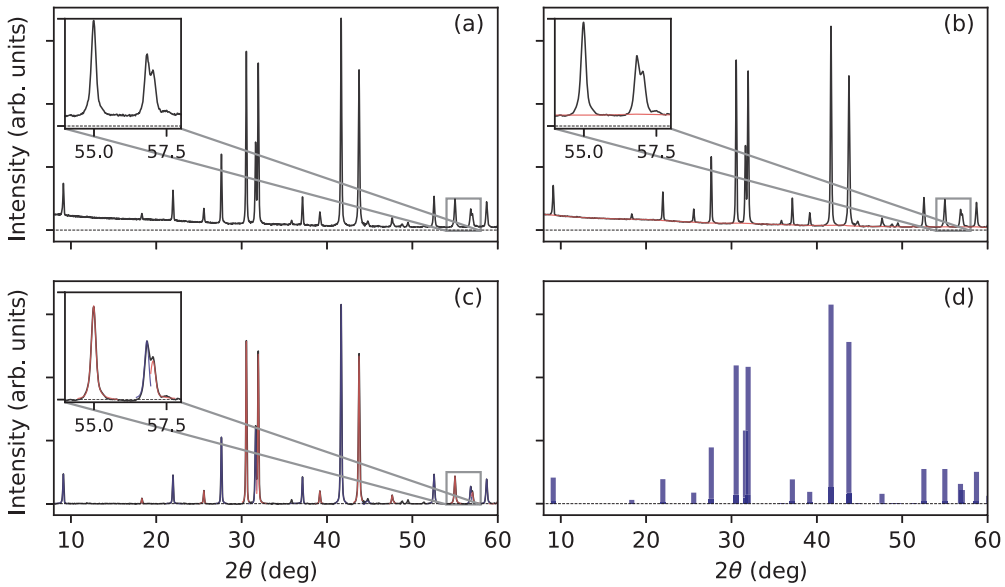

FIG. 2. Illustration of our method for automatically determining peak positions and intensities from powder x-ray diffraction (XRD) patterns. (a) The original XRD pattern. (b) The XRD pattern after noise removal, the automatically detected background signal is shown in red. (c) The XRD pattern with background removed and after peak detection. The pseudo-Voight functions fit to each detected peak are shown in red and blue. (d) Peak locations and integrated intensities.

both the calculated and observed pattern that are not able to be matched are treated as being matched to a hypothetical peak of zero intensity. The computed intensity, I_{calc}, is a function of the factors listed in Eq. (1) (e.g., atomic positions, thermal factors, texturing). We optimize all parameters using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, as implemented in Dlib [35]. Following the procedure described by Pecharsky and Zavalij [12], we sequentially add more fitting parameters (starting with the scale factor) during the optimization. After refinement, we use the optimized R to describe the match to the diffraction pattern.

Rietveld refinement. When matching the computed XRD pattern against the raw intensity measurements, we employ the Rietveld refinement method [36]. In contrast to the previous section where only the integrated intensities of each peak are considered, the computed x-ray pattern of a structure is compared to the entire, unprocessed experimental x-ray diffraction pattern (i.e., without any of the previously-described processing steps) in Rietveld refinement. Here, the match between two patterns is defined by

$$R = \sum_{2\Theta} |I_{\text{calc}} - (I_{\text{obs}} - I_{\text{background}})| \Big/ \sum_{2\Theta} |I_{\text{obs}} - I_{\text{background}}|, \tag{3}$$

where I_{obs} is the observed diffracted intensity, I_{calc} is the computed diffraction intensity, and $I_{\text{background}}$ is the background signal [37]. Unlike in Eq. (2), the intensities here are the measured intensity at a single diffraction angle, and not the integrated intensity of an entire peak. We describe the background signal using Chebyshev polynomials and the shape of each diffraction peak using pseudo-Voight functions. The parameters for the background and peak shape are fit in addition to those that describe peak intensity [i.e., those from Eq. (1)]. As this optimization problem is significantly costlier than when using only integrated peak intensities, we only employ Rietveld refinement when automated pattern processing has failed and also when reporting the match to diffraction data for a proposed structure (i.e., only for the final structure of an FPASS run and not during the structure solution process).

D. Efficiently locating the optimal crystal structure solution—The FPASS method

The FPASS method is based on a genetic algorithm designed to efficiently search through candidate crystal structures. FPASS requires the unit cell parameters and content (i.e., number of atoms of each type) as input, and can use the measured diffraction pattern, space group, and the known position of any atoms in the structure to guide and constrain the search. In particular, including the symmetry as a constraint can dramatically accelerate the solution process [15,19]. With this symmetry and/or diffraction information as constraints, the FPASS algorithm can be used to determine the positions of atoms that minimize the energy. Genetic algorithms, in general, work by mimicking natural selection: better-performing solutions are mixed to create new candidates that are similar to them. As shown in Fig. 3, this process is repeated for several generations until the algorithm converges on an optimal solution. In the following sections, we will describe the two parts of the genetic algorithm that are specific to our implementation of the FPASS method: how the initial population is generated and how new generations are created. We note that these operations are similar to, but not exactly equivalent to those of the original implementation of FPASS [15].

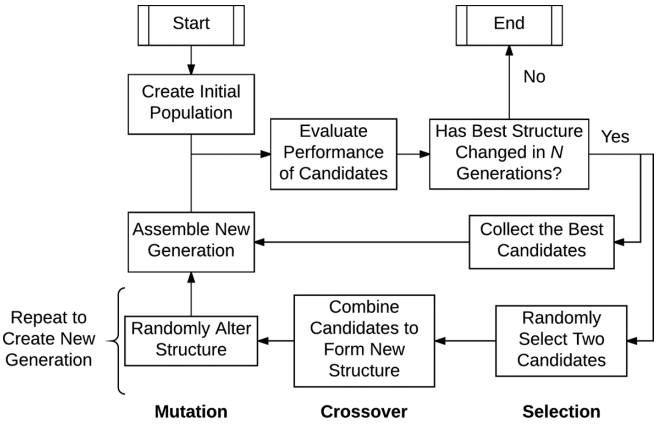

FIG. 3. Flowchart for genetic algorithm used by FPASS. The algorithm starts by generating a random population of candidate crystal structures, and then evaluating their properties. After the initial population, new generations are created by a mixture of the best-performing compounds from the previous generation and compounds created using genetic operators. These new structures are then evaluated, and the process is repeated until the best structure does not change after N generations.

1. Generating initial population

The genetic algorithm used in our implementation of FPASS starts with an initial population of randomly-generated structures. Each initial structure is created by first selecting a combination of Wyckoff sites from the known space group that leads to a unit cell with the correct numbers of atoms. Then, for each Wyckoff site, we assign random positions to each free parameter of that site. This process is repeated until we have the desired number of structures.

To more efficiently search the structure space, we bias the selection of the Wyckoff sites to be statistically similar to information mined from a large number of known crystal structures. We first determined number of Wyckoff sites and minimum possible number of Wyckoff sites consistent with space group and number of atoms for 74680 structures the Inorganic Crystal Structure Database [8]. Of these, 38% of the structures have only one possible number of sites. Of the 50820 structures that have >1 possible number of sites, the majority (56%) have the minimum possible number of Wyckoff sites, as shown in Fig. 4. See the Supplemental Material for further details [22]. To bias our GA to search mostly structures with small numbers of Wyckoff sites when generating new structures, we first generate a list of all possible combinations of Wyckoff sites that will lead to the correct number of atoms in the unit cell, provided the symmetry group given as input to FPASS. Each of those possibilities is assigned a weight related to the fraction of structures in the ICSD with the same ratio between the number of Wyckoff sites in that crystal to the fewest-possible number of sites. We adjust the

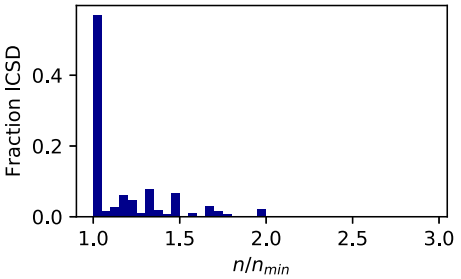

FIG. 4. Distribution of 74680 structures in the ICSD based on the ratio between number of Wyckoff sites (n) in the crystal structure and the minimum possible number of Wyckoff sites (n_0) consistent with the space group and size of the unit cell. Many structures (56%) have the minimum number of possible Wyckoff sites. This analysis only includes structures with more than one possible number of sites.

weights with a single biasing factor that varies between 0 (no biasing) and 1 (weight for a site is equal to the proportion of structures in the ICSD with that ratio). These weights influence the selection of combinations of sites when creating random crystal structures, and can be tuned using one of the input parameters for the genetic algorithm.

2. *Creating new generations*

Each new generation in FPASS is generated from a combination of: (i) the best-performing structures in the previous generation (elitism), and (ii) structures created using genetic operations. Before creating the new generation, we first rank each structure in the previous generation based on the sum of its rank in both energy and match to diffraction pattern (e.g., the structure with the lowest energy and second-best match to the pattern would have a score of 3). The entries with the best score are included in the new generation to ensure the best-performing structure is always considered when making the following generations.

To create the rest of the generation, we generate structures using genetic operations. For each new structure, we first select two parent structures out of the list of better-performing structures from the previous generation. To bias our selection towards better-performing candidates, we select parents by first randomly selecting $0.6\sqrt{n}$ entries from the previous generation, where n is the number of entries in the population, and then selecting the candidate from that list with the highest fitness. Each time we generate a new structure, we repeat this tournament selection process until two different parent structures are selected, then generate a new structure with crossover, and, finally, randomly perturb that structure.

Crossover is performed by combining groups of symmetrically-equivalent atoms (i.e., both the Wyckoff site's identity and free parameters) from either parent. For example, suppose one parent structure has a total of eight Na atoms on two different 4c Wyckoff positions and the second parent has four Na atoms on a 4c position, two atoms on a 2b, and two on a 2a position. Our crossover method could produce a child structure that includes one of the two groups of Na atoms on the 4c position from the first parent and the atoms on the 4c position from the second parent. Or, it could generate structure that contains the second group of Na atoms on the 4c position in the first parent and the atoms on the 2a and 2b positions in the second. In total, there are six possible ways of combining the Na atoms from the two parents that will have the same total number of Na atoms for this example. We randomly select exactly one of these possibilities. This procedure is repeated for each type of atom and will create a new structure with the same space group and composition as the parent structures.

After crossover, for randomly selected children, we perform up to two different mutations: (1) perturbing the atomic positions of a group of symmetrically equivalent atoms, or (2) selecting a different combination of Wyckoff sites and assigning random values to the free parameters of each new site. Both operations are designed to preserve the original symmetry of the structure. The probability of performing either type of mutation is adjustable, and the random magnitude of each perturbation allows the mutations to range from small alterations to completely random structures. It is also important to note that none of these mutations or crossover operations affect the lattice parameters of the structure. We implicitly assume that these lattice parameters were correctly determined during the peak indexing step.

An important caveat for structure creation is that we automatically screen out structures with exceptionally small distances between atoms. Here, we define "exceptionally small" distances as those where two atoms are closer than 50% of the sum of their radii (using radii from Codero *et al.* [38]). To enforce this constraint, we will not accept structures with short bond distances when generating the initial population, performing crossover, or mutating structures. Our exception to this rule is when our software fails to generate a valid structure after 100 attempts. For these rare cases, we use the structure from those 100 attempts that violates our distance rule the least.

III. RESULTS: TESTING VALIDITY AND IMPROVING EFFICIENCY OF FPASS

Before using FPASS to solve unknown structures, we first determined whether FPASS can reliably find the correct crystal structures across a broad variety of test cases. We also adjusted the settings of the algorithm to minimize the computational cost. The results from our validation and tuning efforts are described below.

A. Validating the algorithm

While the FPASS method has been shown to be able to accurately determine the crystal structures in a few test cases [15,19,20], we further validated our new implementation before using it in an automated manner. We determined how often FPASS finds the correct structure for 95 test compounds with a wide variety of symmetry groups and cell sizes. As shown in Fig. 5, we selected compounds from all crystal symmetry families except triclinic and with sizes ranging from between 1 and 30 atoms in the primitive cell. This wide variety of structures also enabled us to study how symmetry and unit cell size affect the performance of FPASS.

For each test case, we ran at least ten individual FPASS calculations to determine how often the algorithm finds the correct structure. In each test, FPASS was supplied with the

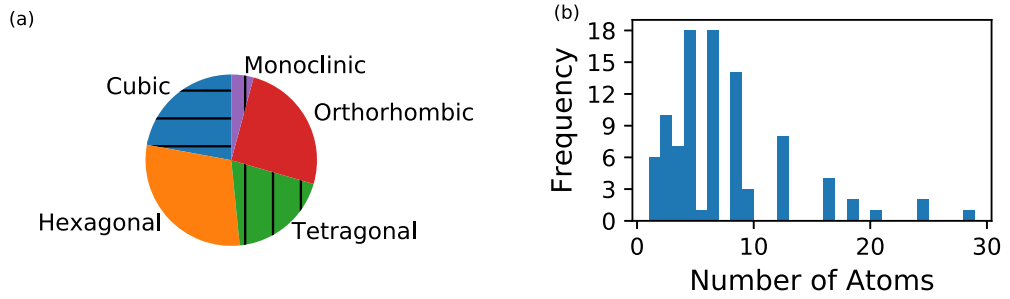

FIG. 5. (a) Distribution of crystal families and (b) number of atoms in the primitive cell for all 95 compounds used to validate FPASS. Candidates were intentionally chosen to sample a wide variety of cell sizes, sizes, and chemistries.

lattice parameters, symmetry group, powder diffraction pattern from the PDF, and the number of atoms of each type in the unit cell. In all test cases, we found that the lowest-energy candidate structure for that compound agreed with the known structure. In other words, FPASS found the correct structure in all 95 test cases. But, we also analyze the "success rate" of the FPASS runs: for each test structure, how many of the runs produced the correct solution. As shown in Fig. 6(a), we found that FPASS returns the correct structure in at least 90% of the runs for 67 (71%) of the 95 test structures. In several of these test cases, the high success rate of FPASS is not surprising because there were fewer than five structures that match the known space group and number of atoms. Even so, FPASS was still able to determine the correct structure of the 29-atom primitive cell of α-Mn 11 out of 15 times.

We also studied the characteristics of structures that correspond to especially low and high success rates. As shown in Fig. 6, the success rate correlates with the quality of the diffraction pattern, and the success rate decreases with larger minimum possible number of symmetrically unique atoms, and larger difference in the maximum and minimum number of Wyckoff sites. In fact, the compound that FPASS solved correctly the least often, $MgNi_2$, has a large number of minimum number of unique atoms (4), a large difference between the maximum and minimum number of unique atoms (4), and a poor-quality x-ray pattern (a "B" rating from the ICDD). Many of the other structures with low FPASS success rates have similar characteristics. In general, we found that FPASS performs best when solving structures with high symmetry and small number of atoms in the unit cell and is

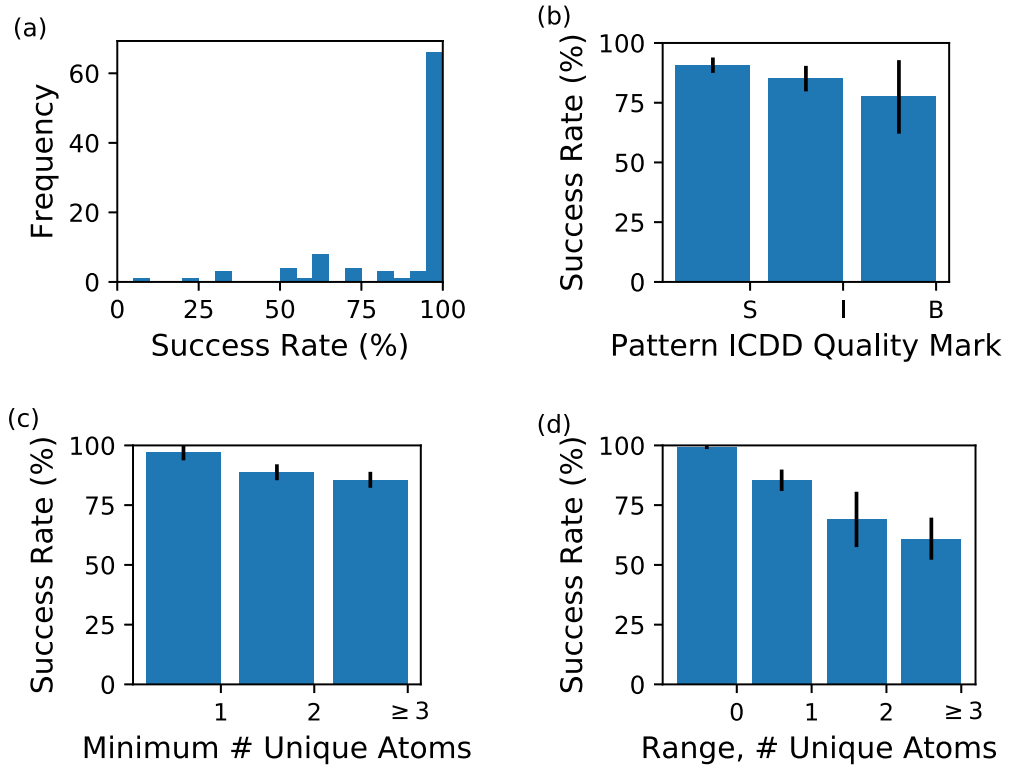

FIG. 6. (a) Histogram of how often FPASS determines the correct structure for all 95 test cases. Note that FPASS correctly solves all structures (success rate > 0) for all 95 cases. (b)–(d) Variation in how often FPASS determines the correct structure as a function of (b) quality of diffraction pattern, (c) minimum possible number of unique atoms, and (d) difference between maximum and minimum number of unique atoms. Generally, the success rate for FPASS is the worst when poor-quality x-ray data is provided and for crystals with large unit cells.

provided with a high-quality x-ray pattern. However, even with these caveats, we note that ten FPASS runs was sufficient in every case to correctly solve the structure.

B. Tuning the algorithm

To improve the reliability of FPASS for difficult-to-solve crystal structures, we adjusted parameters of the genetic algorithm and evaluated changes in success rate for ten of the more difficult test cases. For this process, we selected a set of problems we determined to be more difficult based on the metrics studied in the previous section: α-Mn, β-SiO_2, Mg_2Ni, CdI_2, GeS_2, PdS, CuS, NiS, α-Np, and SiU_3. FPASS had high rates of determining the correct structure for four of these structures (α-Mn, GeS_2, PdS, NiS), even though they were expected to be difficult based on the metrics established in the previous section. For the other six test cases, FPASS had low success rates. By selecting cases with a variety of success rates for FPASS, we can ensure that changes in the parameters that improve the performance on difficult cases do not negatively affect other cases.

We first tuned factors that do not directly affect the computation time: the Wyckoff-site biasing factor and mutation probabilities. As described in the Methods section, these parameters correspond to how much we bias our initial population and how new generations are created. Originally, we used a Wyckoff site and atomic position mutation probabilities of 0.5 and a biasing parameter of 0.5. To tune these parameters, we first adjusted the mutation probabilities and biasing parameter and found that the biasing parameter had the strongest impact on success rate. We then held all mutation parameters fixed, and found that the performance of FPASS was the best with the biasing turned off (i.e., a parameter of 0). We then held the bias factor fixed at 0, and repeated the tuning process to adjust the mutation probabilities and found that the original selections of 0.5 for each type of mutation were optimal. By adjusting mutation probabilities and biasing, we were only able to increase the average success rate of FPASS slightly from 64% of these difficult cases to 68%.

We studied the changes in success rates for individual test cases in order to better understand the effect of the Wyckoff site biasing. We found that the biasing factor can indeed have a significant positive effect in cases where the known structure has the fewest-possible number of Wyckoff sites. For example, the success rate for solving SiU_3 increases from 60 $\pm$ 9% to 74 $\pm$ 9% when the biasing factor is increased from 0 to 0.5. Conversely, the success rates for solving compounds with the maximum-possible number of Wyckoff sites are generally worsened by increasing biasing parameters. However, there are cases where biasing does not have the anticipated effect. For example, the success rate for α-Np with a biasing factor of 0.5 is 60 $\pm$ 11% and 74 $\pm$ 7%, even though α-Np has the fewest-possible number of sites. We hypothesize that this unexpected decrease in success rate could be explained by biasing reducing the diversity in the structures being assessed, which is detrimental to the effectiveness of the GA [39,40]. Alternatively, the observed decreases in success rate could be explained by the positive effect of biasing being smaller than the inherent randomness of FPASS calculations. Regardless of the cause, the results from our tuning suggests that, on the whole, Wyckoff site biasing leads to lower success rates (64 $\pm$ 3% with a biasing of 0.5 vs 68 $\pm$ 2% without biasing) on the more difficult test cases and, therefore, should not be used.

Once we finished tuning the mutation probabilities and biasing factor, we iteratively increased the population size. As the population size directly controls the calculation time, we also considered computational efficiency when adjusting this parameter. We found that by increasing the population size from 10 to 20 we could increase the success rate to 82%. Achieving a one part in a million chance of not finding the correct structure in at least one calculation would require 12 calculations for a success rate of 68% and only 8 for an 82% success rate. However, this increase in population size increases the total time for each calculation at a faster rate. Considering that the average FPASS calculation for a population size of 10 evaluated only 119 structures before converging and the average for a generation size of 20 was 219, the population size of 20 would require evaluating more structures and, thereby, more resources to achieve a certain likelihood of finding the correct structure. For that reason, we recommend a population size of 10 and running large numbers of FPASS calculations for more-difficult solutions. A summary of our recommended values for each parameter are shown in Table I.

IV. RESULTS: AUTOMATED SOLUTION OF CRYSTAL STRUCTURES FROM FPASS

After validating and tuning FPASS, we employed it to solve the structures of several entries that lacked crystal structures in both the Powder Diffraction File and OQMD. We selected a total of 20 compounds from the PDF that had both high-quality diffraction patterns and unit cell parameters, but lacked atomic positions. We preferentially selected compounds with small numbers of atoms per unit cell.

To further automate the solution process, we created a software package designed to automate starting FPASS calculations, checking output for errors, and performing several validation checks. This automation software, named "fpassmgr," is available under an open-source license [41]. Using fpassmgr, we run FPASS at least ten times and until at least five FPASS results are identical to help ensure we capture the ground state. When then relax the candidate solution with DFT and compute the stability and volume with qmpy [10]. Once the calculation is complete, the code generates a webpage summary that includes structure files for the proposed solution and all of the validation results. For most cases, the only human interaction required is starting the calculation and reviewing the validation summary in order to decide whether FPASS has found a correct solution.

A. Strategy to validate structure solutions

Our validation strategy is based on several different tests: (1) agreement between several independent FPASS calculations, (2) match to experimental diffraction pattern, (3) energetic stability, and (4) difference between the experimentally measured and DFT-relaxed volume. First, we run FPASS at least ten times and conclude a solution has been found when at least five calculations agree with the lowest-energy structure. Once we reach this level of agreement, we select

TABLE I. Adjustable parameters for the FPASS algorithm and their recommended values.

Name	Description	Recommendation
gaoptPopSize	Size of population	10
gaoptConverge	Number of generations after which if no better structure is found, the optimization is terminated	7
wyckoffBias	The biasing factor used when selecting new combinations in new Wyckoff site combinations. Larger values of this parameter bias selection towards fewer Wyckoff sites	0
gaoptWyckMutProb	Probability that a new structure will be mutated by selecting new Wyckoff sites	50%
gaoptPosMutProb	Probability that a new structure will be mutated by perturbing atomic positions	50%
gaoptNumToKeep	Number of top entries to retain in new generation	1

the lowest-energy solution as the putative structure, refine the structure to best match the experimental pattern, and measure the R factor. As our Rietveld refinement code is simple, we assume an R of below 0.5 is a satisfactory match (typically, R can be refined to below 0.05).

We also validate the structure by assessing the DFT-computed formation enthalpy and equilibrium volume of the candidate structure. To assess energetic feasibility, we compute the stability with respect to decomposition to all other phases in the OQMD [42]. We then compare this stability to that of all structures from the 36566 inorganic crystal structure database in the OQMD, and determine whether it is within below 2 interquartile ranges above the median (below 128 meV/atom)—as shown in Fig. 7(a). While a low-stability value is no guarantee that our solution is the ground state, it does verify that it is energetically feasible. Once we have computed the formation enthalpy, we also compute the change in volume during relaxation and compare the fractional change to the fractional changes of all compounds in the ICSD [see Fig. 7(b)]. We flag structures as potentially incorrect if the volume change is outside of 2 interquartile ranges from the median (below −8.1% or above 7.9%).

B. Solved structures

The following subsections are descriptions of ten structures we were able to solve using FPASS. The complete structures of some of these compounds had been solved previously but were not present in the OQMD, which made them ideal candidates for further testing of our solution and structure validation strategy. For other compounds, like $NaTmMo_2O_8$ and $LiSbO_3$, the unit cell and symmetry of the structure have been determined, but the atomic positions were not known. In all cases, the solution of the structures with FPASS enabled adding these materials to the OQMD. The full set of computed properties for these materials will be made available in a future public release of the OQMD.

As shown in the summary table, Table II, and Fig. 9, the compounds who structures we solved sample a broad variety of types of materials. Two of these materials are polyanionic compounds (CaCoSO and $Tb_2O_2CN_2$), one is a lithium-containing oxide ($LiSbO_3$), and two are intermetallics ($CeAl_3Pt$ and $Al_3FeGe_2Y_3$). Once we solved the structures and added these compounds to the OQMD, we found that several of the materials are semiconductors with band gap energies in the desired range for photovoltaics or thermoelectrics (Pb_2ZnTeO_6, CaCoSO, Ba_2CdTeO_6). Complete information about the solved structures (e.g., atomic coordinates and computed diffraction patterns) can be found in the Supplemental Material [22].

Solving these structures also provided an opportunity to study the performance of FPASS and to analyze our validation strategy. In terms of performance, we found that the average solution required evaluating 1823 structures and consumed

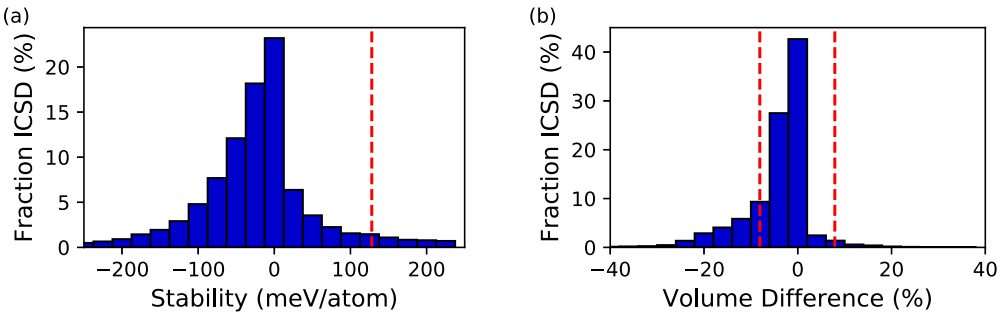

FIG. 7. Distributions of (a) DFT-computed stability and (b) fractional difference between the measured and DFT-computed volume of all 36565 compounds from the Inorganic crystal structure database (ICSD) in the OQMD. Stability was measured as the difference between the computed formation enthalpy of a compound and the minimum-energy combination of all other phases in the OQMD at the same composition. The red, dashed lines indicate two interquartile ranges away from the median, which the threshold used to determine whether a candidate structure solution should be carefully reviewed as potentially incorrect.

TABLE II. Compositions, symmetry groups, and DFT-predicted properties of structures solved using our automated implementation of FPASS. Stability is computed with reference to all other compounds in the OQMD [3,10]. Negative stability indicates that a compound is stable against decomposition into other structures.

Composition	Space group	Volume (Å^3/atom)			ΔH_f (eV/atom)	Stability (meV/atom)	E_g (eV)
		Experiment	DFT	δ (%)			
Al_3CePt	$I4mm$	19.84	20.44	3.0	−0.840	− 32	0
Pb_2ZnTeO_6	$Fm\bar{3}m$	12.78	12.85	0.5	−1.375	70	1.3
Sr_2TaZnO_6	$Fm\bar{3}m$	13.36	12.79	− 4.2	−2.721	58	0
CaCoSO	$P6_3mc$	17.02	16.99	− 0.2	−1.872	46	1.1
$Al_3FeGe_2Y_3$	$P\bar{6}2c$	19.46	19.47	0.1	−0.686	− 20	0
Ba_2CdTeO_6	$Fm\bar{3}m$	14.61	14.73	0.8	−2.159	− 100	1.2
KFe_2Se_2	$I4/mmm$	21.44	20.80	− 3.0	−0.504	31	0
Mo_2NaTmO_8	$I\bar{4}$	12.55	12.24	− 2.5	−2.663	− 55	3.1
$LiSbO_3$	$C2/m$	11.55	11.47	− 0.7	−1.954	1	2.9
$Tb_2O_2CN_2$	$P\bar{3}m1$	14.25	14.03	− 1.5	−2.175	− 206	4.1

5000 h of CPU time (we used a cluster with 3.0 GHz Intel Xeon 5160 s). We also found that the compute times vary significantly between structures. Many FPASS solutions evaluated only the minimum number of structures to achieve convergence (120, given our settings), which means FPASS found the best solution in the first generation of the GA. In general, we found that crystals with small, high-symmetry unit cells (e.g., Ba_2CdTeO_6) completed the most quickly, and that larger structures with low symmetry (e.g., $LiSbO_3$) required the most resources to solve.

Our validation strategy was successful in that there was a clear distinction between solutions we deemed to be correct and those we rejected. As shown in Fig. 8, we concluded all structures with acceptable pattern matches (<0.5) have stabilities below 100 meV/atom to be correct after further analysis (described below). Solutions with diffraction pattern matches near 0.6, on the other hand, are ambiguous. One solution, CaCoSO, has a poor pattern match of 0.65, yet we found our solution to match the experimentally reported structure (see discussion below). On the other hand, $ThAl_2Ni_3$ has a slightly better pattern match (0.62), but we label it is an incorrect solution as we are not sufficiently confident in the match to the pattern. The five structures that failed our volume change criterion also failed the stability and pattern match

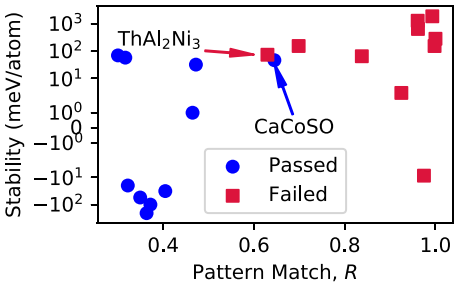

FIG. 8. Stability measured with respect to all other compounds in the OQMD and match to diffraction pattern data for the solutions to 20 structures we attempted to solve in this work. Blue circles indicate structures that passed our validation criteria and we deemed to be correctly solved. Red squares indicate structures that we determined FPASS failed to solve.

checks, which suggests that the volume change criterion is not particularly useful. Our main conclusion from these findings is that we should improve our Rietveld refinement software to be able to more clearly decide whether solutions match available XRD data.

1. $Al_3FeGe_2Y_3$

The crystal structure for $Al_3FeGe_2Y_3$ is not reported in the literature, and we could not even find reports of its synthesis in the literature [43]. As a starting point for our solution, we used the diffraction pattern from the PDF, along with the composition, space group, and lattice parameters. Using FPASS, we solved this structure and found this compound to be isostructural to the chemically similar compound, $Al_3NiGe_2Y_3$ [44]. We repeated the FPASS solution ten times and found that each repetition returned the same structure. The stability (measured with respect to all other competing phases) is negative, which indicates that it is thermodynamically stable at $T = 0$ K. The fractional difference between the experimentally determined and DFT-predicted volume of this structure is well within the distribution of other structures from the ICSD, which also indicates the structure is accurate. Finally, as shown in Fig. 10, the agreement between the computed and measured powder diffraction pattern is qualitatively excellent. Overall, these validation tests suggest our structure is likely the correct solution for $Al_3FeGe_2Y_3$.

2. Al_3CePt

Al_3CePt was originally synthesized in 1994, is known to have the $BaNiSn_3$-type crystal structure [45,46], and our FPASS calculation also finds this structure. While this structure has been solved before, the solution was not present in the Powder Diffraction File when we performed FPASS and is currently listed only under the $ThCr_2Si_2$ structure type in the ICSD. Both $ThCr_2Si_2$ and $BaNiSn_3$ are based on the $BaAl_4$ structure (with Ce in the body center position), but differ in the fact that $ThCr_2Si_2$ is centrosymmetric and $BaNiSn_3$ is not [46]. As in the solution for $Al_3FeGe_2Y_3$, we found the same structure in all ten FPASS calculations, the structure is stable in DFT, and that the DFT and experimental volumes of the structure are in agreement. Also considering the acceptable

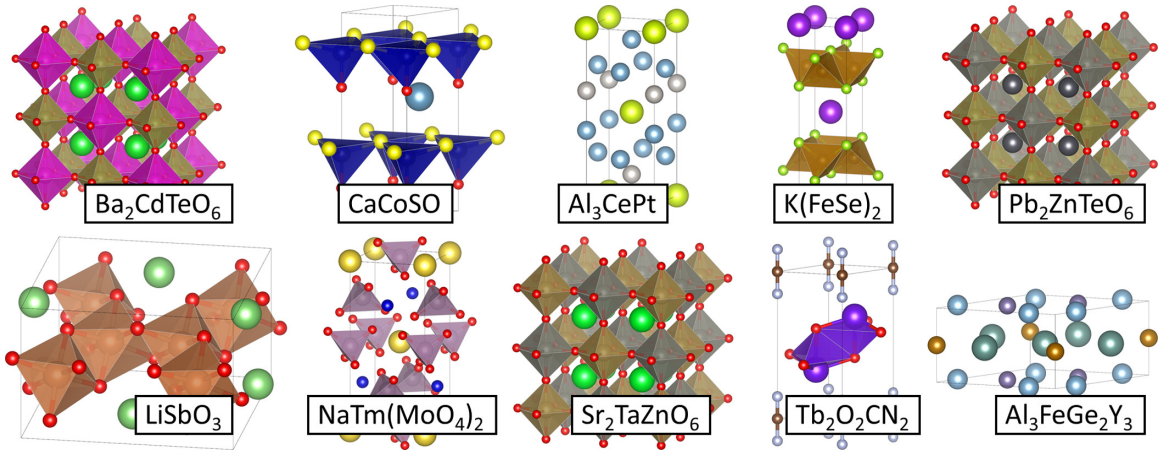

FIG. 9. Crystal structures determined in this work using an automated implementation of the FPASS algorithm.

match to the x-ray diffraction pattern (see Supplemental Material [22]), we conclude our solution is correct. While our work is not the first solution for this structure, our solution does match the literature and we were still successful in adding a missing crystal structure to the OQMD.

3. Pb_2ZnTeO_6, Ba_2CdTeO_6, *and* Sr_2TaZnO_6

We found Pb_2ZnTeO_6, Ba_2CdTeO_6, and Sr_2TaZnO_6 to all have the double perovskite structure. Of these, Pb_2ZnTeO_6 and Ba_2CdTeO_6 are both known to be double perovskites [47,48], but were not present in the ICSD and OQMD. As far as we could tell, Sr_2TaZnO_6 has not yet been reported in the literature and is not present in the ICSD or Crystallography Open Database (COD) [8,49]. For all three cases, FPASS found the same structure ten out of ten calculations, and all of our other validation checks indicate these structures are reasonable. Each structure was either stable or slightly metastable (<75 meV/atom) and the volume difference between the experimental were all within bounds (<5%)—as shown in Table II. According to our FPASS calculations, there are only two possible structures given the $A_2B'B''O_6$ stoichiometry, the known symmetry ($Fm\bar{3}m$), and the number of atoms in the unit cell: a structure where each A atom is coordinated with 12 O atoms, and one where A is coordinated with six. As the only difference between these two structures is the positions of O atoms, these two structures have the same symmetry and have similar diffraction patterns when O scatters x-rays weakly compared to other atoms in the structure. However, the energy difference between the structures can be quite large (>1 eV/atom for Pb_2ZnTeO_6), which makes the solution unambiguous with FPASS and shows the value of adding energetic validation to the structure solution problem.

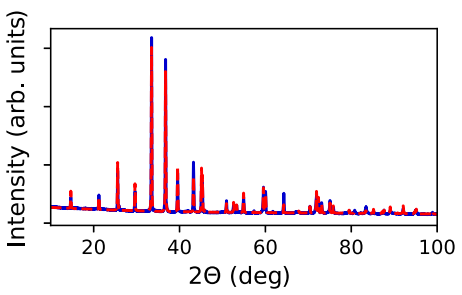

FIG. 10. Calculated (red, dashed line) and measured (blue, solid line) powder diffraction patterns of our proposed solution for the structure $Al_3NiGe_2Y_3$, as calculated using the materials interface (Mint).

4. CaCoSO

According to its entry in the powder diffraction file, CaCoSO is a hexagonal structure with a space group of $P6_3mc$. Using this information and the unit cell parameters available in the PDF, we found the structure to be similar to that of CaClOH, with Co occupying the tetrahedrally coordinated site [50]. This finding is consistent with the very recent structure solution of Salter *et al.* in 2016 [51]. We were unaware of this recent report when we solved the structure with FPASS. We found this structure in nine out of ten FPASS calculations, and we were able to confirm that it is only slightly metastable (46 meV/atom). Therefore, we agree with the structure proposed by Salter *et al.*

5. KFe_2Se_2

KFe_2Se_2 is the nominal composition of a superconducting compound discovered in 2010 and is known to have the $ThCr_2Si_2$ structure [52]. At the time we performed our FPASS calculation, this structure was not available in the OQMD but already established in the literature [52]. We did confirm the $ThCr_2Si_2$ structure with our FPASS calculation and found it to pass all validation checks. As in the solution of Al_3CePt, we did not solve this structure for the first time but were able to improve the OQMD by adding this structure to our database.

6. Mo_2NaTmO_8

We found the first report of Mo_2NaTmO_8 to be in a 1964 paper by Ayala *et al.* [53]. The authors reported the unit cell parameters of the structure and that Mo_2NaTmO_8 was based on the scheelite ($CaWO_4$) structure, but not the atomic positions [54]. Using FPASS, we confirmed the structure of Mo_2NaTmO_8 is based on the scheelite structure and equivalent

to that of Li_2CaHfF_8 [53]. FPASS found this structure in ten out of ten calculations. Our validation DFT calculations found a similar volume to that observed in experiment and that this structure is stable. Considering that the diffraction pattern match is also acceptable, we conclude we have determined the correct structure.

7. $LiSbO_3$

The monoclinic phase for $LiSbO_3$ was discovered by Nalbandyan *et al.* in 2006 [55]. While the authors were unable to determine its crystal structure, they hypothesized $LiSbO_3$ is a distorted rock salt structure [55]. Using the known unit cell parameters and symmetry group ($C2/m$) proposed by Nalbandyan *et al.* as input to FPASS, we found a layered structure similar whose network of Sb-O bonds is similar to the Mn-O network in Li_2MnO_3—a distorted rocksalt structure consistent with the hypothesis of Nalbandyan *et al.* [56]. The Sb atoms form a 2D network of face sharing octahedra, with the Li atoms occupying tetrahedral sites in the space between these layers (as opposed to octahedral sites in Li_2MnO_3). This structure is nearly degenerate with the known, orthorhombic phase of $LiSbO_3$ [57], being only 1 meV/atom higher in energy than the orthorhombic structure according to our DFT calculations. We ran FPASS 30 times and found the layered structure in 8 of the calculations. The other, higher energy solutions, include a version of this structure where the Li is in octahedral sites (as in Li_2MnO_3), which is slightly higher in energy (42 meV/atom). Also considering the acceptable match with the experimental diffraction pattern and small difference between experimental and DFT-predicted volumes, we conclude our solution to $LiSbO_3$—a layered structure with Li on the tetrahedral sites—is correct.

8. $Tb_2CN_2O_2$

According to the powder diffraction file, $Tb_2CN_2O_2$ has a hexagonal unit cell with $P\bar{3}m1$ symmetry. Starting with this information and the lattice parameters listed in the PDF, we found that this compound shares the same crystal structure as 11 other lanthanoid dioxymonocyanamides [58,59]. All ten FPASS calculations we performed found this structure, and we also found it to be stable in the OQMD. The volume change on relaxation and match to diffraction pattern are also reasonable, which lead us to conclude that this is the correct structure for $Tb_2CN_2O_2$.

C. Current limitations of FPASS

We just described ten successful examples of FPASS solutions. However, for ten other compounds we attempted, FPASS failed at least one of the validation tests, each of which could be a result of several factors. First of all, our technique is based on the assumption that the hypothesized lattice parameters and space group are correct. If any of these are inaccurate, our algorithm may converge to an incorrect solution. Also, FPASS may have failed to find the correct solution within the search space—though this is unlikely if the algorithm returned the same structure from multiple runs. Furthermore, our implementation of FPASS currently supports only perfectly ordered materials. If the true solution contains partially occupied sites (e.g., mixing between two types of elements on a single site), FPASS will fail to find the correct structure. Overall, these failures highlight potential avenues for improving FPASS (e.g., accounting for disorder) and the necessity of automated validation tests.

A potential route for accelerating the solution process is to utilize already-known crystal structure types in the search process. Such a route seems especially promising when considering that nine out of ten of our solutions are isostructural with another material. Integrating structure prototypes into our automated tool could be accomplished several ways. The first step to any of these approaches would be to search the ICSD or another database for structures that match the stoichiometry and symmetry of the structure being solved. One route to integrating this information into FPASS is including these candidate structures into the initial population of the genetic algorithm, thereby biasing it toward more-feasible solutions. Alternatively, we could check whether any of the prototype structures are valid solutions before employing FPASS. Hautier *et al.* were able to solve the structures of approximately 200 ternary oxides from the PDF using only known structures, so this approach could be successful for other classes of materials [60]. By combining information about already-known structure types with the ability of FPASS to determine previously unknown structure types (as in $LiSbO_3$), we could further accelerate FPASS and make it possible to solve the structures of even more materials.

Another possible route for improving FPASS is to incorporate techniques developed for Crystal Structure Prediction (CSP) and by other direct-space structure solution tools into our genetic algorithm. Both communities have developed many unique and successful global optimization algorithms that accelerate the search for the crystal structures with desirable properties (e.g., low energy, high agreement with diffraction patterns) by maintaining local structural motifs, including slicing-based mating operations [39], treating clusters of atoms as building blocks [14], "ripple" mutation operations [40], and many others [13,14,17,39,61–65]. Alternatively, techniques such as intelligent random structure searches [66], particle swarm optimization [65,67], and minima hopping [68] could also prove to be viable routes for crystal-structures solution. Exploring how to do so could be a valuable research direction.

V. CONCLUSION

In this work, we described an implementation of the FPASS algorithm capable of being used to automatically solve incompletely determined crystal structures. We validated this algorithm by determining the structures of 95 known crystal structures and found that FPASS identified the correct structure in each case. We found that FPASS performed worse when provided lower-quality XRD patterns and for structures with large numbers of symmetrically unique atoms. Once validated, we tuned the algorithm to increase its reliability for crystal structures that were difficult to solve and then applied it to attempt to solve 20 yet-undetermined structures from the powder diffraction file. To date, we have solved the structures of ten compounds and added these structures to the OQMD,

thereby increasing the completeness of this database. We propose that FPASS can be further improved by accounting for partially occupied structures and including information of known crystal structure prototypes into the solution process.

ACKNOWLEDGMENTS

This work was performed under the financial assistance Award No. 70NANB14H012 from the U.S. Department of Commerce, National Institute of Standards and Technology, as part of the Center for Hierarchical Materials Design (CHiMaD). L.W. also acknowledges partial support by the National Defense Science and Engineering Graduate (NDSEG) Fellowship. Software used to perform the FPASS calculations was created with support by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Grant No. DE-FG02-07ER46433. Calculations were performed at the DoD Supercomputing Resource Center at the Air Force Research Laboratory. This work made use of the J.B. Cohen X-Ray Diffraction Facility supported by the MRSEC program of the National Science Foundation (Grant No. DMR-1121262) at the Materials Research Center of Northwestern University and the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF Grant No. NNCI-1542205).

[1] J. Hafner, C. Wolverton, and G. Ceder, MRS Bull. **31**, 659 (2006).
[2] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, Nat. Mater. **12**, 191 (2013).
[3] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, JOM **65**, 1501 (2013).
[4] H. Chen, G. Hautier, A. Jain, C. Moore, B. Kang, R. Doe, L. Wu, Y. Zhu, Y. Tang, and G. Ceder, Chem. Mater. **24**, 2009 (2012).
[5] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 11002 (2013).
[6] D. H. Snydacker, M. Aykol, S. Kirklin, and C. Wolverton, J. Electrochem. Soc. **163**, A2054 (2016).
[7] S. Kirklin, J. E. Saal, V. I. Hegde, and C. Wolverton, Acta Mater. **102**, 125 (2016).
[8] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, Acta Crystallogr. Sect. B **58**, 364 (2002).
[9] http://www.icdd.com/products/pdf4.htm
[10] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, Npj Comput. Mater. **1**, 15010 (2015).
[11] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, Comput. Mater. Sci. **58**, 227 (2012).
[12] V. Pecharsky and P. Zavalij, *Fundamentals of Powder Diffraction and Structural Characterization of Materials* (Springer US, Boston, MA, 2009).
[13] H. Putz, J. C. Schön, and M. Jansen, J. Appl. Crystallogr. **32**, 864 (1999).
[14] R. Černý and V. Favre-Nicolin, Powder Diffr. **20**, 359 (2005).
[15] B. Meredig and C. Wolverton, Nat. Mater. **12**, 123 (2013).
[16] O. J. Lanning, S. Habershon, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, E. Tedesco, and G. W. Turner, Chem. Phys. Lett. **317**, 296 (2000).
[17] V. Brodski, R. Peschar, and H. Schenk, J. Appl. Crystallogr. **36**, 239 (2003).
[18] V. Favre-Nicolin and R. Černý, J. Appl. Crystallogr. **35**, 734 (2002).
[19] L. Ward, K. Michel, and C. Wolverton, Acta Crystallogr. Sect. A Found. Adv. **71**, 542 (2015).
[20] G. L. W. Hart, L. J. Nelson, R. R. Vanfleet, B. J. Campbell, M. H. F. Sluiter, J. H. Neethling, E. J. Olivier, S. Allies, C. I. Lang, B. Meredig, and C. Wolverton, Acta Mater. **124**, 325 (2017).
[21] https://github.com/materials/mint/
[22] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.1.063802 for a tutorial on using the FPASS software, further analysis on the distribution of Wyckoff sites in crystal structures, and detailed information on each structure solved in this work.
[23] G. Hautier, S. P. Ong, A. Jain, C. J. Moore, and G. Ceder, Phys. Rev. B **85**, 155208 (2012).
[24] V. Stevanović, S. Lany, X. Zhang, and A. Zunger, Phys. Rev. B **85**, 115104 (2012).
[25] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[26] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[27] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14251 (1994).
[28] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[29] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[30] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[31] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[32] M. Ladd and R. Palmer, *Structure Determination by X-Ray Crystallography*, 4th ed. (Kluwer Academic/Plenum Publishers, New York, NY, 2003).
[33] C. Giacovazzo, H. L. Monaco, G. Artioli, D. Viterbo, G. Ferraris, G. Gilli, G. Zanotti, and M. Catti, *Fundamentals of Crystallography*, 2nd ed. (Oxford University Press, New York, NY, 2002).
[34] W. A. Dollase, J. Appl. Crystallogr. **19**, 267 (1986).
[35] D. E. King, J. Mach. Learn. Res. **10**, 1755 (2009).
[36] H. Rietveld, J. Appl. Crystallogr. **2**, 65 (1969).
[37] E. Jansen, W. Schäfer, and G. Will, J. Appl. Crystallogr. **27**, 492 (1994).
[38] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, Dalt. Trans. 2832 (2008).
[39] A. R. Oganov, Y. Ma, A. O. Lyakhov, M. Valle, and C. Gatti, Rev. Mineral. Geochemistry **71**, 271 (2010).
[40] D. C. Lonie and E. Zurek, Comput. Phys. Commun. **182**, 372 (2011).
[41] https://bitbucket.org/wolverton/fpass-manager

[42] A. R. Akbarzadeh, V. Ozoliņš, and C. Wolverton, Adv. Mater. **19**, 3233 (2007).
[43] The PDF references a paper reporting the synthesis of $Al_3NiGe_2Y_3$, which does not mention the creation of the Fe-containing analogue. The powder XRD data was from "grant-in-aid" from the ICDD; no paper was listed.
[44] J. T. Zhao and E. Parthé, Acta Crystallogr. Sect. C Cryst. Struct. Commun. **46**, 2273 (1990).
[45] C. Schank, F. Jährling, L. Luo, A. Grauel, C. Wassilew, R. Borth, G. Olesch, C. Bredl, C. Geibel, and F. Steglich, J. Alloys Compd. **207–208**, 329 (1994).
[46] C. Franz, A. Senyshyn, A. Regnat, C. Duvinage, R. Schönmann, A. Bauer, Y. Prots, L. Akselrud, V. Hlukhyy, V. Baran, and C. Pfleiderer, J. Alloys Compd. **688**, 978 (2016).
[47] S. Vasala and M. Karppinen, Prog. Solid State Chem. **43**, 1 (2015).
[48] E. D. Politova and Y. N. Venevtsev, Dokl. Akad. Nauk SSSR **209**, 838 (1973).
[49] S. Gražulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, J. Appl. Crystallogr. **42**, 726 (2009).
[50] S. Westman, P.-E. Werner, T. Schuler, W. Raldow, and P. H. Nielsen, Acta Chem. Scand. **35a**, 467 (1981).
[51] E. J. T. Salter, J. N. Blandy, and S. J. Clarke, Inorg. Chem. **55**, 1697 (2016).
[52] J. Guo, S. Jin, G. Wang, S. Wang, K. Zhu, T. Zhou, M. He, and X. Chen, Phys. Rev. B **82**, 180520(R) (2010).
[53] A. P. Ayala, C. W. Paschoal, J.-Y. Gesland, J. Ellena, E. E. Castellano, and R. L. Moreira, J. Phys. Condens. Matter **14**, 5485 (2002).
[54] M. M. Schieber, Inorg. Chem. **4**, 762 (1965).
[55] V. B. Nalbandyan, M. Avdeev, and A. A. Pospelov, Solid State Sci. **8**, 1430 (2006).
[56] A. Boulineau, L. Croguennec, C. Delmas, and F. Weill, Chem. Mater. **21**, 4216 (2009).
[57] M. Edstrand and N. Ingri, Acta Chem. Scand. **8**, 1021 (1954).
[58] Y. Hashimoto, M. Takahashi, S. Kikkawa, and F. Kanamaru, J. Solid State Chem. **125**, 37 (1996).
[59] M. Li, W. Yuan, J. Wang, C. Gu, and H. Zhao, Powder Diffr. **22**, 59 (2007).
[60] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Chem. Mater. **22**, 3762 (2010).
[61] W. W. Tipton and R. G. Hennig, J. Phys. Condens. Matter **25**, 495401 (2013).
[62] B. C. Revard, W. W. Tipton, and R. G. Hennig, Predict. Calc. Cryst. Struct. **345**, 181 (2014).
[63] S. M. Woodley and R. Catlow, Nat. Mater. **7**, 937 (2008).
[64] A. Kinaci, B. Narayanan, F. G. Sen, M. J. Davis, S. K. Gray, S. K. R. S. Sankaranarayanan, and M. K. Y. Chan, Sci. Rep. **6**, 24974 (2016).
[65] P. Gao, Q. Tong, J. Lv, Y. Wang, and Y. Ma, Comput. Phys. Commun. **213**, 40 (2017).
[66] C. J. Pickard and R. J. Needs, J. Phys. Condens. Matter **23**, 53201 (2011).
[67] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Phys. Rev. B **82**, 094116 (2010).
[68] M. Amsler and S. Goedecker, J. Chem. Phys. **133**, 224104 (2010).