Multi-Objective Quality-Diversity for Crystal Structure Prediction

Hannah Janmohamed
Imperial College London, InstaDeep
London, UK
hnj21@imperial.ac.uk

Marta Wolinska
Imperial College London
London, UK

Shikha Surana
InstaDeep
London, UK

Thomas Pierrot
InstaDeep
Boston, USA

Aron Walsh
Imperial College London
London, UK

Antoine Cully
Imperial College London
London, UK

ABSTRACT

Crystal structures are indispensable across various domains, from batteries to solar cells, and extensive research has been dedicated to predicting their properties based on their atomic configurations. However, prevailing Crystal Structure Prediction methods focus on identifying the most stable solutions that lie at the global minimum of the energy function. This approach overlooks other potentially interesting materials that lie in neighbouring local minima and have different material properties such as conductivity or resistance to deformation. By contrast, Quality-Diversity algorithms provide a promising avenue for Crystal Structure Prediction as they aim to find a collection of high-performing solutions that have diverse characteristics. However, it may also be valuable to optimise for the stability of crystal structures alongside other objectives such as magnetism or thermoelectric efficiency. Therefore, in this work, we harness the power of Multi-Objective Quality-Diversity algorithms in order to find crystal structures which have diverse features *and* achieve different trade-offs of objectives. We analyse our approach on 5 crystal systems and demonstrate that it is not only able to re-discover known real-life structures, but also find promising new ones. Moreover, we propose a method for illuminating the objective space to gain an understanding of what trade-offs can be achieved.
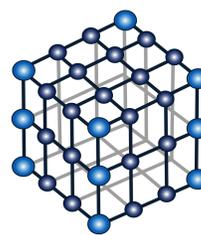
CCS CONCEPTS

• **Theory of computation → Evolutionary algorithms**; • **Applied computing → Multi-criterion optimization and decision-making**.

Figure 1: *Left:* **Illustration of crystal structure: atoms and molecules are arranged in a repeated, ordered pattern.** *Right:* **The crystal structure energy landscape is rugged with several local optima. Relaxation is a form of local optimisation that brings solutions toward local optima.**

1 INTRODUCTION

Inorganic crystal structures are pervasive in our everyday lives, forming the foundation of materials used in industries as diverse as electronics, pharmaceuticals, and energy storage. Consequently, advancements in materials science could have revolutionary effects across a wide variety of sectors. For instance, in the electronics industry, the discovery of novel crystal structures could lead to the development of more efficient semiconductors and superconductors, enabling faster and more energy-efficient electronic devices [28].

Crystal Structure Prediction (CSP) is an optimisation problem that aims to predict stable atomic arrangements within crystalline materials. Traditionally, computational techniques for CSP have focused on finding the global minimum of the energy function, assuming that the crystal structure with the lowest energy is the most stable. Existing methods, such as genetic algorithms [30], random search [8], and particle swarm optimization [9], employ various search strategies to locate this global minimum. While these methods have made significant contributions to the field, they often

overlook potentially valuable regions of the objective space and may miss relevant meta-stable structures in local optima [19].

Consider carbon as an illustrative example. The most stable carbon crystal is graphite, characterised by its layered hexagonal structure. However, there are numerous alternative carbon allotropes, such as diamonds, nanotubes, and fullerenes, each possessing distinct properties with immense technological potential. Traditional CSP methods often fail to explore these alternative structures, as they are solely aimed at finding the lowest energy state.

Quality-Diversity (QD) algorithms [12], primarily used in evolutionary robotics and optimization [11], have recently emerged as a promising approach for addressing CSP challenges [41]. QD algorithms aim to discover not only the single best solution but instead a large collection of different, yet high-quality, solutions across the search space. This approach has many advantages in the context of CSP. Firstly, seeking a collection of solutions could lead to the discovery of multiple local optima, each potentially representing unique and interesting materials. Discovering a multitude of crystal structures not only enhances our understanding of materials but could also provide a valuable reservoir of options for synthesis. In particular, synthesising materials under laboratory conditions can often prove challenging [28], so having alternative candidate solutions offers researchers greater flexibility and choice. Finally, since the energy landscape of CSP is typically rugged [19], having a collection of solutions could aid exploration, as diverse solutions can serve as stepping stones to navigate and escape local optima effectively [18, 35].

Recently, Wolinska et al. [41] applied the well-established MAP-ELITES algorithm to the CSP problem by using domain-specific mutation operators and graph neural-networks as surrogate models for evaluating potential crystal structures. The authors demonstrate that, by using this approach, QD algorithms are capable of discovering a wide variety of crystal structures [41]. Excitingly, their method was not only able to discover structures known by material scientists, but also to discover some promising new structures.

While this is an important advancement, using stability as the sole objective does not allow optimisation of any other desired objectives of the materials, such as magnetism or thermoelectric efficiency. In order to find materials that are useful for downstream applications, it would be extremely valuable to allow for the optimisation of crystal structure stability in tandem with other objectives. With this goal in mind, in this work, we apply Multi-Objective Quality-Diversity algorithms to the field of CSP. Multi-objective Quality-Diversity (MOQD) [23, 29, 33] is a new research effort that aims to find diverse solutions that simultaneously optimise multiple objectives. For CSP this allows us to not only find diverse crystalline materials but also explore trade-offs and properties associated with these crystal structures.

In this paper, we build upon the work of Wolinska et. al [41] by applying Multi-Objective Quality-Diversity algorithms to Crystal Structure Prediction. We evaluate our method on 5 different crystal systems and demonstrate that our method is able to find a large collection of crystal structures that have diverse conductivity properties and deformation-resistance, and that achieve varying trade-offs between stability and magnetism. Moreover, we demonstrate that our approach is not only able to re-discover known structures, but also uncover novel structures that may surpass these known materials. Finally, we illustrate how our approach can be used to illuminate the search space, highlighting the possible trade-offs and properties that crystal structures can exhibit. All of our work is fully containerised and is available at: https://github.com/adaptive-intelligent-robotics/MOQD-CSP.

2 BACKGROUND

2.1 Crystal Structures

Crystals are three-dimensional arrangements of atoms or molecules, characterised by a repeating pattern (see Figure 1) [38]. The entire crystal structure can be summarised by its *unit cell*, which is the smallest repeating structural component of the crystal [31]. The positions of atoms in the unit cell collectively define the crystal's stability and properties.

Finding novel crystal structures experimentally in a laboratory setting can be a technically challenging, resource-intensive and time-consuming task. To overcome these challenges, computational methods have become an indispensable tool for finding new materials. These techniques leverage the power of computers to explore the vast search space of possible crystal structures, assessing their stability through an *energy function* (see Figure 1). The energy function calculates the total energy of a given atomic arrangement within a crystal, taking into account the interactions between atoms, including their bond lengths and angles. In essence, it quantifies the stability of a crystal structure based on the relative positions of its constituent atoms. The aim of Crystal Structure Prediction (CSP) is to find structures which lie at the global minimum of this energy function and are therefore highly stable.

Computing the energy function of a crystal structure is not always a straightforward task. In most applications, to accurately predict crystal structures, researchers often employ Density Functional Theory (DFT) calculations [3]. DFT calculations estimate the properties of materials based on quantum mechanical principles and, while they are highly accurate, they are also computationally expensive, requiring significant computational resources and time. To address this, surrogate models [14] have emerged as an alternative approach for approximating the energy function. These surrogate models aim to reduce the computational burden by providing quick and cost-effective approximations of the energy landscape. While surrogate models offer efficiency gains, they introduce the challenge of accuracy and reliability, as their predictions may not always align perfectly with the true energy values.

The CSP problem is inherently challenging due to the vast high-dimensional space of possible atomic configuration and ruggedness of the energy landscape [19]. Given the complexities of both experimental and computational approaches, crystal structure prediction remains a challenging and critical problem.
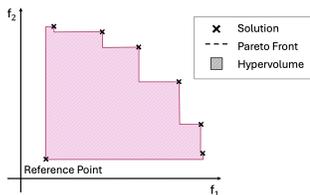
2.2 Quality-Diversity Algorithms

Traditional optimisation methods aim to find a single, high-quality solution, in contrast, the objective of Quality-Diversity (QD) optimisation algorithms is to find a set of solutions that are both diverse and high-performing. In QD, a solution is characterised by a *fitness* score f and a *feature* vector d. The feature vector characterises some aspect of interest of the solution. For example, in robotic locomotion the feature vector could reflect the gait of the robot

[11, 33] or in video game level design this could be the number of enemy characters [24]. The objective of QD is to find a collection of solutions which have diverse feature vectors and high fitness.

In the family of QD optimisation methods, Multi-dimensional Archive of Phenotypic Elites (MAP-ELITES) is a simple yet powerful algorithm used successfully in several prior works [23, 29, 33]. In MAP-ELITES, the feature space is discretised into a grid composed of cells. Each cell corresponds to a feature vector and stores the solution characterised by the same feature. The algorithm begins with initialising the grid and placing randomly sampled solutions in the cells. At each iteration of the MAP-ELITES algorithm, the first step consists of selecting a batch of solutions from the grid. Next, copies of these solutions undergo variation to form offspring solutions, which are evaluated to obtain the fitness scores and the features. Finally, each new solution is added back to the grid if the cell matching the solution's feature vector is empty or if the existing solution within the cell is outperformed by the new solution. By repeating this loop for a specified budget, the grid gradually accumulates diverse and high-quality solutions. The performance of MAP-ELITES algorithms can be assessed via the QD-SCORE which is the sum of the fitness of the solutions in each cell.

2.3 Multi-Objective Optimisation

Figure 2: The Pareto Front represents the set of different trade-offs on objectives. The hypervolume reflects the area between points on the Pareto Front and a reference point.

Multi-objective optimisation considers the problem of maximising a set of different, and often conflicting, objectives at once. Concretely, given a solution $x \in \mathcal{X}$ and an objective $\mathbf{f} : \mathcal{X} \rightarrow \mathbb{R}^k$, the problem is formalised as $\max_{x \in \mathcal{X}}(f_1(x), ..., f_k(x))$.

In reality, for most problems, achieving a solution that maximises all objectives simultaneously is often unattainable. Consequently, the aim becomes identifying the set of solutions which achieve the best possible trade-offs on the set of objectives, known as the *Pareto front* (see Figure 2). A solution belongs to the front if it scores at least as high as the other solutions on the front on all of the objectives, and outperforms them on at least one objective [20]. Solutions that belong to the front are called *Pareto optimal*.

The performance of multi-objective optimisation is often assessed via the hypervolume metric which reflects the area of space between points on the front and a fixed reference point r [4] (see Figure 2). Pareto Fronts that contain solutions that are more evenly distributed and higher-performing on each objective will have a higher hypervolume [20].

2.4 Multi-Objective Quality-Diversity

Multi-Objective Quality-Diversity (MOQD) aims to generate a diverse set of high-performing solutions that maximise a combination of k different objectives $\mathbf{f} = [f_1, f_2, ...f_k]$. Multi-Objective MAP-Elites (MOME) [33], a well-established MOQD algorithm, achieves this by storing a Pareto Front of solutions in each grid cell. The algorithm flow closely matches that of MAP-Elites with one crucial distinction: after evaluating a new solution and identifying its corresponding cell based on its feature, the solution is placed into the cell only if it Pareto-dominates the other existing solutions in that cell. The objective of MOQD algorithms is to find an approximation of the Pareto Front in each cell that has the best set of trade-offs and therefore maximum hypervolume. Hence, the performance of MOQD algorithms is assessed by the sum of the hypervolumes of all the Pareto Front approximations in the grid, referred to as the MOQD-SCORE. This objective translates to finding the best possible trade-offs across the set of objectives for each feature of interest.

Since the Pareto Front can contain infinitely many solutions, in order to control the memory costs and allow for parallelisation, previous works in MOQD use a fixed maximum size for the Pareto Front [23, 33]. Prior works in the field of MOQD build upon the MOME algorithm and have applied it to policy learning robotics tasks and environments. However, to the best of our knowledge, no other MOQD works have been applied to CSP.

3 RELATED WORKS

3.1 Crystal Structure Prediction

Within the field of Crystal Structure Prediction, recent advancements have been heavily influenced by the growing availability of crystallographic databases [22]. For example, data-mining methods analyse these large databases with the aim to uncover insights for predicting new structures [13]. Deep learning methods, on the other hand, harness the power of neural networks to extract features from high-dimensional crystallographic data [36]. Additionally, generative models such as auto-encoders [10] and generative adversarial networks [25] have emerged as a creative means to produce entirely new crystal structures. While these learning-based approaches leverage extensive data sources, they may also be biased towards these databases and therefore could be potentially limited for the discovery of entirely novel crystal structures.

Other efforts in CSP have been driven by evolutionary algorithms (EAs) [30]. These algorithms work by generating new structures through genetic-like operators and gradually improving the structures over multiple iterations by retaining those which are most promising. Most of the work in applying EAs to CSP has been focused on injecting domain knowledge into the evolutionary process. For example, some works explore using alternative crystal representations to reduce the complexity of the search space [1, 34] while others introduce specialised variation operators which, for example, preserve crystal symmetries [19, 27, 28].
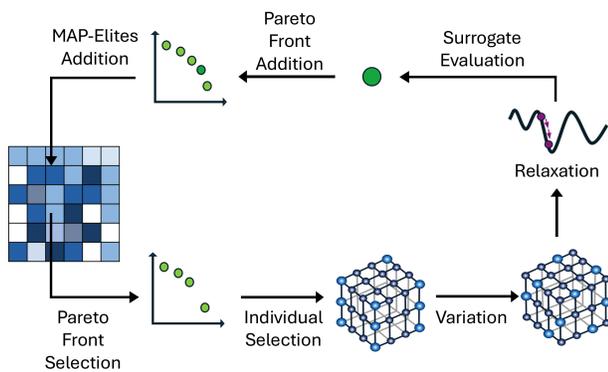
A key challenge of using EAs for CSP arises from the costly nature of evaluating crystal structures. Therefore some methods aim to optimise a surrogate energy model rather than rely on resource-intensive DFT calculations. While these models provide fast estimates of the energy landscape, they come with inherent reliability

concerns. However, recent works have greatly enhanced the accuracy of surrogate models by training them to be conservative in their estimates [34] or leveraging ensemble techniques [28].

3.2 Quality-Diversity for Crystal Structure Prediction

Despite the progress made in CSP, there remains a notable gap in the CSP literature. Most existing methods primarily focus on finding the global minimum of the energy function, largely overlooking the exploration of diverse crystal structures and their associated properties. In a recent study by Wolinska et al. [41], the Map-Elites algorithm (see section Section 2.2) was applied to CSP, resulting in the discovery of a diverse range of crystal structures, including both known materials and promising new configurations. While this represents a significant advancement, this work only focused on finding stable crystal structures and disregarded other objectives for properties of the material, such as the toughness and hardness. On the other hand, other works have used multi-objective evolutionary algorithms [1] to find such trade-offs, but do not include specific diversity-maintenance mechanisms which facilitate the discovery of diverse structures. In contrast, our approach utilises Multi-Objective Quality-Diversity algorithms, allowing us to not only uncover diverse crystalline materials but also explore the associated trade-offs and properties within these crystal structures.

4 METHOD

Figure 3: MOME-X for CSP method overview. At each iteration, solutions are selected from the grid and undergo domain-specific variation. We then perform a fixed number of relaxation steps. The new solution is evaluated via a surrogate model and added back to the archive if it belongs to the Pareto Front of the cell corresponding to its feature.

4.1 Overview

In this section we present our method, as visualised in Figure 3). Overall, our method adheres the general flow of the MOME algorithm which maintains a Pareto Front in each cell of a MAP-ELITES grid: at each iteration, solutions are selected from the grid, undergo variation and are added back to the grid if they belong to the corresponding Pareto Front. However, we also utilise the crowding-based selection and addition exploration mechanisms from MOME-PGX

[23], hence we term the approach MOME-X. These crowding-based mechanisms are employed in order to improve exploration and ensure the maintenance of different trade-offs (see Section 4.3).

In order to apply MOME-X to the CSP problem, following Wolinksa et. al [41], we inject domain knowledge into the evolutionary process. Further details are provided in subsequent sections.

4.2 Initialisation

As explained in Section 3.1, crystal structures can be defined by their *unit cell*, which corresponds to the smallest repeating configuration of atoms. Therefore, many works in CSP use the *unit cell parameters* as the genotype within the evolutionary process [19, 41]. The unit cell parameters comprise: 3 lattice parameters a, b, c which describe the length of the bounding box of the unit cell, 3 parameters α, β, γ which describe the angles of this bounding box and $3 \times (N-1)$ parameters that describe the (x, y, z) co-ordinates of each of the atoms in the unit cell, where N is the number of atoms in the unit cell. However, one significant challenge in CSP is the dimensionality of the search space: there are approximately 10^N distinct structures for a structure with N atoms. Of these arrangements, many candidates may not achieve the intricate symmetrical properties that crystal structures must obey for stability. Therefore, if we were to initialise the population of structures randomly, it is highly likely that this would result in highly unstable structures which would be hard for the optimisation process to bootstrap from. To address this, evolutionary algorithms may prefer to generate an initial population using domain-specific heuristics to ensure that the initial candidates are realistic structures [19, 41]. We choose to use the `pyxtal` package [17] to create initial candidates. The `pyxtal` package generates structures that conform to the permissible symmetries within a specific crystal system.

4.3 Crowding-based Exploration

After generating the initial candidate structures, the algorithm enters the main algorithmic flow: at each iteration solutions are selected from the grid, mutated and evaluated for potential re-addition to the MAP-ELITES grid. In this work, we incorporate crowding-based selection and addition mechanisms, first introduced in MOME-PGX [23]. At each iteration, after selecting a Pareto Front from the grid with uniform probability, a solution from this front must be selected for variation. The crowding-based selection operator biases the selection of solutions which lie in sparser regions of the Pareto Front which encourages exploration across all objectives. Similarly, the crowding-based addition mechanism is used to encourage a uniform spread of solutions on the Pareto Front of each cell. In particular, since each cell's Pareto Front has a fixed maximum size, if a new solution is added to the front of a cell which is already at maximum capacity another solution must be replaced. In the MOME algorithm, a random solution from the front is replaced, which could result in a loss of coverage of the objective space and a decrease in MOQD score. By contrast, the crowding-based addition operator ensures that the solutions which would provide the most even distribution across the Pareto Front are kept in each cell, ensuring that the best set of possible trade-offs are achieved. We refer the interested reader to the MOME-PGX paper [23] for more details about these operators.

4.4 Mutations

The use of traditional genetic algorithm GA variation operators poses a significant challenge in CSP since many of these operators have the potential to disrupt these symmetrical properties of crystal structures. To address this issue, numerous works in CSP employ domain-specific mutation operators [1, 19, 28, 41], which are tailored to preserve these essential symmetries. In our research, we adopt two domain specific variation operators from the Atomic Simulation Environment, `ase` package [26]: 1) *strain* mutations, capable of compressing or expanding crystal structures, and 2) *permutation* mutations, which facilitate the rearrangement of atoms of distinct types within the structure. While alternative variation operators are available [1, 26–28], an extensive study of alternative choices is beyond the scope of this work. Moreover, previous studies have demonstrated that these selected operators are effective, especially when initiated from well-structured starting points [19].

4.5 Relaxation

Despite the use of domain-specific variations operators tailored to CSP, there remains the potential to inadvertently disrupt the structural crystal symmetries when applying mutations. To mitigate this challenge, after the application of mutation operators, we apply *relaxation* [2]. As illustrated in Figure 1, relaxation is a form of local optimisation, leveraging the energy function's gradient to guide the solution towards its nearest local optima while preserving critical symmetrical properties. The gradient of the energy function represents the force on each of the atoms and minimising the force by moving toward a local optimum leads to a more stable structure.

Relaxation is a highly effective procedure for preventing the generation of redundant crystal structures, making it a prevalent practice in CSP methodologies [1, 19, 28, 30, 41]. However, since relaxation is a form of local optimisation it requires performing additional energy function evaluations. Therefore, this process comes with a computational cost and reduces the overall budget of evaluations typically available in evolutionary algorithms. To manage the computational budget, we apply a fixed number of relaxation steps to each offspring solution. Then we use a surrogate model to estimate the force on the atoms of the offspring solution and apply a loose filtering system: if the force is still very high after these relaxation steps, it is removed from the offspring candidates.

4.6 Evaluating Solutions

Evolutionary algorithms typically require evaluating many thousands of solutions at each iteration. In the context of CSP, accurate energy calculations could be achieved through DFT. However, the sheer volume of evaluations typically needed in evolutionary algorithms makes this approach prohibitively expensive. To address this challenge, our method employs pre-trained neural networks [5–7, 14], trained on extensive materials databases as surrogate models for evaluating the fitness and features of solutions. Although these models are less accurate compared to DFT, we contend that this limitation is manageable for several reasons.

Firstly, recent advancements in deep learning have demonstrated a trend towards improving neural network accuracy in various fields, including CSP. Indeed, recent works which employ techniques such as ensemble training [28] and conservative loss functions [34] have already facilitated enhanced surrogate model accuracy. Thus, we anticipate that these models will continue to improve and will soon be a competitive alternative to DFT calculations.

Secondly, we hope that future work could extend our method as a tool for fine-tuning surrogate models through active-learning strategies, following a similar approach to GNoME [28]. For example, we found that the surrogate model predicted some solutions to have a negative fitness, which is not feasible in real-life. In this work, we simply filtered out these unrealistic solutions. However, future works could re-evaluate these out-of-distribution solutions via DFT and then using them as training examples to further refine the surrogate model. While this lies beyond the scope of our present work, it offers an intriguing avenue for future CSP research.

5 EXPERIMENTAL SET UP

5.1 Experiments

We evaluate our method on 5 crystal systems: Carbon (C), Silicon (Si), Silicon Carbide (SiC), Silicon Dioxide (SiO_2) and Titanium Dioxide (TiO_2). For each of the systems, we use the following features:

(1) The *band gap* [37] reflects the energy required to move an electron from the valence band to the conduction band and therefore is an indication of the conductivity of a material. We estimate the band gap via the `"MEGNet-MP-2019.4.1-BandGap-mfi"` model from the Materials Graph Library, `matgl` [5].

(2) The *shear modulus* [42] of a material quantifies its resistance to deformation under shear stress, providing information about its stiffness and ability to withstand forces. We estimate the shear modulus using the `"logG_MP_2018"` model from the Materials Graph Network Library (`MEGNet`) [6, 7].

The aim of the method is to find crystal structures which are diverse in their band gap and shear modulus, and are high-performing on a set of objectives. In this work, for all systems, we consider the following objectives [41]:

(1) The *stability* of the crystal structure, which we quantify via the energy function, is estimated by `CHGNet` [15]. The objective is to minimise the total energy of the crystal structure, reflecting the configuration that represents the most stable state for the crystal. The stability of a structure is actually assessed by its value on the energy function, with more negative values being more stable. For simplicity, we take the negative energy as fitness and seek to maximise this.

(2) We also use `CHGNet` to estimate the *magnetic moment* of the structures. The magnetic moment is a vector quantity that characterises magnetic moment associated with each atom in a crystal lattice. We combine the moments for each atom into a single magnetism score, known as the total magnetic moment. For simplicity, we refer to the total magnetic moment as *magnetism* henceforth.

We note that the choice of objectives in this work serve as a starting point for applying MOQD to CSP. Future research avenues could explore additional material properties such as toughness and hardness [1]. Alternatively, we might consider using the uncertainty of an ensemble of the surrogate model predictions as a meta-objective,
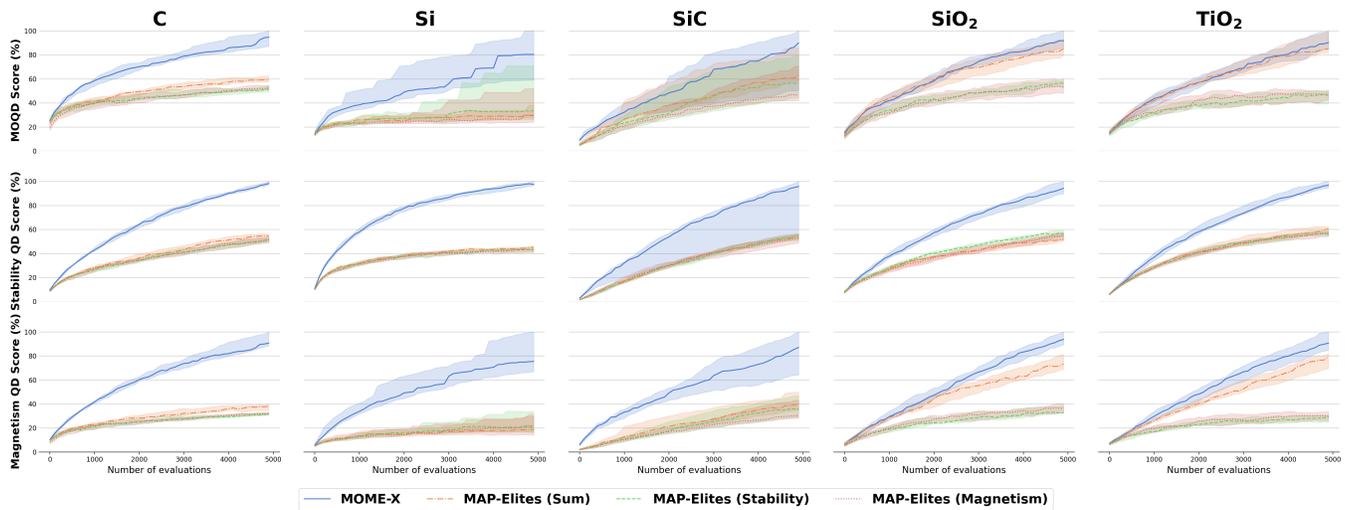
Figure 4: Median performance of 15 seeds, the shaded regions show the inter-quartile range.

enabling optimisation for both stability and the confidence in that stability prediction. These possibilities illustrate the flexibility and potential for expanding the objectives within the MOQD framework to enhance crystal structure discovery.

5.2 Baselines

We evaluate our method, MOME-X, against three MAP-ELITES baselines: 1) ME-ENERGY 2) ME-MAGNETISM and 3) ME-SUM. Each of these baselines corresponds to applying the MAP-ELITES algorithm to CSP, optimising for stability of structures, magnetism of structures and an additive sum of the two respectively. We use the same mutation operators, initialisation and relaxation procedures for all of the baselines and MOME-X. However, it should be noted that a straightforward comparison between MOME algorithms and MAP-ELITES algorithms is non-trivial, since MAP-ELITES algorithms store at most one solution per cell whereas MOME algorithms can store several. In particular, if we use an archive tessellated into c cells for MOME-X, each with a maximum Pareto Front length p, MOME-X could have a maximum population size of $c \times p$. Therefore, for fairness, in all MAP-ELITES baselines we tessellate the archive into $c \times p$ cells.

On the other hand, calculating MOQD metrics on grids with different tessellation sizes would also not be a fair comparison. Therefore, to report metrics for the baselines, we also maintain passive MOME archives alongside the main MAP-ELITES archives [23, 33]. At each iteration, all of the solutions from the MAP-ELITES archives are added to these passive MOME archives using the normal Pareto Front addition rules. The passive archives are then used to calculate and report metrics, but do not interact with the main algorithm flow.

5.3 Hyperparameters

For MOME-X, we use an archive with a CVT tessellation [39] of 200 cells, each with a maximum Pareto Front length of 10. Therefore, for MAP-ELITES baselines we use a CVT tessellation of 2000 cells. We run all experiments for a total of 5000 evaluations, selecting 100 solutions from the grid at each iteration. For all systems, we consider

structures with 24 atoms, with an initial unit cell volume of 450 A^3 and ratio of covalent radii of 0.4. For multi-atom systems (SiC, SiO_2, TiO_2) we apply strain and permutation mutation operators with equal probability and for single-atom systems (Si, C) we only apply strain mutation operators. After applying the variation operators we use 100 relaxation steps and filter out solutions with a force greater than 1.0 eV/A. We use a reference point of $[0, 0]$ to calculate the hypervolume. Each experiment was repeated for 15 replications.
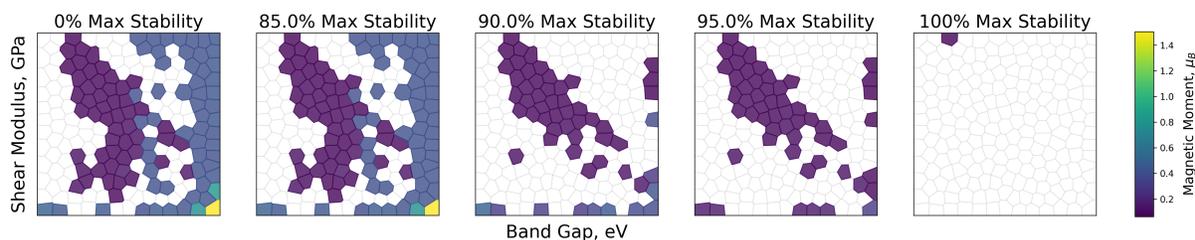
6 RESULTS

6.1 Multi-Objective Quality-Diversity Results

We first evaluate MOME-X against the MAP-ELITES baselines using the following MOQD metrics:

(1) MOQD-SCORE: the sum of the hypervolumes of each Pareto Front of the MOME archive. This metric aims to reflect whether our approach is able to find solutions which are diverse in the feature space, that achieve different trade-offs on each of the objectives (see Section 2.4).

(2) ENERGY-QD-SCORE: the sum of the energy scores of all solutions in the archive. This is what the ME-ENERGY baseline is explicitly trying to maximise.

(3) MAGNETISM-QD-SCORE: the sum of the magnetism scores of all solutions in the archive. This is what the ME-MAGNETISM baseline is explicitly trying to maximise.

We verify the statistical significance of our results by reporting the p-value from a Wilcoxon signed-rank test [40] using a Holm-Bonferroni correction [21].

Firstly, Figure 4 shows that MOME-X outperforms or matches the performance of all of the baselines across all of the metrics in all of the systems. Crucially, this shows that MOME-X is able to find the largest selection of possible trade-offs of stability and magnetism for varying band-gaps and shear-moduli. Interestingly, MOME-X outperforms ($p < 0.03$) all baselines on the MOQD-SCORE for single-atom systems (C, Si) but matches the performance of ME-SUM on multi-atom systems (SiC, TiO_2, SiO_2). This suggests

Figure 5: Archive plot for Silicon Carbide, colour coded to show the maximum magnetism fitness in each cell. Different plots show different threshold levels for the minimum stability of solutions.

that scalarisation of the objectives may not always achieve a wide variety of possible trade-offs.

Figure 4 shows also that MOME-X also outperforms ($p < 0.03$) all other baselines on the ENERGY-QD-SCORE and MAGNETISM-QD-SCORE. This is a particularly intriguing result, given that these metrics are what ME-ENERGY and ME-MAGNETISM are respectively explicitly aiming to maximise. This corroborates previous observations that simultaneously optimising over several objectives may provide helpful stepping stone solutions that aid exploration and, in turn, can help to maximise each objective separately [23]. Additionally, we observe that MOME-X also outperforms ME-SUM on these metrics. This could suggest that ME-SUM succeeds in finding solutions that are balanced across the two objectives, but is less able to find solutions that are high-performing on each objective separately and therefore provides fewer possible trade-offs for each cell.

Finally, we note two further metrics that can be used to asses MOQD performance: the COVERAGE and the GLOBAL-HYPERVOLUME [23, 33]. The COVERAGE is the proportion of cells in the archive that have at least one solution and aims to assess whether the algorithm can find solutions that are diverse in the feature space. The GLOBAL-HYPERVOLUME is the hypervolume of the Pareto Front of all the solutions from the entire archive and aims to assess the multi-objective performance of the algorithm, when disregarding the solutions' features. We find that these metrics are similar for most algorithms so report the results in Appendix A. We note that this result is expected since all baselines are designed to find solutions cover the feature space and the global Pareto Front is attained by most methods. However, we emphasise that only MOME-X was able to solve the combinatorial problem of finding the best Pareto Front for each feature, as reflected in its high MOQD-SCORE.

6.2 Illumination

In Section 6.1 we demonstrated that MOME-X surpasses all MAP-ELITES baselines across various MOQD metrics, highlighting its ability to discover an extensive collection of diverse crystal structures which achieve distinct trade-offs of stability and magnetism. However, we recognise that in many CSP applications, the importance of these objectives may not be equal. Therefore, in order to further understanding what trade-offs are possible, in this section we introduce a visualisation method that portrays the levels of magnetism that are attainable for varying degrees of stability.

To construct this representation, we first computed the minimum and maximum stability scores within the entire archive of one
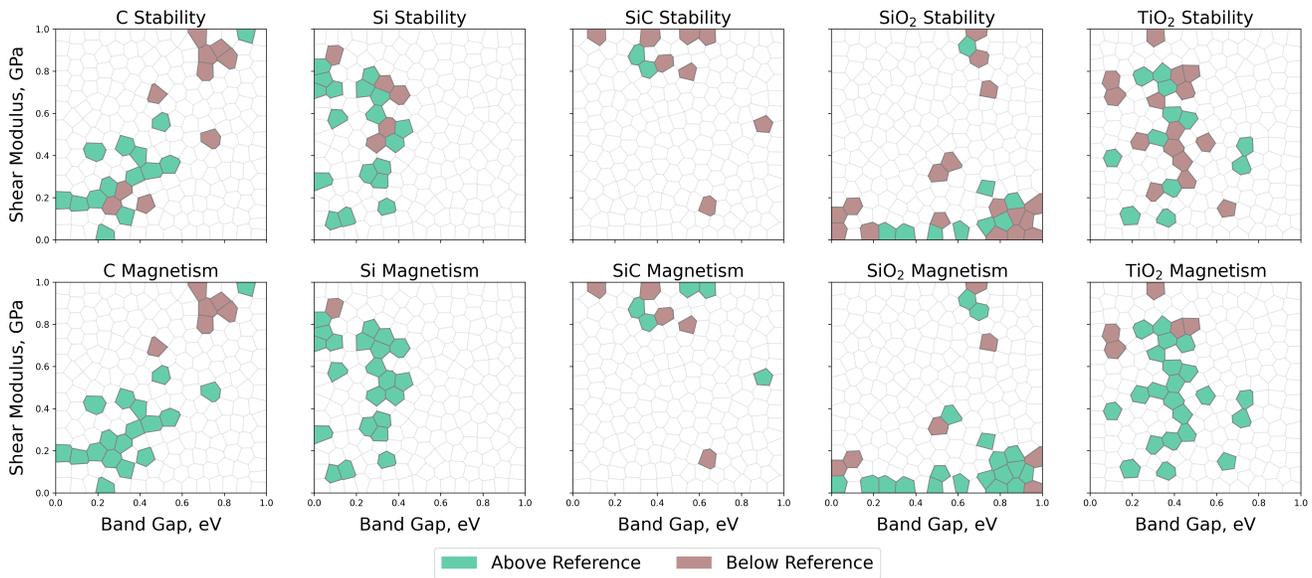
replication. Then, we visualise the archive, colour-coded according to the highest magnetism score that is achieved by solutions for varying degrees of interpolation between this minimum and maximum stability. For example, a 90% interpolation shows the best possible magnetism scores in each cell for solutions which achieve at least 90% of the maximum stability of the entire archive.

Figure 5 shows an example of this plot for the MOME-X run that achieved the median MOQD score for Silicon Carbide (visualisations for other crystal systems are provided in Appendix B). This visualisation reveals when the minimum threshold for stability is lower, the magnetism of solutions is higher. This highlights that there is a genuine trade-off between the stability and magnetism of solutions. However, Figure 5 also shows there are many solutions that perform highly on both objectives, given that there are still many solutions with high magnetism scores, even at high interpolation levels of 85%, 90%, and 95%. Therefore, if the end-user is willing to sacrifice some level of stability for a highly magnetic material, Figure 5 suggests that there are crystal structures that would satisfy this. We emphasise that this visualisation method, while used to illustrate trade-offs between stability and magnetism, could be adapted for exploring trade-offs with other objectives (as detailed in Section 5) and finding solutions that achieve targeted properties.

6.3 Evaluation Against Reference Structures

While utilising MOQD metrics offers valuable insights into the performance of MOME-X, we also wish to validate that our approach is useful for the discovery of real-life crystal structures that could have practical use in materials science applications. This approach is pivotal in confirming that our method is not just yielding solutions that deceive the neural-network surrogate models but rather uncovering genuinely realistic configurations. To accomplish this, we cross-reference the structures found by each of the baselines against those contained in the *Materials Project Database* [22], which is an extensive database of approximately 150,000 materials. To verify whether a solution found by an algorithm matches a real-life reference structure, we use the `StructureMatcher` method from the `pymatgen` [32]. This method provides a series of checks on the unit-cell parameters to determine if there is a match. We consider a solution to match the reference structure if it 1) is a match according to the `StructureMatcher` and 2) lies in the same centroid as the reference structure.

Table 1 shows the median number of reference structure matches found by each of the algorithms. Interestingly, we find that MOME-X

Figure 6: Archive visualisation with cells colour-coded to indicate whether MOME-X found a solution in the same cell as the reference solution that was higher-performing. Top row: stability objective. Bottom row: magnetism objective.

Table 1: Median number of reference structure matches

	MOME-X	ME-ENERGY	ME-MAGNETISM	ME-SUM
C	**3.0**	2.0	2.0	2.0
Si	**6.0**	2.0	3.0	3.0
SiC	0.0	0.0	0.0	0.0
SiO_2	**7.0**	1.0	3.0	3.0
TiO_2	**4.0**	2.0	2.0	1.0

is able to find more real-life structures than other baselines. We hypothesise that this could be because optimising over several objectives provides diverse stepping stone solutions that can aid exploration and thus improve the discovery of realistic structures.

While it is promising to see that MOME-X is able to re-discover real-life structures, we also wanted to verify whether any of the other solutions found by the algorithm outperform these reference structures, on either of the objectives. Figure 6 shows the archives of the runs that score the median MOQD-SCORE in each of the systems. The cells of the reference structures are colour coded according to whether, for each structure, MOME-X finds a solution in the same cell that outperforms the reference structure on either of the objectives. In other words, Figure 6 visualises whether MOME-X is able to find solutions that have similar features to the reference solutions, but are either more stable or have higher magnetism. Excitingly, these results show that MOME-X finds many promising structures that could be better than the known reference materials.

While we note that it is likely that some of these structures could score highly due to inaccuracies of the surrogate model, we argue that this presents an opportunity for future research, rather than a limitation of our method. Indeed, as explained in Section 4.6, we

believe that finding solutions which are able to fool the surrogate model lays the foundation for a very interesting line of future research which use these types of solutions as training examples to improve surrogate models in CSP via active learning. This process could be employed in an adversarial loop: surrogate models could aid the discovery of novel crystal structures and then these structures could be re-assessed via DFT calculations and used to create even better surrogate models. In summary, Figure 6 shows numerous solutions that are either a) higher-performing than known materials or b) could be used to improve surrogate model accuracy, both of which are advantageous outcomes.

7 CONCLUSION

In this work, we applied Multi-Objective Quality-Diversity to Crystal Structure Prediction. Using this approach, we were able to find a range of crystal structures that achieved different trade-offs of magnetism and stability, and we proposed a method for visualising these trade-offs. We found that MOME-X discovered more real-life structures than MAP-ELITES baselines and also had a higher performance on the ENERGY-QD-SCORE and MAGNETISM-QD-SCORE, suggesting that optimising over many objectives may improve exploration. Finally, we explained how our method could fit into a larger framework for active learning, and could be used to optimise for other objectives such as toughness [19] or meta-objectives such as surrogate model uncertainty. Future work could consider improving the efficiency of our method by using gradient-based mutation operators [16, 23]. Alternatively, we could apply our approach to similar problems in other fields, such as the discovery of protein folds [43].

ACKNOWLEDGEMENTS

REFERENCES

[1] Zahed Allahyari and Artem R Oganov. 2020. Coevolutionary search for optimal materials in the space of all possible compounds. *npj Computational Materials* 6, 1 (2020), 55.

[2] Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. 2006. Structural relaxation made simple. *Physical review letters* 97, 17 (2006), 170201.

[3] Philippe Blanchard, Erwin Brüning, Philippe Blanchard, and Erwin Brüning. 2015. Density functional theory of atoms and molecules. *Mathematical Methods in Physics: Distributions, Hilbert Space Operators, Variational Methods, and Applications in Quantum Physics* (2015), 563–573.

[4] Yongtao Cao, Byran J Smucker, and Timothy J Robinson. 2015. On using the hypervolume indicator to compare Pareto fronts: Applications to multi-criteria optimal experimental design. *Journal of Statistical Planning and Inference* 160 (2015), 60–74.

[5] Chi Chen and Shyue Ping Ong. 2022. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* 2, 11 (2022), 718–728.

[6] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. [n. d.]. Supplementary Information Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. ([n. d.]).

[7] Chi Chen, Yunxing Zuo, Weike Ye, Xiangguo Li, and Shyue Ping Ong. 2021. Learning properties of ordered and disordered materials from multi-fidelity data. *Nature Computational Science* 1, 1 (2021), 46–53.

[8] Guanjian Cheng, Xin-Gao Gong, and Wan-Jian Yin. 2022. Crystal structure prediction by combining graph network and optimization algorithm. *Nature communications* 13, 1 (2022), 1492.

[9] Maurice Clerc. 2010. *Particle swarm optimization*. Vol. 93. John Wiley & Sons.

[10] Callum J Court, Batuhan Yildirim, Apoorv Jain, and Jacqueline M Cole. 2020. 3-D inorganic crystal structure generation and property prediction via representation learning. *Journal of Chemical Information and Modeling* 60, 10 (2020), 4518–4535.

[11] Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. 2015. Robots that can adapt like animals. *Nature* 521, 7553 (2015), 503–507.

[12] Antoine Cully and Yiannis Demiris. 2018. Quality and Diversity Optimization: A Unifying Modular Framework. *IEEE Transactions on Evolutionary Computation* 22, 2 (2018), 245–259. https://doi.org/10.1109/TEVC.2017.2704781

[13] Stefano Curtarolo, Dane Morgan, Kristin Persson, John Rodgers, and Gerbrand Ceder. 2003. Predicting crystal structures with data mining of quantum calculations. *Physical review letters* 91, 13 (2003), 135503.

[14] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. 2023. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* 5, 9 (2023), 1031–1041.

[15] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. 2023. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* 5, 9 (2023), 1031–1041.

[16] Matthew Fontaine and Stefanos Nikolaidis. 2021. Differentiable quality diversity. *Advances in Neural Information Processing Systems* 34 (2021), 10040–10052.

[17] Scott Fredericks, Kevin Parrish, Dean Sayre, and Qiang Zhu. 2021. PyXtal: A Python library for crystal structure generation and symmetry analysis. *Computer Physics Communications* 261 (2021), 107810.

[18] Adam Gaier, Alexander Asteroth, and Jean-Baptiste Mouret. 2019. Are Quality Diversity Algorithms Better at Generating Stepping Stones than Objective-Based Search? *(GECCO '19)*. Association for Computing Machinery, New York, NY, USA, 115–116. https://doi.org/10.1145/3319619.3321897

[19] Colin W Glass, Artem R Oganov, and Nikolaus Hansen. 2006. USPEX—Evolutionary crystal structure prediction. *Computer physics communications* 175, 11-12 (2006), 713–720.

[20] Conor F Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M Zintgraf, Richard Dazeley, Fredrik Heintz, et al. 2022. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems* 36, 1 (2022), 26.

[21] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.

[22] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. 2013. The materials project: A materials genome approach to accelerating materials innovation, APL Mater. (2013).

[23] Hannah Janmohamed, Thomas Pierrot, and Antoine Cully. 2023. Improving the Data Efficiency of Multi-Objective Quality-Diversity through Gradient Assistance and Crowding Exploration. *arXiv preprint arXiv:2302.12668* (2023).

[24] Ahmed Khalifa, Scott Lee, Andy Nealen, and Julian Togelius. 2018. Talakat: Bullet hell generation through constrained map-elites. In *Proceedings of The Genetic and Evolutionary Computation Conference*. 1047–1054.

[25] Sungwon Kim, Juhwan Noh, Geun Ho Gu, Alan Aspuru-Guzik, and Yousung Jung. 2020. Generative adversarial networks for crystal structure prediction. *ACS central science* 6, 8 (2020), 1412–1420.

[26] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. 2017. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* 29, 27 (2017), 273002.

[27] David C Lonie and Eva Zurek. 2011. XTALOPT version r7: An open-source evolutionary algorithm for crystal structure prediction. *Computer Physics Communications* 182, 11 (2011), 2305–2306.

[28] Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. 2023. Scaling deep learning for materials discovery. *Nature* (2023), 1–6.

[29] Anna Nickelson, Nicholas Zerbel, Gaurav Dixit, and Kagan Tumer. 2023. Shaping the Behavior Space with Counterfactual Agents in Multi-Objective Map Elites. (2023).

[30] Artem R Oganov, Andriy O Lyakhov, and Mario Valle. 2011. How Evolutionary Crystal Structure Prediction Works and Why. *Accounts of chemical research* 44, 3 (2011), 227–237.

[31] Michael O'Keeffe and Bruce G Hyde. 2020. *Crystal structures*. Courier Dover Publications.

[32] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. 2013. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* 68 (2013), 314–319.

[33] Thomas Pierrot, Guillaume Richard, Karim Beguir, and Antoine Cully. 2022. Multi-Objective Quality Diversity Optimization. *arXiv preprint arXiv:2202.03057* (2022).

[34] Han Qi, Xinyang Geng, Stefano Rando, Iku Ohama, Aviral Kumar, and Sergey Levine. 2023. Latent Conservative Objective Models for Data-Driven Crystal Structure Prediction. *arXiv preprint arXiv:2310.10056* (2023).

[35] Chao Qian, Ke Xue, and Ren-Jian Wang. 2024. Quality-Diversity Algorithms Can Provably Be Helpful for Optimization. *arXiv preprint arXiv:2401.10539* (2024).

[36] Kevin Ryan, Jeff Lengyel, and Michael Shatruk. 2018. Crystal structure prediction via deep learning. *Journal of the American Chemical Society* 140, 32 (2018), 10158–10168.

[37] Frank H Stillinger. 1999. Exponential multiplicity of inherent structures. *Physical Review E* 59, 1 (1999), 48.

[38] Richard JD Tilley. 2020. *Crystals and crystal structures*. John Wiley & Sons.

[39] Vassilis Vassiliades, Konstantinos Chatzilygeroudis, and Jean-Baptiste Mouret. 2018. Using Centroidal Voronoi Tessellations to Scale Up the Multidimensional Archive of Phenotypic Elites Algorithm. *IEEE Transactions on Evolutionary Computation* 22, 4 (2018), 623–630. https://doi.org/10.1109/TEVC.2017.2735550

[40] Frank Wilcoxon. 1992. *Individual comparisons by ranking methods*. Springer.

[41] Marta Wolinska, Aron Walsh, and Antoine Cully. 2024. Illuminating the property space in crystal structure prediction using Quality-Diversity algorithms. arXiv:2403.03511 [cond-mat.mtrl-sci]

[42] David H Wolpert and William G Macready. 1997. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation* 1, 1 (1997), 67–82.

[43] Tom Zahavy, Vivek Veeriah, Shaobo Hou, Kevin Waugh, Matthew Lai, Edouard Leurent, Nenad Tomasev, Lisa Schut, Demis Hassabis, and Satinder Singh. 2023. Diversifying AI: Towards Creative Chess with AlphaZero. arXiv:2308.09175 [cs.AI]