LATENT CONSERVATIVE OBJECTIVE MODELS FOR OFFLINE DATA-DRIVEN CRYSTAL STRUCTURE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

In computational chemistry, crystal structure prediction (CSP) is an optimization problem that involves discovering new crystal structures. This problem is challenging for machine learning (ML) methods: it requires discovering globally optimal designs that attain the smallest energy in complex non-Euclidean manifolds. One approach to tackle this problem involves building simulators based on density functional theory (DFT), but these simulators are painfully slow. More recent approaches are exploring the alternate paradigm of relying on learned graph neural networks (GNNs) surrogate models as a proxy for simulation. We propose a method that leverages GNNs to reduce the complexity of the problem. Concretely, we reduce the non-Euclidean optimization search space to a standard vector one with Graph Variational Autoencoders (GVAEs), and we combine that with techniques from offline model-based optimization. This prevents the optimization procedure from producing unstable structures that erroneously appear to have low energies under the learned model. We show that this procedure outperforms current alternatives, both in terms of success rate of structure prediction, and computational cost. In addition, it provides a generic recipe to apply offline optimization techniques for optimizing in non-Euclidean spaces.

1 INTRODUCTION

Data-driven optimization problems arise in many areas of science and engineering. In these settings, we have an unknown function that we would like to optimize with respect to its inputs, provided only with a dataset of input-output pairs. Examples include drug design, where inputs might be molecules and outputs are the efficacy of a drug, protein design, where inputs correspond to protein sequences and outputs are some metric such as fluorescence (Sarkisyan et al., 2016) or, as in our experiments, prediction of crystal structures, where inputs consist of crystal structures and outputs correspond to their formation energy. Such data driven optimization problems present several challenges. First, naïvely training a predictive model to predict the output from the input and then optimizing against such a model may lead to exploitation and adversarial examples: a sufficiently strong optimizer can typically discover inputs that lead any learned model to extrapolate erroneously, and then exploit these errors to find inputs that "fool" the model into making the desired predictions. Second, even if a model can be suitably robustified against such exploitation, many of the most important design and optimization problems in science and engineering, including crystal structure prediction, require optimizing over complex constrained sets and non-Euclidean manifolds, such that naïvely applying gradient-based methods in the input space is unlikely to lead to good results.

In this paper, we study these challenges in the context of crystal structure prediction (CSP) Woodley & Catlow (2008). Crystals are a class of solid-state materials characterized by the periodic placement of atoms. These structures form the basis of a wide variety of applications such as designing super-conductors, batteries Yamashita et al. (2016), and solar cells Walsh et al. (2012). Computationally identifying stable crystal geometries given a particular molecule is an important problem for practitioners, and typically involves minimizing (an estimate of) the crystal's formation energy to find the minimal energy structure. Conventional approaches to this problem rely on slow and compute-intensive DFT simulators Chermette (1998), but more recent machine learning approaches dispense with DFT-based simulators and use databases of structures and their corresponding energies to train models that estimate crystal formation energy directly Gasteiger et al. (2021); Klicpera

et al. (2020). However, the CSP problem suffers from both issues outlined above: crystal structures typically exhibit highly complex geometries characterized by periodicity of the lattice that forms the crystal and discrete (e.g., number and types of atoms in the chemical compound) and continuous features (e.g., positions of atoms in 3D space), which make it hard to produce reliable estimates of energies across the entire manifold of possible structures. Optimizing the structure using such inaccurate models then bears the risk of the optimization procedure "exploiting" these inaccuracies, resulting in structures that erroneously appear promising in this learned model but do not actually attain stable energies. Even if we can avoid such errors locally around existing structures, our goal is to discover *globally* optimal structures reliably, which still remains hard.

In this paper, we aim to develop a data-driven optimization algorithm that can overcome these challenges. First, to avoid the complexities associated with optimization over non-Euclidean spaces, our optimization procedure utilizes advances in variational graph auto-encoders (VGAEs) Kipf & Welling (2016) to convert crystal structures into smooth latent representations, which are much more amenable to simple gradient-based optimization methods. Second, to prevent the optimizer from getting "fooled" by the errors in the learned surrogate model, we apply a robustification procedure to our learned energy prediction function that explicitly pushes down on erroneously over-estimated out-of-distribution inputs in the latent space. Using a combination of these techniques, we develop a method for finding stable crystal structures that alleviates the time and compute costs associated with using DFT simulators, while also addressing the inaccuracies in a purely ML-driven approach for designing crystal structures.

Our main contribution is a data-driven optimization algorithm that operates on the latent space of a variational graph auto-encoder (VGAE), applied to the problem of crystal structure prediction for solid materials. Our method leverages both advances in generative modeling over periodic solid-state materials for latent space learning and recent advances in model-based optimization for robustifying the learned model to make it amenable to direct optimization of formation energies. We summarize the method in Figure 1. We instantiate our method, which we refer to as latent conservative objective models (LCOMs), using crystal diffusion variational auto-encoders (CD-VAE) Xie et al. (2021) for learning the latent space and conservative objective models (COMs) Trabucco et al. (2021b) for optimization. We show that this strategy gives us crystal structures with accuracy of 17/25, using 25 typical compounds for testing. This not only matches the state-of-art method for molecule optimizations, but it also improves on them in terms of computational time. A single optimization cycle in our framework takes an average of 20 minutes and, thanks to the architecture of modern CPUs and GPUs which are highly efficient at vectorized operations, in a single optimization cycle we can optimize over more 100 compounds (in our experiments we worked with a batch size of 256). That allows our model to provide predictions for more than 100 compounds in 20 minutes, which is a timing that no other previous work can compare to.

2 Background and Definitions for Crystal Structures and Materials

In this section, we present the background definitions and terminology associated with crystals and solid-state materials and then discuss the crystal structure prediction problem (CSP) that we study in this work in the next section.

A crystal is a solid-state material characterized by a periodic placement of its constituents, which are chemical elements. The stoichiometry or the composition of a crystal, like NaCl, consists of the elements that make up the solid-state material (i.e., Na and Cl in this case) and in what ratio. In real-world applications of solid-state materials it is not enough to develop a material with a suitable chemical composition, but we must also account for the crystalline periodic structure of the solid and the atoms' positions with respect to it to assess the synthesizability and stability of a given compound.

Mathematically, we can describe the periodic structure of a material via characterizing its lattice L in the 3D space, a set of points that repeat themselves periodically. To characterize a lattice, we define its base vectors $\mathbf{v}, \mathbf{w}, \mathbf{z}$. Every point in the lattice is a linear combination of the three vectors using only integers coefficients.

$$p \in L \iff \exists n, m, k \in \mathbb{Z} \,|\, p = n\mathbf{v} + m\mathbf{w} + k\mathbf{z}. \tag{1}$$

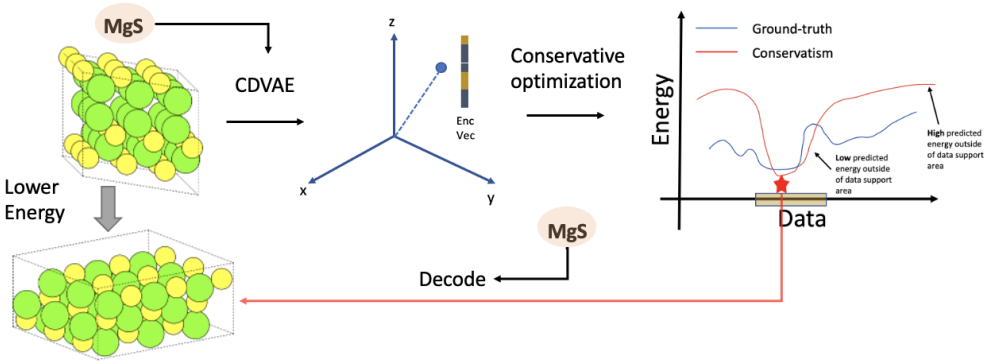

Figure 1: **Overview of LCOMs.** We train a graph-based VAE to construct a latent space the represents crystal structure, conditioned on the molecular structure of the compound. Different points in this latent space correspond to different crystal structures, and we can then optimize the structure with simple gradient-based optimization methods within this latent space. The formation energy is predicted based on the position in the latent space, and the energy prediction model is trained via a conservative optimization strategy that makes it robust on out-of-distribution inputs, preventing the optimizer from discovering latent space points far from the training data for which the energy predictions might extrapolate and yield erroneously low energies. Since the entire optimization is performed in the latent space, the comparatively complex encoder and decoder only need to be used once during optimization (to encode the initial structure guess and decode the final one).

Given a lattice L, the unit cell is the volume of 3D space contained between the base vectors, defined formally as follows:

$$\left\{p \in \mathbb{R}^3 \,|\, \exists x, y, z \in [0, 1]\,,\, p = x\mathbf{v} + y\mathbf{w} + z\mathbf{z}\right\}. \tag{2}$$

We can obtain the global geometry of a crystal by periodically repeating this unit cell in 3D space, which gives rise to the entire lattice.

Given a lattice in 3D space, a crystal is additionally characterized by how many atoms n are in the unit cell. We observe that the number n is always a multiple of the number of elements in the chemical composition of the material. For example, for a formula MgO_3, which consists of 4 atoms, we can have 4, 8, 12, or generally $4k$ elements in a unit cell (one element corresponding to one atom), but not 3 or 6 elements, which are not a multiple of 4.

Finally, we can describe atoms' types and positions with two matrices $A \in \mathbb{R}^{n\times 128}$, $X \in \mathbb{R}^{n\times 3}$. The matrix A identifies different atoms in the unit cell with a one-hot encoding strategy. Specifically, A_i is a vector with a 1 at position Z corresponding to the atomic number of the i-th element, and 0 everywhere else. The matrix X provides fractional coordinates for the atoms. These are coordinates between 0 and 1 with respect to the basis defined by the lattice base vectors. More specifically, the vector X_i tells us that the i-th element is at $X_i^1\mathbf{v} + X_i^2\mathbf{w} + X_i^3\mathbf{z}$ in the unit cell.

To summarize, a crystal is defined by three quantities: **(1)** A 3×3 matrix L representing the crystal lattice, the rows of which corresponding to the base vectors of the lattice; **(2)** number (n) and types ($A \in \mathbb{R}^{n\times 128}$) in the chemical; and **(3)** atoms positions $X \in \mathbb{R}^{n\times 3}$, which is specified in terms of fractional coordinates between 0 and 1.

In the following discussion, we will denote a crystal with the variable $\mathbf{x}$ and we will use subscripts when need to refer, respecitvely, to its lattice parameters $\mathbf{x}_L$, atoms' types $\mathbf{x}_A$, and fractional coordinates $\mathbf{x}_X$.

3 Problem Statement, Dataset and Evaluation

Majority of crystal and lattice configurations for a given chemical compound are "unstable": this means that its formation energy is neither at a local minimum nor a global minimum. Consequentially, if such structure is synthesized, this crystal would have the affinity to quickly collapse to a different, more stable lattice structure. It would be idea if we restricted our study to only stable structures, but such a subspace is complex and not easily characterizable. When we sample a crystal by taking a random lattice matrix and random atoms types and coordinates, the result is almost certainly an unstable design.

Crystal structure prediction (CSP) is the problem of finding a crystal of a given chemical composition (e.g. NaCl, or MgO_3) that is "synthesizable". While it is not yet possible to precisely quantify what makes a crystal synthesizable, it is conventionally believed that a synthesizable crystal is one that forms a global minimum of the formation energy and thus cannot collapse into other structures once synthesized. Therefore, with the goal of discovering a synthesizable crystal, CSP aims to discover such a structure for a chemical composition that attains the smallest formation energy.

[**Crystal structure prediction**] Given a chemical composition c, find the crystal $\mathbf{x}^* = (L^*, A^*, X^*)$ of lattice matrix, atoms' types, atoms' coordinates whose represented crystal globally minimizes the formation energy function E.

$$\mathbf{x}^* = \text{argmin}_{\mathbf{x}} E(\mathbf{x}, c).$$

Why is solving CSPs hard? Note that only one or very few crystal structures are actually synthesizable for a given chemical compound, which makes this equivalent to searching for a "needle in a haystack". The difficulty of solving a CSP is further compounded by the fact that the search space for all possible crystal structures for a given chemical composition is non-Euclidean. This is because there is no one-to-one correspondence between 3×3 matrices and lattices L. Given a lattice matrix L, every other matrix that is rotationally or permutation-equivalent to it represents the same lattice. Moreover, reducing the search space from all possible structures to only stable ones is a procedure that changes the manifold of designs considerably.

We intend to solve this problem using only a static dataset of that shows several (sub-optimal) crystal structures for a variety of chemical compounds along with their corresponding formation energies. With no access to the simulator, our goal is to find a globally optimal crystal structure for a new, previously unseen chemical compound. We describe our procedure for constructing the dataset to train on and our evaluation protocol next.

3.1 Datasets for Training

For training, we use the OQMD dataset (Saal et al., 2013), which consists of the crystal structures and formation energies for more than 1 million materials, obtained via DFT simulations. Every sample of the dataset represents a stable crystal $\mathbf{x}$ which has been obtained through relaxation using the VASP simulator Hafner (2008), together with its chemical composition c and formation energy $E(\mathbf{x}, c)$. We chose this dataset for our task because of its large size, and the availability of more than one stable structure per chemical composition, all of which are not at their global optimum.

3.2 Held-Out Evaluation Datasets

We will evaluate our data-driven, offline optimization approach in terms of its efficacy in discovering the globally optimal structure for a given chemical compound. Hence, to validate the efficacy of our method, we want to construct a held-out test set that consists of globally optimal crystal structures for a held-out set of compounds. Additionally we also want this held-out test dataset to provide us with a locally optimal structure that could be used for kickstarting our iterative offline optimization procedure.

While in principle, we could have utilized a subset of samples from the OQMD dataset or the materials project database (Jain et al., 2013) as the initial designs for our optimization procedure, this however, leads to evaluations that are biased and unreliable. The reason is that samples from these databases are already highly optimized structures and hence, these samples already present strong starting point, that they leave little room for finding an improved crystal structure. To overcome the problem, as part of our work, we have collected a new dataset of stable structures for a variety of different compounds. We utilize these structures serve as our initial samples for kickstarting data-driven optimization. To build this dataset, we used the GPAW Enkovaara et al. (2011) simulator, an open-source DFT simulator fully integrated into python packages for chemistry like ase and pymatgen. We utilized the following protocol for collecting this dataset: **(i)** for every compound, we first select the number of atoms corresponding to the optimal compound design as listed in the materials project database, **(ii)** we then randomly initialize the lattice matrix and the atom coordinates, and **(iii)** we then run the process of structure relaxation in the simulator to obtain the closest local minimum (i.e., a closeby structure that is stable).

3.3 EVALUATION PROTOCOL

The evaluation protocol of our model follows the same procedure as prior works Xie et al. (2021); Cheng et al. (2022). With particular reference to the second work, we test our method on 25 typical compounds of the 29 compounds in Cheng et al. (2022), where the 4 remaining compounds are omitted because they cannot be simulated with GPAW. The compounds are listed in Table 1.

For each of these compounds, we run our optimization process to convergence and check the final energy of the optimized design. We compare the optimized energy against the known global minimum energy. The optimal crystal structure is successfully predicted if the final energy recovers that of the know global minimum, up to a predefined threshold of noise.

4 LCOMS: LATENT CONSERVATIVE OBJECTIVE MODELSA FOR STRUCTURE PREDICTION

To be able to acquire crystal structures with the lowest possible energy entirely from a dataset of previously-conducted energy measurements, we aim to utilize techniques from the area of data-driven offline model-based optimization. Directly applying these techniques for optimizing crystal structures is non-trivial as conventional techniques for offline optimization (Trabucco et al., 2021b; Yu et al., 2021; Qi et al., 2022) typically employ some sort of optimization schemes that iteratively make local changes to the design (e.g., gradient-based updates or random mutations) to optimize a surrogate estimate of the objective function. Such iterative procedures are more prone to fall short when optimizing over non-smooth geometries and non-Euclidean manifolds like that of crystals. To alleviate this issue, in this section, we present a data-driven optimization approach that first aims to learn a latent vector representation for a crystal structure, then performs data-driven optimization in this vector space, and finally maps back the resulting outcome to a valid crystal structure.

4.1 TRANSFORMING CRYSTAL STRUCTURES TO A LATENT SPACE

Which kinds of latent representations are more suitable for data-driven offline optimization for crystals? Since one of the biggest challenges with optimizing crystal structures is that the optimizers are likely to find high-energy, unstable or potentially even infeasible structures, it is very desirable if our latent representations only encode valid and stable structures. Learning such a latent space presents two benefits: **(1)** once we learn such a space, we can directly perform offline optimization in this latent space, and potentially surpass the difficulties of optimizing in the space of crystal structures that lie on complex manifolds, and **(2)** if we can ensure that every possible latent vector corresponds to a stable crystal structure, then we are guaranteed to at least prune out the possibility of finding infeasible structures during the optimzation process.

To this end, we train a crystal diffusion variational auto-encoder (CD-VAE) on a database of stable crystal structures for various chemical compositions and then run offline optimization in the latent space of this auto-encoder. By construction, the decoder of a CD-VAE is guaranteed to decode any latent representation into a valid stable structure, which reduces the fidelity requirement on the optimizer thereby attaining the benefits mentioned above.

Formally, a CD-VAE is composed of three components: a GNN encoder PGNN_{ENC} that takes a crystal $\mathbf{x}$ as input and outputs a latent vector representation, an NN predictor MLP_{AGG} that outputs lattice parameters $\mathbf{x}_L$ from its encoded representation $\text{PGNN}_{\text{ENC}}(\mathbf{x}, c)$, and a GNN diffusion denoiser PGNN_{DEC} that takes a random noisy crystal $\tilde{\mathbf{x}}$ and a latent encoding $\text{PGNN}_{\text{ENC}}(\mathbf{x}, c)$ as inputs, and outputs forces to apply on the atoms coordinates $\tilde{\mathbf{x}}_X$ to build the original crystal $\mathbf{x}$ by a diffusion process. More in detail, the encoder PGNN_{ENC} is a DimeNet++ Klicpera et al. (2020) architecture. The decoder PGNN_{DEC} is a GemNet-dQ Gasteiger et al. (2021) architecture and it works by initializing a structure with random lattice and coordinates, and using Langevyn dynamics Song & Ermon (2019) to gradually recover the stable structure represented by the latent vector.

To make the notation less cumbersome in the following, we will refer to the PGNN_{ENC} as ϕ, and thus the latent representation of a crystal $\mathbf{x}$ with chemical composition c will be denoted by $\phi(\mathbf{x}, c)$. Moreover, with a slight abuse of notation, we will use the notation $\text{PGNN}_{\text{DEC}}(z)$ to denote the structure obtained after applying the denoising process with latent vector z.

The three components are trained in an unsupervised fashion, by looking at the crystal structures $\mathbf{x}$, but not their formation energy. The loss function for training the model is a combination of three quantities $\mathcal{L}_{\text{AGG}}$, $\mathcal{L}_{\text{KL}}$, and $\mathcal{L}_{\text{DEC}}$.

The objective $\mathcal{L}_{\text{AGG}}$ is an MSE between ground-truth and predicted lattice parameters

$$\mathcal{L}_{\text{AGG}}\left(\text{MLP}_{\text{AGG}}\left(\phi(\mathbf{x}, c)\right), \mathbf{x}_L\right) = \left\|\mathbf{x}_L - \phi(\mathbf{x}, c)\right\|^2. \tag{3}$$

The objective $\mathcal{L}_{\text{KL}}$ is the KL divergence loss term for the VAE Kingma & Welling (2013) part of the model

$$\mathcal{L}_{\text{KL}}\left(\phi(\mathbf{x}, c)\right) = D_{\text{KL}}\left(\mathcal{N}\left(\phi(\mathbf{x}, c)\right) \,||\, \mathcal{N}\left(0, 1\right)\right). \tag{4}$$

Where by $\mathcal{N}\left(\phi(\mathbf{x}, c)\right)$ we mean the normal distribution with mean and variance estimated from observations of ϕ for different crystals $\mathbf{x}$.

Lastly, the objective $\mathcal{L}_{\text{DEC}}$ is a denoising loss

$$\mathcal{L}_{\text{DEC}}\left(\tilde{\mathbf{x}}, \mathbf{x}, \phi(\mathbf{x}, c)\right) = \frac{1}{2L}\sum_{j=1}^{L}\left[\mathbb{E}_{\sigma_j}\left\|\text{PGNN}_{\text{DEC}}\left(\tilde{\mathbf{x}}, \phi(\mathbf{x}, c)\right) - \frac{d\left(\mathbf{x}_X, \tilde{\mathbf{x}}_X\right)}{\sigma_j}\right\|\right]. \tag{5}$$

Where $\{\sigma_j\}_{j=1}^{L}$ are scalars in a geometric sequence whose factor is greater than 1.

We remark that the objective functions we have used and reported are slightly different from the ones in the original CD-VAE article, in that they do not include loss terms for the number $\mathbf{x}_n$ and types of atoms $\mathbf{x}_A$. That is because we want our model to be conditional on those quantities, rather than encoding them in the latent representation.

4.2 Conservative Optimization in Latent Space

Once a latent space is learned, we can then train a surrogate model $\widehat{E}_\theta(\phi(\mathbf{x}, c), c)$, via supervised regression, to approximately predict the formation energy E of a crystal structure $\mathbf{x}$ for a given chemical composition, c, by using a neural network parameterized on top of the learned representation $\phi(\mathbf{x}, c)$. However, as prior works (Kumar et al., 2020; Trabucco et al., 2021b) note, this naïve strategy often falls to succeed at finding optimized designs due to the exploitation of errors in the learned surrogate model by the optimization procedure. To address this issue, we utilize the conservative objective models (COMs) technique from Trabucco et al. (2021b) for optimizing crystals in the learned latent space.

Training latent space conservative models. To prevent the optimization procedure from exploiting inaccuracies in this learned surrogate model, we apply an additional regularizer from Trabucco et al. (2021b); Kumar et al. (2021) to robustify the surrogate model. This regularizer mines for adversarial vectors in the latent space, z^+ that appear to have very low energies $\widehat{E}_\theta(z^+, c)$ under the learned surrogate model and regularizes the model $\widehat{E}_\theta(z^+, c)$ output to be larger on such z^+. Following the COMs approach (Trabucco et al., 2021b), we interleave the training of the learned surrogate model $\widehat{E}_\theta$ with an optimization procedure $\mathsf{Opt}(\widehat{E}_\theta, c)$ that seeks to find adversarial vectors z^+ that optimize the current snapshot for the surrogate model, for a given chemical composition c. Once such adversarial examples are obtained, our method regularizes the learned model $\widehat{E}_\theta(z^+, c)$ to explicitly learn higher energy values on such adversarial points. To compensate the effects of increasing the learned values in an unbounded manner, we additionally balance the regularizer by pushing down the energy values on the latent representations induced by crystal structures in the data. This idea can be formalized into the following loss for training $\widehat{E}_\theta$:

$$\begin{aligned}\min_\theta \;\; & \mathbb{E}_{c,\mathbf{x}\sim\mathcal{D}}\left[\left(\widehat{E}_\theta(\phi(\mathbf{x}, c), c) - E(\mathbf{x}, c)\right)^2\right] \\ & - \alpha\left(\mathbb{E}_{c,\mathbf{x}\sim\mathcal{D}}\left[\mathbb{E}_{z^+\sim\mathsf{Opt}(\widehat{E}_\theta, c)}[\widehat{E}_\theta(z^+, c)] - \widehat{E}_\theta(\phi(\mathbf{x}, c), c)\right]\right).\end{aligned} \tag{6}$$

We will discuss the precise formulation for Opt below.

Optimizing in the latent space. Once a conservative surrogate model $\widehat{E}_\theta(z, c)$ is obtained using the above training procedure, we must now optimize this model to obtain learned structures. The

optimization procedure, Opt that was used to obtain adversarial latent vectors in Equation 6 can be repurposed to obtain optimized latent vectors once the latent conservative model is trained. Since the latent space z is a continuous Euclidean vector space, for any given chemical composition c, our choice of Opt is to run T rounds of gradient descent on the surrogate energy $\widehat{E}_\theta(z, c)$ with respect to the latent vector z starting from the latent vector z_0 obtained by encoding a randomly chosen crystal structure from the dataset. For a given c, this procedure can be formalized as follows:

$$z_{k+1} \leftarrow z_k - \alpha \nabla_z \widehat{E}_\theta(z, c), \quad \text{where} \quad z_0 \sim \phi(\mathbf{x}_0, c), \;\; \mathbf{x}_0 \sim \mathcal{D}. \tag{7}$$

Once we run this optimization procedure is run for T steps, we pass the final latent vector z_T to the decoder of the pre-trained CD-VAE to obtain the optimized crystal structure: $\widehat{\mathbf{x}}^* = \text{PGNN}_{\text{DEC}}(z_T)$.

Implementation details. The encoder and decoder are trained independently from the optimization model. DimeNet++ adapted for periodicity is used for the encoder and GemNet-dQ is used for for the decoder. The training for encoder/decoder is followed with the implementation in Xie et al. (2021). Then the molecule structures from dataset OQMD are encoded to vectors, as inputs for the optimization model training. For training LCOMs, we model a conservative objective model $\widehat{E}_\theta(\phi(\mathbf{x}, c), c)$ as a neural network with two hidden layers of size 2048 each and leaky ReLU activations. The optimizer choice is followed in Trabucco et al. (2021b). When training the $\widehat{E}_\theta(\phi(\mathbf{x}, c), c)$, we follow the loss function in Equation 6. For computing z^+, we do one gradient descent step over vector z from input latent space. Finally, for optimization, we encode a group of molecules with certain compositions and stable but maybe not optimal structures. We do 50 gradient steps and get a group of optimized vectors representing optimal structures in latent space. The decoder is then used to decode them, producing corresponding optimal molecules with the same compositions as inputs but better structures which achieve lower energy.

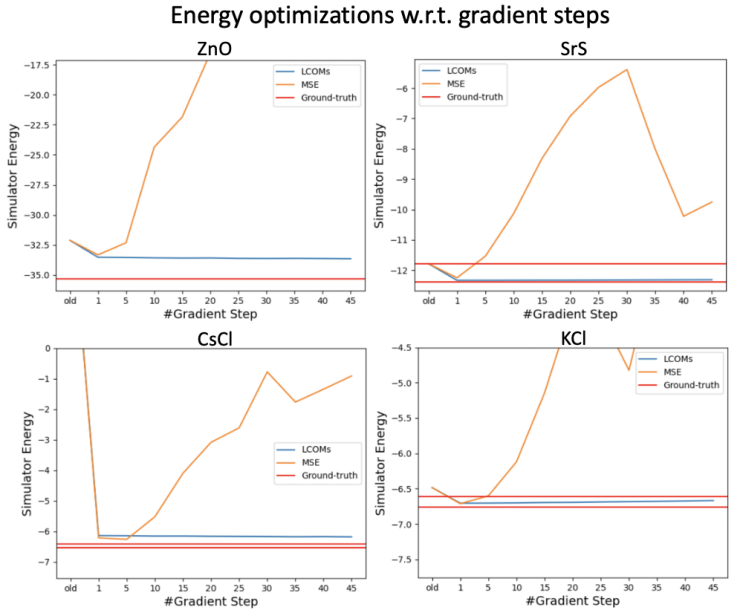

Figure 2: Optimization curves comparing LCOMs, MSE, and ground-truth energy for known structures. LCOMs (blue curve) enables the optimized structures to achieve lower energies after gradient descent, and the final structures are approaching the ground-truth optimal results. Gradient descent with the non-conservative MSE model quickly increases the true energy values, as the optimizer exploits flaws in the model.

5 Related Work

Our work combines methods from offline model-based optimization, representation learning with graph neural networks, and crystal structure prediction (CSP). Before the widespread use of modern deep networks, one widely studied optimization approach to CSP focused on genetic algorithms Lonie

Compounds	RAS	PSO	BO	LCOMs	MSE
LiF		✓	✓	✓	
NaF	✓	✓	✓		✓
KF	✓		✓	✓	
RbF				✓	✓
LiCl					
NaCl	✓	✓	✓	✓	✓
KCl	✓			✓	
RbCl	✓			✓	
CsCl	✓		✓	✓	
BeO		✓	✓	✓	
MgO	✓		✓		
CaO	✓		✓	✓	
SrO	✓		✓	✓	
BaO	✓		✓	✓	
ZnO		✓	✓	✓	
CdO					
BeS	✓	✓	✓	✓	
MgS	✓			✓	✓
CaS	✓		✓	✓	
SrS	✓			✓	✓
BaS	✓		✓		
ZnS	✓		✓	✓	
CdS			✓		
C	✓				
Si					
Accuracy	17/25	6/25	16/25	17/25	5/25

Table 1: Comparison of LCOMs (our approach) and prior crystal structure prediction methods for 25 typical compounds. Check marks indicate successful recovery as per the criteria in the text.

& Zurek (2011); Oganov et al. (2011). One important work in that direction is USPEX Glass et al. (2006), an algorithm that uses three variation strategies for evolving new structures at each generation. Heredity takes parts of two structures and combines them; mutation applies a slight variation to the lattice parameters of a single structure; permutation exchanges atoms' types and positions. During the evolutionary process in USPEX, every candidate is relaxed with simulators to guarantee that the search space operates on only stable structures. The extensive use of simulation makes such approaches computationally very expensive, and is part of the reason why more recent works have sought to explore ML-based techniques.

More recent work has applied ML techniques and deep network to CSP, in part spurred by the availability of public datasets like the Materials Project Database, or the Open Catalyst Project Chanussot et al. (2021). In terms of methods, Cheng et al. (2022) is particularly relevant, being an overview of different optimization algorithms built on top of a GNN predictor: random search, particle swarm optimization Clerc (2010), and Bayesian optimization Pelikan et al. (1999). The main difference from our work is that the authors leverage GNNs only as a surrogate model for estimating energy. The sampling of crystals in the search space is performed independently and depending on the optimization strategy employed.

The work of Kim et al. (2020) is an approach to the problem based on GANs rather than VAEs. However, this work proposes a generative model, but not an optimization strategy. The querying of a property like formation energy is still done with DFT simulators.

A number of other works have investigated GNNs and other related models in the context of computational chemistry. Much of this research studies prediction of formation energy with GNNs, together with some other properties when available in the training dataset. The main architectures that we cite are the Crystal Graph Convolutional Network Xie & Grossman (2018), DimeNet Gasteiger et al. (2020), DimeNet++ (which we use as the encoder for the CD-VAE) Klicpera et al. (2020), and GemNet (which we use as the decoder for the CD-VAE) Gasteiger et al. (2021). A significant

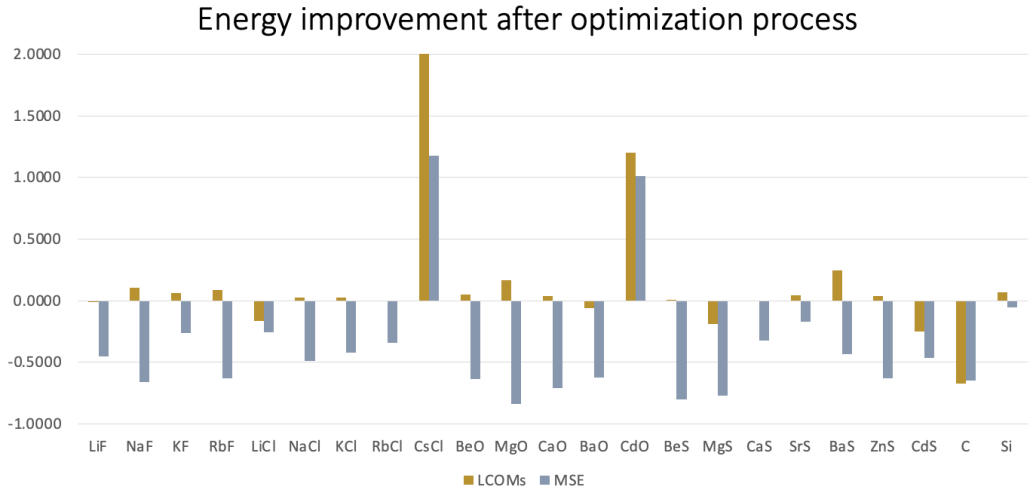

Figure 3: Comparisons of energy improvements achieved with LCOMs and the non-conservative MSE baseline. Note that with LCOMs, most structures achieve better formation energy after optimization, while the non-conservative MSE model actually leads to structures with worse energy values (negative improvement). These results indicate that conservative training is essential for successfully instantiating a method with gradient-based latent space optimization for CSP.

distinction with our work is that we focus not on proposing a new architecture for modeling formation energy, but an algorithm for optimizing crystal structures that combines state-of-the-art architectures and conservative model-based optimization algorithms.

Model-based optimization (MBO) refers to the problem of optimizing an unknown function by constructing a surrogate model. Bayesian optimization represents one of the most widely known classes of MBO methods (Snoek et al., 2015), but classically MBO methods require iteratively sampling new function values, which can be very expensive when evaluating a crystal structure's energy requires an expensive simulation process. More recently, offline MBO methods, sometimes referred to as data-driven optimization, have been proposed to optimize designs based entirely on previously collected *static* datasets (Brookes et al., 2019; Trabucco et al., 2021a; Kumar & Levine, 2019; Trabucco et al., 2021b). Our work builds on these methods, and is most closely related to the COMs algorithm proposed by Trabucco et al. (2021b). However, while prior offline MBO methods focus on robustifying the surrogate model while optimizing in the original design space, we integrate these approaches with a latent space optimization method that makes it possible to optimize over the complex and non-Euclidean design space of crystal structures, while using simple and efficient gradient descent methods.

6 Experiments

In this section, we provide an empirical evaluation of LCOMs for crystal structure prediction. In designing the experiments, we aim at answering to the following questions: (1) Can LCOMs successfully optimize in the latent representation space? (2) Does LCOMs manage to recover the optimal energy up to a pre-defined threshold of accuracy? (3) How do the hyperparameters of LCOMs influence its performance? To answer these questions, we evaluate LCOMs as well as baseline methods following the protocol detailed in Section 3.

Comparing LCOMs with baselines and prior methods. We compare LCOMs with three methods from prior work Cheng et al. (2022), including random searching (RAS), particle-swarm optimization (PSO) and Bayesian optimization (BO). We also include one additional baseline method, where the design is optimized with gradient descent in the latent space of the CD-VAE, as in our method, but the energy prediction model is trained without any conservatism, but rather with standard mean-squared error (MSE), to evaluate the importance of conservative training. We evaluate these methods on the 25 chemical compounds in our evaluation dataset, and present the results in Table 1. We mark each crystal structure as recovered successfully if the energy reported for the recovered structure is close

to the energy of the ground truth structure according to the simulator, with the threshold defined as $(E(\mathbf{x}, c^*) - E(\mathbf{x}, \hat{c}))\,/|E(\mathbf{x}, c^*)|$ to account for imprecision in the simulator, where c^* is the ground truth optimal crystal and $\hat{c}$ is the model optimized crystal.

Our method optimizes the crystal structure in the latent space of the CD-VAE, using simple gradient descent methods. One advantage of this approach is computational efficiency, since the complex graph-based component of the pipeline is only used during the encoding and decoding stages at the beginning and end of the optimization, rather than at each optimization step. The conservatism term in LCOMs ensures that this process does not produce out-of-distribution adversarial inputs into the formation energy prediction model. The results in Table 1 show that our method improves significantly over naïvely optimizing in the CD-VAE latent space without conservative training (MSE), and also that it is competitive with the prior state-of-the-art methods RAS, PSO, and BO, exceeding the performance of the PSO and BO baselines and matching RAS. Considering the benefit of efficiency for our method, it has huge potential usage in molecule structure optimization.

How does conservative training influence optimization? In the next set of experiments, we aim to understand how conservative training of the formation energy prediction model influences the behavior of gradient descent optimization in the CD-VAE latent space. If the energy model is trained naïvely (i.e., with standard empirical risk minimization methods), it will be liable to make arbitrarily erroneous predictions on out-of-distribution inputs that differ too much from the training data. Some of these errors will be positive, and some negative, and therefore a strong optimizer would be able to exploit these errors to find points in the latent space for which the model erroneously predicts arbitrarily favorable energies. To compare how conservative and naïve models behave during latent space optimization, we train both types of models, and then perform 50 gradient steps on the learned energy models starting from an initial structure. Although our full method only decodes the structure at the end, in this experiment we save out the decoded structure after every gradient step for analysis, and evaluate their energy according to the simulator (which we treat as the ground truth). The results are shown in Figure 3, where improvement is given by $(E(\mathbf{x}, c_0) - E(\mathbf{x}, \hat{c}))\,/|E(\mathbf{x}, c_0)|$, where c_0 denotes the initialization structure and $\hat{c}$ denotes the model optimized structure, and the larger improvement values are better.

The results in Figure 3 show that LCOMs generally produces positive improvement, while the non-conservative MSE model leads to *negative* improvement – i.e., the optimized structures are generally worse than the starting point. This indicates that conservative training is critical for latent space optimization to work. To provide more fine-grained analysis, we plot the energy for each method at each of the 50 gradient steps for a few example compounds in Figure 2. The orange line, representing the non-conservative MSE model, shows that the energy increases over the course of optimization (lower is better), while the blue curve corresponding to LCOMs is much closer to the result from optimizing the ground-truth energy (shown in red).

7 Discussion

In this work, we presented a method for data-driven optimization that uses the latent space of a variational graph autoencoder to perform smooth gradient-based optimization of complex structures, with application to crystal structure prediction. Our method combines concepts from conservative objective models, which robustify predictive models to make them amenable to gradient-based optimization over their inputs, with variational graph autoencoders, which provide us with a latent space over crystal structures that overcomes the complex geometry of the design space, enabling simple gradient-based methods to be used. Our method can optimize crystal structures after training on a static dataset of previously collected structure-energy tuples.

Empirical evaluation shows that our method can successfully optimize the formation energy and recover the optimal structure of a chemical compound with a good level of accuracy, comparing favorably with existing approaches. Moreover, aside from the training dataset, our pipeline is based entirely on deep neural networks and does not make use of simulators or other external models. This enables our method to perform optimization in a much more rapid and parallelizable way, effectively leveraging the compute capacity of modern hardware.

REFERENCES

David Brookes, Hahnbeom Park, and Jennifer Listgarten. Conditioning by adaptive sampling for robust design. In *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 2019. URL http://proceedings.mlr.press/v97/brookes19a.html.

Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021.

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

H Chermette. Density functional theory: a powerful tool for theoretical studies in coordination chemistry. *Coordination chemistry reviews*, 178:699–721, 1998.

Maurice Clerc. *Particle swarm optimization*, volume 93. John Wiley & Sons, 2010.

Jussi Enkovaara, Nichols A Romero, Sameer Shende, and Jens J Mortensen. Gpaw-massively parallel electronic structure calculations with python-based software. *Procedia Computer Science*, 4:17–25, 2011.

Johannes Gasteiger, Janek Gro
, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34: 6790–6802, 2021.

Colin W Glass, Artem R Oganov, and Nikolaus Hansen. Uspex—evolutionary crystal structure prediction. *Computer physics communications*, 175(11-12):713–720, 2006.

Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1):011002, 2013.

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

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

Johannes Klicpera, Shankari Giri, Johannes T Margraf, and Stephan Günnemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020.

Aviral Kumar and Sergey Levine. Model inversion networks for model-based optimization. *NeurIPS*, 2019.

Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. Conservative q-learning for offline reinforcement learning. *arXiv preprint arXiv:2006.04779*, 2020.

Aviral Kumar, Amir Yazdanbakhsh, Milad Hashemi, Kevin Swersky, and Sergey Levine. Data-driven offline optimization for architecting hardware accelerators. *arXiv preprint arXiv:2110.11346*, 2021.

David C Lonie and Eva Zurek. Xtalopt: An open-source evolutionary algorithm for crystal structure prediction. *Computer Physics Communications*, 182(2):372–387, 2011.

Artem R Oganov, Andriy O Lyakhov, and Mario Valle. How evolutionary crystal structure prediction works and why. *Accounts of chemical research*, 44(3):227–237, 2011.

Martin Pelikan, David E Goldberg, Erick Cantú-Paz, et al. Boa: The bayesian optimization algorithm. In *Proceedings of the genetic and evolutionary computation conference GECCO-99*, volume 1, pp. 525–532. Citeseer, 1999.

Han Qi, Yi Su, Aviral Kumar, and Sergey Levine. Data-driven offline decision-making via invariant representation learning. *arXiv preprint arXiv:2211.11349*, 2022.

James E Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and Christopher Wolverton. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd). *Jom*, 65:1501–1509, 2013.

Karen S. Sarkisyan, Dmitry A. Bolotin, Margarita V. Meer, Dinara R. Usmanova, Alexander S. Mishin, George V. Sharonov, Dmitry N. Ivankov, Nina G. Bozhanova, Mikhail S. Baranov, Onuralp Soylemez, Natalya S. Bogatyreva, Peter K. Vlasov, Evgeny S. Egorov, Maria D. Logacheva, Alexey S. Kondrashov, Dmitry M. Chudakov, Ekaterina V. Putintseva, Ilgar Z. Mamedov, Dan S. Tawfik, Konstantin A. Lukyanov, and Fyodor A. Kondrashov. Local fitness landscape of the green fluorescent protein. *Nature*, 533(7603):397–401, May 2016. ISSN 1476-4687. doi: 10.1038/nature17995. URL https://doi.org/10.1038/nature17995.

Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat, and Ryan Adams. Scalable bayesian optimization using deep neural networks. In *Proceedings of the 32nd International Conference on Machine Learning*. PMLR, 2015.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf.

Brandon Trabucco, Xinyang Geng, Aviral Kumar, and Sergey Levine. Design-bench: Benchmarks for data-driven offline model-based optimization, 2021a. URL https://github.com/brandontrabucco/design-bench.

Brandon Trabucco, Aviral Kumar, Xinyang Geng, and Sergey Levine. Conservative objective models for effective offline model-based optimization. In *International Conference on Machine Learning*, pp. 10358–10368. PMLR, 2021b.

Aron Walsh, Shiyou Chen, Su-Huai Wei, and Xin-Gao Gong. Kesterite thin-film solar cells: Advances in materials modelling of cu2znsns4. *Advanced Energy Materials*, 2(4):400–409, 2012.

Scott M Woodley and Richard Catlow. Crystal structure prediction from first principles. *Nature materials*, 7(12):937–946, 2008.

Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters*, 120(14), apr 2018. doi: 10.1103/physrevlett.120.145301. URL https://doi.org/10.1103%2Fphysrevlett.120.145301.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

Tomoki Yamashita, Hiroyoshi Momida, and Tamio Oguchi. Crystal structure predictions of naxc6o6 for sodium-ion batteries: First-principles calculations with an evolutionary algorithm. *Electrochimica Acta*, 195:1–8, 2016.

Sihyun Yu, Sungsoo Ahn, Le Song, and Jinwoo Shin. Roma: Robust model adaptation for offline model-based optimization. *Advances in Neural Information Processing Systems*, 34, 2021.