Artificial Intelligence Simulation and Design of Energy Materials with Targeted Properties

uOttawa

by

Ericsson Tetteh Chenebuah

Thesis submitted to the University of Ottawa
in conformity with the requirements for the degree of

Doctor of Philosophy (Ph.D.) in Mechanical Engineering

Ottawa-Carleton Institute for Mechanical and Aerospace Engineering
Department of Mechanical Engineering
University of Ottawa
Ottawa, Canada.

Abstract

The discovery of new energy materials is fundamental to addressing numerous technological challenges. At the forefront of discoverable materials is the perovskite crystal structure, which is an emerging multifunctional material class with several attractive properties that are utilized in many engineering applications. They include high ionic conductivity for solar cells, good dielectric response for piezoelectric devices, and strong catalytic activity for batteries or hydrogen production. Considering both the chemical flexibility of perovskite stoichiometries and the possibility of polymorphism, the estimated number of perovskite compounds could potentially exceed ten million. Conventional discovery techniques often involve Edisonian-based synthetic chemistries and/or first-principles quantum mechanical calculations. Although such techniques have achieved substantial successes in the past, their application can be difficult, unpractical, uneconomical and computationally expensive for complex systems and extremely large chemical search spaces. In this regard, the current thesis explores Artificial Intelligence (AI) as potentially a more reliable, inexpensive and rapid alternative over conventional techniques. In the process of making such discovery, two design schemes are sequentially addressed, namely forward design simulation and inverse design simulation. The forward design is focused on simulating the structure → property relationship by accurately mapping perovskite materials to their deterministic target properties. Conversely, the inverse design simulation is aimed at simulating the property → structure relationship by discovering novel perovskite materials that possess deterministic target properties of interest. In the present research, the forward design is addressed by developing a hybridized Deep Learning (DL) framework comprised of a two-dimensional convolutional neural network (Conv2D) model and a support vector machine (SVM) model. The hybridized Conv2D-SVM approach is demonstrated to out-perform periodic benchmark representations in the Coulomb matrix, Ewald-sum matrix, and Sine matrix by about 70%, 75% and 66% on stability energy, formation energy and bandgap targets, respectively. In addressing the inverse design, three Deep Generative Modelling (DGM) pipelines are designed: (1) Target-Learning Variational Autoencoder (TL-VAE) model; (2) Evolutionary Variational Autoencoder for Perovskite Discovery (EVAPD) model; and (3) Lattice-Constrained Materials Generative Model (LCMGM). In total, 265 new perovskite materials are designed using the developed DGM pipelines, of which about 60% are newly discovered (i.e. unique and novel) chemical compositions, while the remainders are polymorphs of already known chemical compositions. The new materials are validated using Density Functional Theory (DFT) technique and are openly archived in materials database repositories. Overall, the current research demonstrates efficient AI innovative pathways for advancing the rapid search for perovskite energy materials. The newly discovered perovskites are the subject of ongoing follow-up research on their synthesization, characterization and testing.

Main Contributions of the Thesis

The current thesis is paper based, presented as a series of five original research articles that have been published (or currently being considered for publication) in peer-reviewed scientific journals and an academic conference. They are all fully and exclusively based on my Ph.D. research work, and feature me as first and leading author, with co-authorship contribution from my supervisors: Dr. Michel Nganbe[1] and Dr. Alain Beaudelaire Tchagang[1,2]. The articles are presented and included in Chapters 4 and 5. They are listed as follows:

1. Comparative analysis of machine learning approaches on the prediction of the electronic properties of perovskites: A case study of ABX_3 and $A_2BB'X_6$. *Materials Today Communications*, 27 (2021) 102462. https://doi.org/10.1016/j.mtcomm.2021.102462

2. A Fourier-transformed feature engineering design for predicting ternary perovskite properties by coupling a two-dimensional convolutional neural network with a support vector machine (Conv2D-SVM). *Materials Research Express,* 10 (2023) 026301. https://doi.org/10.1088/2053-1591/acb683

3. Target-learning the latent space of a variational autoencoder model for the inverse design of stable perovskites. *In the 36th proceedings of the Canadian Conference on Artificial Intelligence*, (2023). https://doi.org/10.21428/594757db.07402193

4. An evolutionary variational autoencoder for perovskite discovery. *Frontiers in Materials*, (2023) 10:1233961. https://doi.org/10.3389/fmats.2023.1233961

5. A deep generative modelling architecture for designing lattice-constrained perovskites. (*Journal article under review for publication*). Mendeley data repository: https://doi.org/10.17632/m262xxpgn2.1

[1] Department of Mechanical Engineering, University of Ottawa, 161 Louis-Pasteur, Ottawa, ON, K1N 6N5, Canada.

[2] Digital Technologies Research Center, National Research Council of Canada, 1200 Montréal Road, Ottawa, ON, K1A 0R6, Canada.

Acknowledgments

I wish to first acknowledge my supervisors Dr. Michel Nganbe and Dr. Alain Tchagang, for their unwavering support throughout the course of my doctorate degree journey. In particular, I am grateful to Dr. Nganbe – a person who I've come to deeply admire – for his quality inspection on all my research documents, in addition to his administrative role in ensuring I fully concentrate on my research. I am equally grateful to Dr. Tchagang for proposing the thesis topic and integrating me into the Digital Technologies Research Center of the National Research Council (NRC) of Canada. The support I received at NRC considerably aided my research progress, especially with respect to using their High Performance Computing (HPC) clusters for simulating most of my complex calculations. For this purpose, I also wish to extend my gratitude to the head of the Scientific Data Mining (SDM) team at NRC, Dr. Miroslava Cuperlovic-Culf and research support Ms. Anu Surendra. I am further grateful to the Department of Mechanical Engineering at the University of Ottawa and the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support in the form of Teaching Assistantships and graduate admission scholarships. Moreover, I am sincerely grateful to each member of my PhD examination committee for reviewing my thesis and providing suggestions that are intended to further improve the quality of the thesis. Finally, the entirety of this research is dedicated to my parents: Samuel and Christie for bringing me up in a practicing Christian and God-fearing home. They have been a huge inspiration as it was always their dream to see me complete my doctorate.

Table of Contents

List of abbreviations - - - - - - - - - vii

List of tables and figures - - - - - - - - - viii

CHAPTER 1: INTRODUCTION - - - - - - - - 1

1.1 Background overview - - - - - - - - - 1

1.2 Purpose of research - - - - - - - - - 4

1.3 Novelty of research - - - - - - - - - 5

1.4 Scope of research - - - - - - - - - 7

1.5 Research outline - - - - - - - - - 8

CHAPTER 2: SUMMARISING LITERATURE REVIEW - PEROVSKITE ENERGY MATERIALS - - - - - - - - - - - - - 10

2.1 The perovskite crystal structure - - - - - - - - 10

2.1.1 Ternary perovskites of the ABX_3 stoichiometry - - - - 10

2.1.2 Double perovskites of the $A_2BB'X_6$ and $AA'BB'X_6$ stoichiometries - - 12

2.2 Perovskite multifunctionality and applications - - - - - 14

2.3 Perovskite determination via first-principles (*ab initio*) techniques - - - 15

2.4 Perovskite synthesization and characterization - - - - - 16

CHAPTER 3: OVERALL METHODOLOGY – PEROVSKITE DESIGN SPACE AND AI APPLICATION - - - - - - - - - 19

3.1 Chemical combinatorial design space for perovskite discovery - - - 19

3.2 AI simulation for perovskite discovery - - - - - - 20

3.2.1 Simulating the forward design using supervisory models - - - 21

3.2.2 Simulating the inverse design using variational autoencoders - - 23

3.2.3 Simulating the inverse design using generative adversarial networks - 25

3.3 Spherical Linear Interpolation (SLERP) for sampling operation - - - 26

3.4 Genetic algorithm for multi-objective target-search optimization - - - 27

3.5 Feature engineering for multi-stoichiometrical perovskite representation - - 29

3.6 Data sources and preprocessing measures - - - - - 30

3.7 Density Functional Theory (DFT) validation and Bayesian optimization - - 32

CHAPTER 4: SCIENTIFIC PUBLICATIONS ON THE FORWARD DESIGN SIMULATION 36

4.1 Improving the prior art for higher target property prediction accuracy - - 36

4.2 Investigations on target property prediction for perovskite structures- - - 38

4.2.1 Journal publication 1 – Comparative analysis of different ML approaches on the prediction of the electronic properties of perovskites: A case study of ABX_3 and $A_2BB'X_6$– Published in *Materials Today Communications* - - - 38

4.2.2 Journal publication 2 – Fourier-transformed feature engineering design for predicting ternary perovskite properties by coupling a two-dimensional convolutional neural network with a support vector machine (Conv2D-SVM) – Published in *Materials Research Express* - - - - 40

CHAPTER 5: SCIENTIFIC PUBLICATIONS ON THE INVERSE DESIGN SIMULATION- 42

5.1 Improving the prior art on generative inverse modelling for materials discovery - 42

5.2 Investigations on inverse design pipelines for novel perovskite discovery - - 43

5.2.1 Conference publication – Target-learning the latent space of a variational autoencoder model for the inverse design of stable perovskites – Published in the *Proceedings of the 36th Canadian Conference on AI* - - - 45

5.2.2 Journal Publication 3 – An evolutionary variational autoencoder for perovskite discovery – Published in *Frontiers in Materials* - - - - 46

5.2.3 Journal Publication 4 – A deep generative modelling architecture for designing lattice-constrained perovskites – Journal article *under review for publication*- 48

CHAPTER 6: OVERALL DISCUSSION - - - - - - 50

6.1 Newly discovered perovskites and property determination - - - 50

6.1.1 New perovskites generated by the TL-VAE model - - - 51

6.1.2 New perovskites generated by the EVAPD model - - - 52

6.1.3 New perovskites generated by the LCMGM model - - - 53

6.2 Application in photovoltaics and optoelectronics - - - - - 55

CHAPTER 7: CONCLUSION AND FUTURE RESEARCH - - - - 58

7.1 Conclusion - - - - - - - - - 58

7.2 Future research - - - - - - - - - 60

References - - - - - - - - - - 62

Appendix - - - - - - - - - - 74

List of Abbreviations

A-GAN	Auxiliary Generative Adversarial Network	**LCMGM**	Lattice-Constrained Materials Generative Model
AI	Artificial Intelligence	**MAE**	Mean Absolute Error
AUC	Area Under Curve	**ML**	Machine Learning
BO	Bayesian Optimization	**MLP**	Multi-Layer Perceptron
CBM	Conduction Band Minimum	**MP**	Materials Project
CGCNN	Crystal Graph Convolutional Neural Network	**MSE**	Mean Squared Error
CIF	Crystallographic Information File	**NOMAD**	Novel Materials Discovery
Conv2D/CNN	Convolutional Neural Network	**OQMD**	Open Quantum Materials Database
CS	Crystal System	**PAW**	Projector Augmented Wave
DFT	Density Functional Theory	**PBE**	Perdew-Burke-Ernzerhof
DGM	Deep Generative Model	**PCA**	Principal Component Analysis
E_f	Formation Energy	**PCE**	Power Conversion Efficiency
E_g	Energy bandgap	**QE**	Quantum Espresso
E_s/E_{hull}	Stability Energy/Energy above convex hull	**R^2**	Coefficient of Determination
E_{rel}	DFT-computed Relative energy	**ROC**	Receiver Operating Characteristics
E_{tot}	DFT-computed Total energy	**RMSE**	Root Mean Squared Error
EVAPD	Evolutionary Variational Autoencoder for Perovskite Discovery	**SLERP**	Spherical Linear Interpolation
FTCP	Fourier Transformed Crystal Property	**SS-VAE**	Semi-Supervisory Variational Autoencoder
GA	Genetic Algorithm	**SVM**	Support Vector Machine
GAN	Generative Adversarial Network	**t-SNE**	t-Distributed Stochastic Neighbor Embedding
GGA	Generalized Gradient Approximation	**TL-VAE**	Target-Learning Variational Autoencoder
HOIP	Hybrid Organic-Inorganic Perovskite	**VAE**	Variational Autoencoder
ICSD	Inorganic Crystal Structure Database	**VASP**	Vienna Ab initio Simulation Package
KL	Kullback-Leibler Divergence Function	**VBM**	Valence Band Maximum

List of Tables and Figures

List of Tables

Table 1: Perovskite crystal compounds and their respective applications - - - 15

Table 2: Inorganic materials database and their available data sizes - - - 31

Table 3: Benchmark evaluation from past studies as related to formation energy and band gap predictions for general inorganic crystalline structures - - - 37

Table 4: Comparison of developed DGMs for perovskite discovery - - - 50

Table 5: Special class of materials with potential serviceability as host perovskites for photovoltaic applications - - - - - - - 56

List of Figures

Fig. 1: Schematic of modelling flow for the forward versus inverse design simulative approaches 6

Fig. 2: An ideal cubic perovskite crystal structure of the ABX_3 stoichiometry - - 11

Fig. 3: The double perovskite crystal structure - - - - - - 13

Fig. 4: Frequency-modelled periodic table from the preprocessing of data - - 20

Fig. 5: Interested region in the latent space for conducting SLERP sampling operation - 27

Fig. 6: Integrated Genetic Algorithm (GA) for ranking high-quality perovskite candidates - 29

Fig. 7: Graphical abstract on the comparative analysis of machine learning approaches on the prediction of the electronic properties of perovskites - - - 40

Fig. 8: Graphical abstract on the Fourier-transformed feature engineering design for predicting ternary perovskite properties using the Conv2D-SVM setup - - 41

Fig. 9: Modelling architecture of the Target-Learning Variational Autoencoder (TL-VAE) model 44

Fig. 10: Modelling architecture of the Evolutionary Variational Autoencoder for Perovskite Discovery (EVAPD) model - - - - - - 47

Fig. 11: Modelling architecture of the Lattice-Constrained Materials Generative Model (LCMGM) 49

Fig. 12: New perovskite candidates from the TL-VAE inverse design pipeline - - 52

Fig. 13: Comparison of geometrical coordination between a newly generated Ca_2YOsO_6 perovskite by the EVAPD model and subsequent DFT optimization/relaxation - 53

Fig. 14: LCMGM designed perovskites with the best model-predicted formation energy with respect to the considered crystal systems and stoichiometries - - 54

Fig. 15: Band-structure and Projected Density of States (P-DOS) for three LCMGM-designed perovskites with potential serviceability as host materials in solar cells - - 57

CHAPTER 1: INTRODUCTION

1.1 Background overview

Traditionally, discovering new materials is known to be possible by experimental and/or first-principles deterministic means. Depending on the desired property of interest, experimental techniques (i.e. synthetic chemistries) normally begin with the initial fabrication of the material samples, followed by mechanical or functional testing. First-principles (or *ab initio*) deterministic methods, on the other hand, are proven substitutes equally used in the discovery and determination of a material's properties. Despite both techniques having achieved huge successes in the past [49, 79, 82, 84, 119], they have considerable limitations, which can hinder their generalizability or applicability. For instance, experimental synthesization is typically Edisonian-based (i.e. trial-and-error) and requires a vast amount of resources and considerable time. On the other hand, direct first-principles deterministic methods will strongly rely on the systematic screening of promising candidates by performing noble ionic substitutions on new chemical combinations. Such screening processes are typically high-throughput quantum-mechanically demanding and, as such, will involve solving the costly Schrödinger equation on a many-body Hamiltonian system. Moreover, first-principles techniques, such as Density Functional Theory (DFT), incorporate some underlying challenges, including choosing the right exchange correlation functional, assuming reaction barriers and van der Waals interaction, addressing delocalization and static correlation errors, among others [1]. With the advent of big data and supercomputing, Artificial Intelligence (AI) methods are now gaining significant interest among scientists and engineers, and are proven to tackle cumbersome challenges associated with traditional materials discovery approaches. AI principles are purely data-driven, which makes them fast, intrinsic and low-cost. This has led to AI's exponential growth in several applications, including organic drug design [2], autonomous vehicles [3], robotics [4], object/image recognition [5], speech detection [6], fraud detection [7], and many more. Examples of AI branches include Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), Computer Vision, and Reinforcement Learning (RL).

Today, Machine Learning (ML) plays a crucial role in understanding numerous scientific phenomena and filling gaps between experimental and theoretical approaches by learning from

well-curated datasets. As specific to computational materials sciences, ML seeks to solve modelling challenges in two broad ways: (1) *the forward design approach* and (2) *the inverse design approach*. In the forward design approach, relevant target properties are accurately predicted by using features that best describe an objective material. In doing so, a direct mapping function is learnt, which assimilates hidden relational behaviours by correlating input features to unknown target properties. The prediction performance of the mapping function is evaluated using comparative standardized metrics between the real targets (from a held-out test set) and their corresponding predicted outcomes. Examples of some interesting literature that applies the forward design approach in computational material sciences include the prediction of deterministic target properties for solid-state materials [8-9], classifying complex inorganic structures into their respective crystal classes or space groups from X-ray diffraction patterns [10-11], and machine-learned interatomic potential for accelerating atomic scale prediction [12]. In order to assess the reliability of such forward design models for practical applications, their predictive performances are usually compared against accuracy limits from standard laboratory approaches or first-principles simulation. Considering thermochemical target properties for example, the experimental *chemical accuracy* is generally within 1 Kcal/mol, which is equivalent to 0.0434 eV/atom in Mean Absolute Error (MAE) [121]. As a result, researchers in the field focus on developing newer and improved descriptor designs and/or modelling architectures that can compete or even surpass such accuracy threshold. This target is not always met, as reported on the case of a periodical description of materials using a sine-induced Ewald-sum matrix representation [70, 92] to develop a forward design ML algorithm comprising a Laplacian kernel and a Manhattan norm. The modelling architecture was demonstrated to predict the formation energy of diverse crystals at 0.37 eV/atom MAE [65], which is a considerably lower accuracy as compared to experimental chemistry. In contrast, a Crystal Graph Convolutional Neural Network (CGCNN) was architectured to directly learn material properties from their interatomic connections with a realized Mean Absolute Error (MAE) score on formation energy at 0.039 eV/atom [63], which exceeds the chemical experimental accuracy, thus, highly efficient for realistic predictive purposes.

Considering the inverse design approach, the objective is directed towards the discovery of unknown materials that exhibit certain functionalization predefined by optimized targets. Most often, solving the inverse design challenge will progressively involve a sequence of reconstructive

ML algorithms that are systematically assembled for materials regeneration. The new materials, as obtained from the inverse design pipeline, are normally of close variance to the originals in the training set. As such, the inverse design approach is crucial for materials discovery and its process may include a single or a combination of several ML strategies such as Deep Generative Modelling (DGM), evolutionary learning, metric learning, adaptive learning, and Bayesian optimization, among others. For instance, the discovery of new metastable Vanadium Oxide (VO) materials in their polymorphic forms was made possible by exploring the continuous latent space of a Variational Autoencoder (VAE) model [13]. In a different study, a semi-supervised VAE (SS-VAE) generative model was engineered for discovering new inorganic photovoltaic materials that are of a general chemical composition [14]. Moreover, the efficient sampling of a chemical composition latent space using a constrained Deep Convolutional Generative Adversarial Network (DC-GAN) produced new materials of the Bismuth Selenide (Bi-Se) type [15]. The successes achieved in the aforementioned studies open up opportunities for advancements emerging from newer methodologies. Besides, it is believed that the possible number of theoretical materials could be as high as a googol (i.e. 10^{100}) [16]. With only a very small fraction of this estimate discovered so far, it is hypothesized that ML can contribute to accelerating the discovery of novel materials, particularly for emerging applications in energy, healthcare, biological, defense, and agricultural industries.

As a result, this thesis primarily aims at bypassing experimental and first-principles methods by designing and implementing AI-driven ML and DGM solutions for discovering unknown energy materials. In particular, this research investigates perovskites, which is an emerging material class due to their exceptional multifunctional properties. Specifically, perovskites are among the most versatile energy materials with serviceable functionalities that can be utilized in areas related to superconductivity [17], catalysis [18], ferroelectricity [19], piezoelectricity [20], optoelectronics [21], photovoltaics [22], biosensors [23], and many more. Their multifunctionality is credited to the wealth of the design space afforded by their stoichiometry as multiple chemical elements across the periodic table can occupy distinctive ionic sites within the crystal structure. As such, this study develops novel solutions that address two areas in perovskite determination. The first area is the forward design simulation and is directed at accurately predicting standard perovskite targets, which are of high interest to material scientists

and engineers. This forward design approach follows a supervisory learning framework and involves mapping known perovskite compounds to their corresponding targets. The second area is the inverse design simulation and is directed at discovering novel perovskite materials that meet predefined stability and functionalization requirements based on a target-structured optimization/latent space. Here, the approach is based on semi-supervisory learning, whereby optimized target properties are the subjective inputs, while the chemical composition and the configuration geometry of matching perovskites are the output.

1.2 Purpose of research

The current research aims at developing Deep Generative Modelling (DGM) techniques towards the discovery of new synthesizable perovskite energy materials that possess optimized target qualities. To achieve this objective, focus is given to four critical targets that characterize the formability, stability and functionality of perovskites. They include stability energy (or energy above convex hull), formation energy, energy bandgap and geometrical lattice constraints. By definition, the stability energy indicates the thermodynamic state of a theoretical chemical compound with respect to decomposition, and is crucial for screening synthesizable candidates [24]. The formation energy quantifies the formability of a chemical structure from its disintegrated form and is necessary for developing phase diagrams [24-25]. The energy bandgap quantifies the energy region between the valence band and the conduction band, and as such, is a useful indicator for characterizing the electronic state of a material (i.e. insulating, semiconducting or conducting material) [26]. In crystallography, lattice constraints (such as lattice parameters and crystal system) provide a geometrical description of the lattice structure and can help assimilate precise symmetrical behaviours within the microstructure of the material. Hence, the core of this thesis is to efficiently design robust DGM frameworks (i.e. inverse design pipelines) with target-specific constraints for multi-objective search optimizations, as it relates to the generation of novel perovskites. To validate the chemical and geometrical feasibility of the new candidates emerging from the DGM algorithms, the Density Functional Theory (DFT) [27] is applied for performing structural relaxations on the variable-cell of the generated perovskite candidates. Perovskite

materials that are confirmed to converge upon spin-polarized DFT calculations are subjected to property determination and recommended for further analysis and/or synthesization.

1.3 Novelty of research

The novelty of the current research is documented in a total of three published peer-reviewed scientific journal articles, one further journal article submitted for publication, as well as one peer-reviewed full conference article with oral presentation. As illustrated in Fig. 1, the contributions advance knowledge in two major ways:

(1) Developed a forward design model with improved target property prediction accuracy

The present study contributes to advancing benchmark predictive accuracy on perovskite targets. Given perovskite's functionality over a wide range of potential applications [17-23], accurately estimating their properties is of great interest to researchers and industry. Standard tools for perovskite target determination often involve DFT techniques and laboratory experiments. Despite their reasonable successes in the past, their application can be cumbersome, expensive and time-consuming. Hence, this research develops efficient and inexpensive modelling architectures with improved target predictive performance compared to conventional ML approaches in the field. By correlating the structural – property relationship of perovskite materials, ML is used to infer an optimizable mapping function that is trained on a well-labelled dataset. To evaluate the target modelling accuracy on the mapping function, standardized statistical tools that compare the actual target property to the predicted target property are used. Common evaluation metrics for regressive modelling include the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the coefficient of determination (R^2 value) [30-31]. For classification analysis (i.e. distinguishing between different crystal classes), evaluation metrics include the F1-score, Area Under Curve (AUC), and the Receiver Operating Characteristics (ROC) [31]. So far, benchmark evaluations (from previous literature) have considerably contributed to pushing the predictive boundary towards first-principles or experimental accuracy levels. The current research makes further modelling improvements through newer methodologies. The focus is on crucial perovskite

targets including the stability energy, the formation energy, the bandgap, and geometrical constraints.

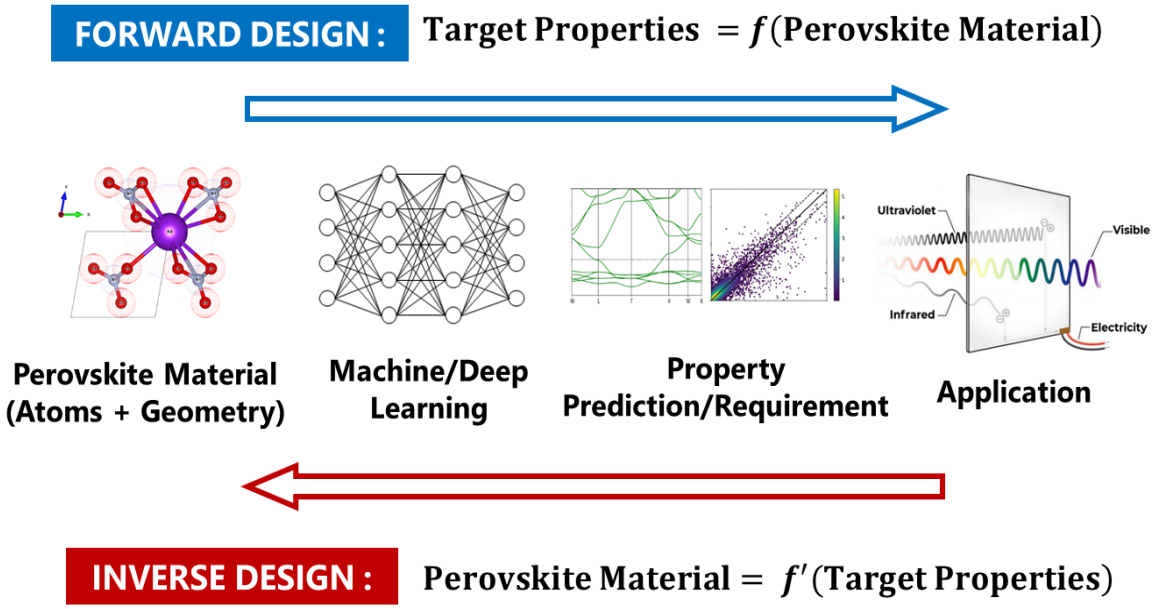

Figure 1: Schematic of modelling flow for the forward versus inverse design simulative approaches. The forward design approach begins by feature-engineering perovskite materials in a training dataset, which are correlated to their corresponding target properties. The model can then be used to predict properties of similar materials not included in the training set. The inverse design approach equally begins by feature-engineering perovskite materials in a training dataset and correlating them to their corresponding target properties using a *generative* ML/DL model. The model is then conversely used to produce new materials featuring the targeted property requirements.

(2) Developed an inverse design model for the discovery of novel perovskites based on predefined and optimized target properties

The material science community reported about 2×10^5 inorganic material compounds within the past century for various applications that improve everyday living [28-29]. Considering

only perovskites configurational permutations afforded by their stoichiometry, while mindful of the possibilities of charge imbalances and ionic-swapping forms, over 10^7 distinct structures are presumed to potentially exist. Therefore, there is currently great drive and need to explore novel perovskites that are yet to be discovered from this large galaxy of possibilities. Previous efforts for exploring novel materials were primarily based on chemical intuition, such as Edisonian-based synthetic chemistries. Other conventional methods have been derived from empirical rules or first-principles deterministic approaches, such as DFT and Molecular Dynamics (MD). Despite some reasonable successes achieved by such evolutional methods, their long-term impact for materials discovery in extremely large search spaces remains inadequate. Therefore, the current research makes a contribution to the new frontier based on innovative AI. In the process of discovering novel perovskites, the developed modelling designs solve the inverse design simulative challenge, as enabled by an active and learnable latent space of a DGM. The implemented approach yields substantial improvement as compared to existing ML-based techniques that have often been limited to straightforward tabular-dataset models and compositional phase-field representations.

1.4 Scope of research

This thesis explores novel approaches for perovskite material discovery using an interdisciplinary approach that includes materials science and engineering, solid-state physics, applied numerical methods, quantum mechanics, and computer modelling and simulation. Specifically, theories derived from solid-state physics are used for formulating the structure – property relationships in a perovskite material, which ensures that ML algorithms can adequately assimilate underlying data patterns in the process of making decisions. Moreover, most challenges encountered in engineering analysis are often non-linear. As such, this thesis engages ideas from applied numerical methods for addressing some computational challenges. For instance, crystal materials in general are known to be periodic with regular ordering across a lattice. Hence, implementing numerical techniques in the Fourier domain is beneficial for modelling the periodic effect of perovskites in the reciprocal space of their unit cell lattice. Based on the conservation of crystal momentum, the Fourier transforming operation emerges into a spatial arrangement of atomic distribution, which emerges into the expression of an approximated structure factor

equation for indexing geometrical properties such as crystal systems, space groups and lattice parameters [32]. As a result, Fourier transforming applications are demonstrated in this research for describing perovskite's crystal periodicity, as it relates to analogous feature engineering. Finally, the present study verifies all geometrical and ground-state electronical properties of novel perovskite candidates using principles derived from quantum mechanics, as implemented using the Quantum Espresso Density Functional Theory (DFT) package [33]. In essence, DFT are quantum-mechanical simulative techniques tools that are used to determine the geometrical and electronical behaviour of atomic and molecular systems, based on the approximation of electron densities. Overall, the proposed study is strongly data-driven and High Performance Computing (HPC) demanding. The source codes for simulating the DGMs are scripted using the Python programming language on a Keras TensorFlow Graphical Processing Units (GPU) backend [109], and are made openly available to the research community.

1.5 Research outline

This thesis is compiled into six broad chapters covering the introduction, literature review, methodology, forward design simulation, inverse design simulation, and conclusion, respectively. A brief overview of each chapter is provided as follows:

Chapter 1 introduces the research background and describes the scientific challenge in the field that this research seeks to address. In addition, the expected outcomes in novelty as well as the techniques required to obtain them are explained.

Chapter 2 is the summarising literature review section that expands beyond and links the specific scopes of the scientific publications. It highlights state-of-the-art traditional techniques for discovering perovskites, as they relate to first-principles techniques and synthetic experiments. The chapter describes the perovskite crystal structure, the types of stoichiometries, the multi-functionality afforded by perovskite stoichiometries, and the techniques for perovskite determination via first-principles and synthesization.

Chapter 3 discusses the overall research methodology and the role of AI for accelerating perovskite discovery. The chapter further elucidates the quasi-unlimited chemical combinatorial

space for designing perovskites, the working mechanisms of the applied AI tools, and the different types of descriptors implemented in the study for representing multi-stoichiometrical perovskite compositions. The chapter finally highlights available crystal data platforms and their reliability for carrying out perovskite data mining.

Chapter 4 presents the first two publications that resulted from the forward design simulations for improving benchmark prediction accuracy on target property determination.

Chapter 5 presents the subsequent three publications that resulted from the core research objective of this thesis, which focusses on the generative inverse design. The chapter proposes three new DGM frameworks (i.e. inverse design pipelines) for generating novel perovskites. The newly discovered materials are subjected to DFT validation and property determination.

Chapter 6 is the overall discussion. The chapter discusses the generative functionalities of the developed DGM models for perovskite discovery. The chapter finally highlights some special perovskite candidates with potential applications in photovoltaics and/or optoelectronics.

Chapter 7 is the conclusion. The chapter summarizes the findings of this thesis and suggests future research pathways in the field of novel materials discovery using sustainable AI technologies.

CHAPTER 2: SUMMARISING LITERATURE REVIEW – PEROVSKITE ENERGY MATERIALS

2.1 The perovskite crystal structure

Originally discovered in 1839 and named after the Russian mineralogist Lev Perovski [34], perovskites are widely recognized as some of the most versatile energy materials with multifunctional properties that are applied in engineering fields such as superconductivity [17], catalysis [18], ferroelectricity [19], piezoelectricity [20], optoelectronics [21], photovoltaics [22, 117-118], biosensors [23], and so on. The multifunctionality characteristic of perovskites is the main reason why they have attracted significant research interest in several interdisciplinary fields of sciences and engineering. Historically, the first perovskite chemical compound is credited to the naturally occurring calcium titanate ($CaTiO_3$) mineral in its crude form. As such, other general perovskite crystal materials are hereditary to the ancestral $CaTiO_3$, and are chemical compounds with similar stoichiometry, geometrical configuration, and ionic composition. The possibility of direct chemical substitution of constitutive ions within distinctive crystallographic sites associated with their configuration is what makes perovskites chemically diverse. As such, the interesting properties of perovskite materials are mostly associated with their stoichiometry, which influences the nature of crystallinity, geometrical formation, morphology, and functionalization. Common perovskite stoichiometrical types include ternary ABX_3 [35], double A- and B- sites [36, 42-43], hybrid organic-inorganic [22, 37], and anti- or inverse-perovskites [38].

2.1.1 Ternary perovskites of the ABX_3 stoichiometry

The ternary ABX_3 compound is the most fundamental and prevalent stoichiometry for perovskite structures. It consists of three distinctive chemical sites within a crystal lattice, which are orderly arranged from the primitive unit cell to the overall crystal structure. In general, the lattice sites are divided into two non-equivalent cationic positions in the A- and B- sites, and one anionic position in the X-site. The A- and B- site elements are in most cases metallic and can therefore alternate over a wide mix of possible cations, spanning across groups I–XV of the

periodic table, including lanthanides and actinides. For the X-anionic site, typical chemical elements may include oxides (O^{2-}), halides (i.e. F^-, Cl^-, Br^-, and I^-), hydrides (H^-), selenides (Se^{2-}), and tellurides (Te^{2-}). As illustrated in Fig. 2, the ABX_3 perovskite structure in its ideal form can be geometrically described as a perfect symmetrical $Pm\bar{3}m$ cubic structure. The B-site ion occupies the center of a corner-sharing BX_6 octahedron and is coordinated by six X-site elements, whereas the A-site ion is situated in a twelve-fold cavity in between an undistorted polyhedral of a three-dimensional setting [35]. Using the ideal $SrTiO_3$ cubic perovskite as an example, the structural formula based on equilibrate ionic stoichiometry and anionic coordination is given as

$$Sr^{2+}Ti^{4+}O_3^{2-}\left\{Sr_{[m3m]}^{XII}\ Ti_{[m3m]}^{VI}\ O_{3[4/mm]}^{(II)+IV}\right\} \quad (1)$$

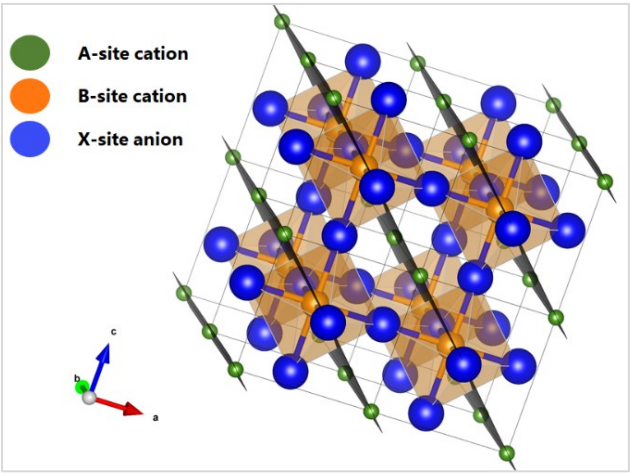

Figure 2: An ideal cubic perovskite crystal structure of the ABX_3 stoichiometry.

It should be noted that a large variety of other derivative forms of the ABX_3 perovskites exist with non-idealized (distorted) geometries, which are known to exhibit specially-tailored properties for functionalized applications. Largely due to their unsymmetrical crystal structure alignment, the deviation from cubic ideality results into different crystal systems or space-groups. In general, the main reasons for perovskite structural deviation from simple cubic to other forms are attributed to the following three reasons: (1) BX_6 octahedral tilting; (2) first-order Jahn-Teller (J-T) distortion effects; and (3) second-order J-T ionic displacement effects. The BX_6 octahedral

tilting/rotation is known as the most prominent distortional type that is found in naturally occurring perovskites. This tilting in octahedral formation is due to the differences in relative ionic sizes of the constitutive elements within the BX_6 octahedron. The first-order J-T effect describes the distortion of the BX_6 octahedral based on electronic instabilities that are associated with the compositional ions. The geometrical distortion in octahedral formation depends on the non-linearized axial and equatorial bonds that are observed among octahedral complexes. As a result, octahedral distortion occurs to stabilize the perovskite by minimizing the axial bond energy in the form of elongation or compression. The second-order J-T effect is rooted in cationic displacements within a three-dimensional perovskite polyhedral setting. This effect primarily depends on the relative oxidation states or charge differences of the A- and B- sites cations. Moreover, a combination of the three aforementioned mechanisms of deviation from the ideal structure may simultaneously occur in some perovskite structures. The magnitude and orientation in structural deviation or distortion depend on the cationic-anionic pair, local cationic ordering, and size effect of the constitutive ions. Hence, non-idealized perovskites generate polymorphic variations, which play key roles in characterizing the geometrical (stable), physical and functional features of the materials class [39-41].

2.1.2 Double perovskites of the $A_2BB'X_6$ and $AA'BB'X_6$ stoichiometries

Double perovskites are higher derivatives of the single ternary ABX_3 and are formed due to cationic displacements or local ionic-site sharing within the A- and/or B- sites of a perovskite crystal structure. The two most common double-site perovskites are the double A-site and the double B-site compounds. In a conventional double B-site perovskite, the B-site location relative to the ideal ABX_3 setting is shared by two cations to form a rock salt ordering structure with stoichiometry defined as $A_2BB'X_6$ or in some cases $A_3BB'X_9$ forms. The stoichiometry ensures that the oxidation states of the constitutive ions chemically satisfy the formula of the structure in consideration. This opens up versatile avenues for numerous choices in B-site cations that are available across the periodic table of chemical elements. Considering the $A_2BB'O_6$ (double B-site oxide-perovskite), an average A-site cation could be divalent (+2), whereas at the B-sites, each cation can possess oxidation states of +4 to conserve ionic neutrality. A typical example of the

$A_2BB'X_6$ configuration is illustrated in Fig. 3 for ferromagnetic $Sr_2^{2+}Fe^{4+}Mo^{4+}O_6^{2-}$ revealing distinctive ionic sites across the compound [36]. Although less common, the double A-site perovskite is also of high interest to material scientists and engineers due to their unique properties from the sharing of multiple cationic positions. The double A-site perovskites can be viewed as an extension to the double B-site compound, but with additional cationic sharing at the A-site of a simple ternary stoichiometry. For most double A-site cases, the stable chemical formula can be represented as $AA'BB'X_6$. A typical example is $Na^{1+}La^{3+}Mg^{2+}W^{6+}O_6^{2-}$ where Na and La share the A-sites whereas Mg and W share the B-sites [42]. The most prevalent defect that occurs in double perovskite stoichiometries is cationic anti-site disordering, which is a case when the positions of B- and B'-sites are swapped in an unfavorable pairing disorder [36, 39, 43]. This creates more complexity in the structure as the overall B-site cations are separated by antiphase boundaries. Such an anti-site disorder can influence the functional properties of the material.

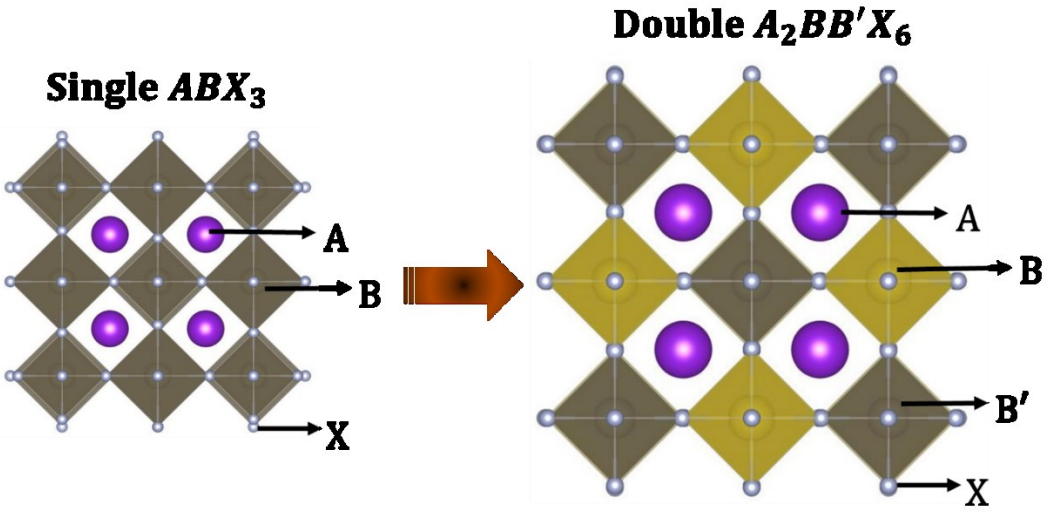

Figure 3: The double perovskite crystal structure. $A_2BB'X_6$ is a derivative of the fundamental ABX_3 compound with ionic sharing characteristics at the B-site of the chemical configuration.

2.2 Perovskite multifunctionality and applications

Perovskite materials are well known for their multi-functionality, which is primarily due to the collaborative effect associated with the diverse chemical elements occupying specific positions in the crystal structure. Considering the chemical flexibility associated with some complex forms of perovskites, uncommon behaviours related to multidimensional quantum degrees of freedom (e.g. lattice, spin, charge, orbital overlaps, etc.) can lead to several eclectic possibilities of interesting functionalities. Table 1 provides a summary of some perovskite chemical compounds in relation to their crystal system and determined functional properties that extend to conductivity (i.e. transport phenomenon), magnetism, and catalysis [44]. Moreover, perovskite-based devices have emerged as key candidates for light harvesting applications [22, 45]. The optical advantages associated with perovskites are strongly related to their higher solar absorptivity and tunable direct bandgaps. Unlike silicon-doped conventional solar cells with larger thickness, perovskite cells are considered to be more cost-effective by allowing thin-film designs with better absorptivity effects. Prime candidates for perovskite-based optical devices are pure inorganic ABX_3 and hybrid organic-inorganic perovskites (HOIP) [22, 37]. The flexibility with choosing from a pool of different chemical ions at the A-, B- and X-sites even creates substantial opportunities for tunable band structures. According to the Goldschmidt's tolerance [46], the size of the A-site cation can affect the bandgap range of a given perovskite material. By increasing the size of the A-site cation, the energy bandgap can be gradually decreased for certain HOIPs [45]. However, there can be challenges with perovskite optical properties due to low power conversion efficiencies (PCE) in some compositional makeups and the detrimental effect of oxidation in tin-iodide perovskites. For instance, some tunable HOIPs are reported at just ~5% in PCE estimations, which is substantially lower than the Shockley and Queisser (SQ) calculations that ideally predict 33% PCEs for operational band gaps at ~1.3 eV [22, 115-116]. This suggests the need for further research advancements in improving PCEs for perovskite devices. In this regard, a potential solution is in the area of multijunction perovskite solar cells, which can be achievable due to their tunable bandgap. Other application examples of perovskite devices include low-temperature superconductors, image and data storage, transducers, sensors and actuators, fuel cell membranes, photochromic materials, electrochromic materials, signal processing, lasers, capacitors, etc.

Table 1: Perovskite crystal compounds and their respective applications [44]

Compound	Crystal system	Property	Application
$BaZrO_3$	Cubic	Proton and ionic conductivity	Protonic fuel cell and hydrogen separation membrane
$(Ba,K)BiO_3$	Cubic	Superconductivity	Superconductor
$AgSbO_3$	Cubic	Photocatalytic	Visible-light sensitive photocatalyst
$BaTiO_3$	Tetragonal	Ferroelectricity	Multilayer capacitor
$Pb(Zr,Ti)O_3$	Tetragonal	Piezoelectricity	Piezoelectric transducer
$PbTiO_3$	Tetragonal	Pyroelectricity	Pyroelectric infrared detector
$La(Cr,Fe)O_3$	Orthorhombic	Mixed conductivity	Solid oxide fuel cell (SOFC) cathode
$GdFeO_3$, $LaMnO_3$	Orthorhombic	Magnetic	Magnetic memory and ferromagnetism
$YAlO_3$, $KNbO_3$	Orthorhombic	Optical	Laser
$BiFeO_3$	Rhombohedral	Multiferroic	Spintronics and memory devices
$Na_{0.5}Bi_{0.5}TiO_3$	Rhombohedral	Piezoelectricity	Lead-free piezoelectric
$LaAlO_3$	Rhombohedral	Catalytic	Industrial catalyst for oxidative coupling of methane (OCM)

2.3 Perovskite determination via first-principles (ab initio) methods

First-principles (or ab initio) methods are typically high-throughput techniques that are based on quantum mechanical simulations, which involves solving the many-body Schrödinger equation. Popular examples of some proven first-principles methods include density functional theory (DFT) [27], molecular dynamics [47], and Monte Carlo [48] simulations. For most applications, DFT is the de facto method of choice, which is due to its reasonable electronic-structure approximation and appreciable accuracy for determining ground state properties. Examples of some ground state properties that can be simulated using DFT include structural, mechanical, optoelectronic, magnetic, thermoelectric and thermodynamic properties. Specifically, coupling the Tran-Blaha-modified Becke-Johnson (TB-mBJ) functional [110] with the Generalized Gradient Approximation (GGA) pseudopotential was proven to accurately simulate the $RbGeI_3$ perovskite in agreement with laboratory experimental results [49]. In a study on novel materials discovery using purely DFT means, a dataset of 1,346 hybrid organic-inorganic

perovskites (HOIP) was prepared via a combination of atomic search method and DFT calculations [79]. The HOIP structures were optimized using GGA and the hybrid Heyd-Scuseria-Ernzerhof (HSE06) functional [80] and were further subjected to property determination for estimating the bandgap, dielectric constant, and relative energies. Moreover, several well-established crystallographic databases utilize DFT in their material's property determination. For instance, platforms such as Materials Project (MP) [56-57] and Open Quantum Materials Database (OQMD) [25, 58] both generate their hypothetical materials using the Vienna Ab initio Simulation Package (VASP) [50] on a Projector Augmented Wave (PAW) pseudopotential [78], as parameterized using the Perdew-Burke-Ernzerhof (PBE) functional [77]. Their achieved results are demonstrated to be in close agreement with experimentally verified Inorganic Crystal Structure Database (ICSD) standards [28-29]. Although perovskite property determination and discovery via DFT has been successfully implemented in many studies, several drawbacks exists that limit their applications. For instance, DFT and first-principles deterministic methods in general can be computationally expensive, which results in minimal scalability for accelerated discovery. In addition, the high complexities with respect to deciding and approximating exchange-correlations and functionals can lead to poor choices that can affect the authenticity of the simulated results [1].

2.4 Perovskite synthesization and characterization

The method in which a perovskite chemical compound is synthesized may considerably affect the defining properties of the material. As a result, there exist several manufacturing techniques for growing perovskites, which are highly influenced by the physical states of the raw materials. In general, perovskite synthesization processes can be divided into three types: (1) synthesis by solid-state reaction; (2) synthesis by wet chemical methods; and (3) synthesis by gas-phase reaction. Solid-state reactions require that the raw materials are processed in their solid physical states and may involve a sequence of manufacturing techniques that extend to powder/nanoparticle fabrication, ball-milling, compaction, sintering, calcination, and secondary thermomechanical treatments. For instance, the synthesization of stable proton conducting $BaCeO_3$ and $BaCe_{0.95}Yb_{0.05}O_{3-\delta}$ perovskites has been shown to be experimentally possible by ball-milling $BaCO_3$, CeO_2 and Yb_2O_3 solid precursors followed by calcination and grain refinement at

temperatures reaching 1600ºC [81]. Although solid-state perovskite synthesization techniques are generally straightforward, they are prone to several drawbacks such as inhomogeneity, defects, and chemical impurities, which makes them unsuitable for manufacturing sensitive coatings. Wet chemical methods, on the other hand, are solution-based methods and are preferred over solid-state reactions due to lower processing temperature, improved reactivity, single-phase homogeneity, better stoichiometrical control, refined (purified) particle size, larger surface area, and greater flexibility in forming thin films [82-84]. Some proven solution-based synthesization techniques include sol-gel preparation, co-precipitation of metallic ions using precipitating agents, hydrothermal treatment, and Pechini method, among others. Common precipitating agents used to prepare certain perovskite compositions include oxalate-based acids for producing $BaTiO_3$; hydroxide-based solutions for producing $BaZrO_3$ and $LaCoO_3$; cyanide-based solutions for preparing rare-earth (RE) orthoferrites/cobalt compounds (i.e. $RE[Fe,Co]O_3$); and acetate-based solutions for producing $La_{1-x}Sr_xCoO_3$ ($x = 0, 0.2, 0.4, 0.6$) double perovskites [82]. The drawbacks with utilizing wet chemical methods include toxicity, solvent compatibility, ionic conformity, and solubility concerns, which altogether influence the usability and quality of the final product. For producing perovskite thin-films with nano-crystallinity [73], gas-phase techniques are the most preferred manufacturing processes due to the ease in controlling operational thickness and specific composition. Examples of some famous gas-phase techniques that have been applied in perovskite synthesization include laser ablation, electron beam evaporation, thermal evaporation, molecular beam epitaxy, dc sputtering, and magnetron sputtering [82]. For example, thermal evaporative techniques have been successfully employed in producing specially refined $YBa_2Cu_3O_7$ double perovskite thin-films with single-phase microstructure by evaporating Y, Cu, and BaF_2 chemical targets under a controlled oxygenated environment [85].

Material characterization techniques are commonly used to analyze the microstructural properties of perovskite crystals, and are broadly grouped into three types: (1) structural characterization, (2) morphological characterization, and (3) thermal analysis [82]. Structural characterization techniques, such as single-crystal X-ray powder diffraction (XRD), X-ray photoelectron spectroscopy (XPS), and Fourier transform infrared spectroscopy (FTIR), are used to identify and distinguish the different phases of synthesized perovskites. XRD, in particular, plays an important role in determining the structural and crystallographic features, as it relates to

phase orientations, lattice parameters, theoretical density, and particle size. On the other hand, morphological characterization techniques are used to examine the topological behaviour of perovskites. For example, Scanning Electron Microscopes (SEM) and Transmission Electron Microscopes (TEM) are widely employed for analyzing the surface morphology of general crystals. In a previous study on SEM analysis, for instance, it was observed that the porosity in $LaBO_3$ ($B = Ni, Co, Fe$ or Mn) perovskite microstructures is largely influenced by the constitutive ionic metal occupying the B-site location [86]. Such SEM investigation was suggested to be in good agreement with Goldschmidt's tolerance factors [46] on structural ideality with respect to the rationale on why some constitutive B-site ions are more formidable than others [108]. Finally, on the aspect of thermal analysis, in-situ thermochemical stability has been calibrated for $SrPdO_3$ perovskites using Thermo-Gravimetric Analysis (TGA) [82]. The information obtained from the TGA experiment was helpful in prescribing optimum temperatures with respect to decomposition. Other thermal analysis methods that have been applied in perovskite characterization include thermos Differential Scanning Calorimetry (DSC) and Differential Thermal Analysis (DTA) for evaluating formability and stability.

CHAPTER 3: OVERALL METHODOLOGY – PEROVSKITE DESIGN SPACE AND AI APPLICATION

3.1 Chemical combinatorial design space for perovskite discovery

The possibility of several chemical elements across the periodic table that can occupy distinctive ionic sites allowed by perovskite stoichiometries is responsible for perovskite's vast design space. For instance, exclusively permuting the 94 naturally occurring chemical elements, while mindful of anti-perovskite stoichiometrical possibilities [38, 107] and charge imbalances from Jahn-Teller electronic instabilities [42], the potential numbers of ABX_3, $A_2BB'X_6$ and $AA'BB'X_6$ structures are estimated at $\mathbb{C}_3^{94} = 134{,}044$, $\mathbb{C}_4^{94} = 3{,}049{,}501$ and $\mathbb{C}_5^{94} = 54{,}891{,}018$, respectively. Besides, this rough estimation does not take into account the possibility of polymorphic variants characterized by the same chemical composition but different physical phases and, as a result, exhibit behaviours that are unrelated to their compositional duplicate peers [111]. As an example, Fig. 4 is a frequency-modelled periodic table revealing the chemical combinatorial design space for the ABX_3 configuration from the Open Quantum Materials Database (OQMD) [25, 58]. Surveying through the OQMD for proven ABX_3 perovskites provides about 30,000 chemical compounds that have been explored using DFT. Other proven and computationally reliable databases such as the Materials Project (MP) [56-57] database provides only about 4358 and 3880 investigated entries, for proven ABX_3 and $A_2BB'X_6$ perovskite stoichiometries, respectively. These known materials are only a small fraction of perovskite compounds, given the number of unknowns that are still yet to be discovered. As a result, a quasi-unlimited number of unknown perovskite materials potentially exist, which makes data-driven technologies the most promising approaches for exploring the vast design space to facilitate accelerated and inexpensive novel perovskite discovery.

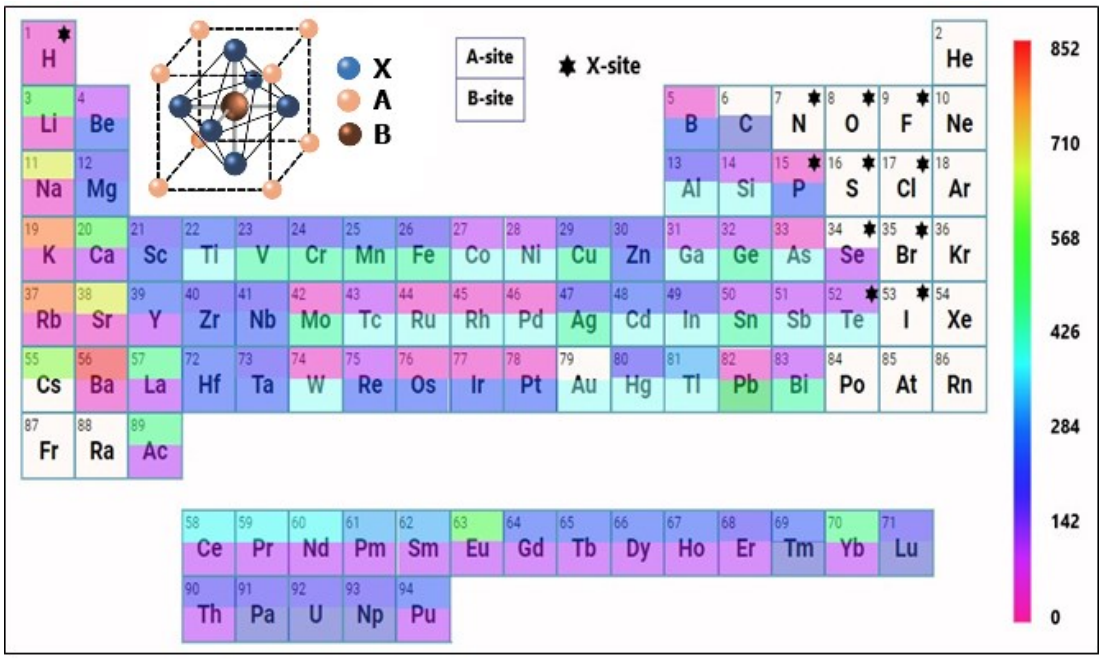

Figure 4: Frequency-modelled periodic table from the preprocessing of data. The distribution in A- and B-site cations, in addition to asterisked X-site anions, can be seen for ternary perovskites of the ABX_3 stoichiometry.

3.2 AI simulation for perovskite discovery

The thesis implements two modelling approaches for simulating perovskites using AI techniques. In the first, the forward design is applied for accurate target property prediction. The modelling approach is strictly supervisory, as the ML models learn from a well-curated dataset that contains target labels. The target properties of interest that were addressed in the forward design simulation include stability energy (E_s) otherwise referred to as energy above convex hull (E_{hull}), formation energy (E_f), and energy bandgap (E_g). The second approach is the inverse design simulation and implements Deep Generative Modelling (DGM) techniques towards the discovery of novel perovskites. The DGM learning process autonomously assimilates key similarities and differences within the data structure, which facilitates the generation of perovskite variants in a material's latent space [54]. The standard DGMs considered in the present research include variational autoencoder (VAE) [55, 74] and generative adversarial network (GAN) [72].

Both models are architectured to be semi-supervisory for actively generating new perovskite materials while simultaneously being conditioned on target properties.

3.2.1 Simulating the forward design using supervisory models

For simulating perovskite target properties, a supervised ML model is used that can be generally formulated as $f(.) : X \mapsto Y$. Here, Y is the target perovskite property of interest (i.e. E_s, E_f, or E_g); X is the crystal structure (i.e. known perovskite as deduced from an input space); and $f(.)$ is a function that models the relationship between the perovskite X and the output Y. Therefore, the goal of a supervised ML workflow is to determine $f(.)$ using a well-defined labelled perovskite dataset known as the training set, and to use the trained model to predict Y for the novel (not included in the training set) perovskites. The inferred function $f(.)$ to be minimized can be expressed using Equation 2:

$$f = \text{argmin}\, \mathbb{E}_{(x,y)\sim P}\big[c\big(x, y, f(x)\big)\big] \tag{2}$$

$\mathbb{E}_{(x,y)\sim P}[.]$ is the expectation over test data drawn from a perovskite dataset P; and $c(.)$ is a loss/cost function that determines the impact of making a prediction. As simulated in this research, the forward design challenge is preliminarily addressed by carrying out a comparative analysis of twelve tabular dataset ML models. The ML models include Ada-Boost Regression (ABR), Bayesian Ridge Regression (BRR), Decision Tree Regression (DTR), Gradient Boosting Regression (GBR), Gaussian Process Regression (GPR), K-nearest Neighbor Regression (KNR), Kernel Ridge Regression (KRR), Multi-layer Perceptron (MLP), Passive Aggressive Regression (PAR), Random Forest Regression (RFR), Stochastic Gradient Descent (SGD), and Support Vector Regression (SVR). Moreover, for improving the accuracy in prediction capability, a conjoint deep learning framework that involves a feature-extracting two-dimensional convolutional neural network (Conv2D) [51] and a prediction-enhancing support vector machine (SVM) [52, 89] is designed. Inspired by the human neuroanatomical system, the Conv2D is a deep learning model specially designed for analyzing and processing high-dimensional visual data through a sequence of algorithms that spatially perform convolutional and pooling operations. They have found wide applications in computer vision for simulating image classification, object

detection, and image segmentation [87]. In the present research, the Conv2D model learns the regressive forward problem by performing non-linear operations in convolutions on a described perovskite representation. Given the specific modelling application, the perovskites in the training set are represented using image-based input features that uniquely describe the thermochemical, geometrical, and crystallographic properties associated with the chemical compound. In general, a neuron-like Conv2D processing unit can be described using Equation 3 [53]:

$$E = \emptyset\left(\sum_{i=1}^{N} w_i X_i + b\right) \tag{3}$$

E is the pre-trained target property; w_i are the updated weights that are associated with each hidden layer i; X_i represent the input features to the unit; b is a bias; and $\emptyset$ is the activation function. For the concerned hidden neuron (i.e. feature extraction layer), the activation function is thus non-linearly effected by introducing a hyperbolic tangent (*tanh*) function, as described in Equation 4:

$$h = tanh(w_h X + b_h) \tag{4}$$

The learnable model's weights are updated by back propagation as in Equation 5:

$$w_h^{new} = w_h^{old} - \eta\left[\frac{\partial \mathcal{L}}{\partial w_h}\right] \tag{5}$$

h represents the feature-extracted layer, which is optimized by enabling a gradient descent with respect to a loss function $\mathcal{L}$ and learning rate η over several epochs. The h feature is one-dimensional (first-rank tensor) in size with vector length ($h \Rightarrow \mathbb{R}^R$: $1 \times R$) and is bounded with values between 1 and -1 due to the non-linear effect of the *tanh* function used in activating the extracted hidden layer [88]. The learned h features are fed to a SVM model specially designed for analyzing lower-dimensional representations of the original perovskite input. In general, SVM models are supervisory tabular-dataset models that are well suited for learning decision boundaries in hyperplane projections. In the present modelling approach, SVM, as enabled by the radial basis kernel, learns the function $f(h) \rightarrow Y$, which takes as input the extracted Conv2D output (h) and

predicts the unknown perovskite target (Y) in a hyperdimensional feature space. The epsilon-SVM regressive function estimates the target energy property as given in Equation 6 [52, 89]:

$$Y = \sum_{i=1}^{l} w_i \varphi(h_i) + b \tag{6}$$

$\varphi(h_i)_{i=1}^{l}$ is a mapping subset of extracted Conv2D outputs; and w_i and b are coefficients. All coefficients are based on the updated feature set $\{(h_i, y_i)\}_{i=1}^{l}$, where y_i is the targeted perovskite variable, and h_i represents the extracted Conv2D feature vector.

3.2.2 Simulating the inverse design using variational autoencoders

Inspired by the Bayes theorem, Variational Autoencoders (VAEs) are Deep Generative Models (DGMs) that are capable of capturing the latent space of a standard autoencoder using a probability distributed function. Unlike standard autoencoders that generally output the encoded input representations into a fixed latent space, VAEs output the described perovskite inputs into a smooth and continuous latent space using Gaussian distribution variables [55, 74]. Given a dataset of unlabeled perovskite input points $X = \{x_1, x_2, \dots, x_n\} \epsilon \mathbb{R}^D$, where D can be a higher dimensional tensor, the task with unsupervised VAE is to determine a function $f'(.)$ that accurately recognizes distinctive patterns within the unlabeled data points. For a known set of original perovskite samples (i.e. $\{x_i\} \subseteq X \in \mathbb{R}^R$), the encoded VAE latent vectors (i.e. $\{z_i\} \subseteq Z \in \mathbb{R}^Q$) are obtained using a probabilistic recognition (i.e. encoding) network, where $Q \ll R$ denotes the dimensionality reduction or feature extraction. The goal of VAE is therefore to approximate the true posterior $p_\theta(z|x)$ in the decoding phase by learning the distribution $q_\phi(z|x)$ at the encoding phase. The learning process is enabled by a reparameterization technique for generating new data points that are of close variance to the original x_i encoded in the latent space. Mathematically, the reparameterization can be expressed as given in Equation 7:

$$z = \mu + \sigma \odot \epsilon, \text{ where } \epsilon \sim \mathcal{N}(0, \mathrm{I}) \tag{7}$$

z is the sampled perovskite latent vector; μ and σ are deterministic vectors denoting mean and standard deviation, respectively; and ϵ is a random variable, which can be generated using a standard Gaussian (normal) distribution $\mathcal{N}$. Due to the competing nature of the encoding and decoding functions, two predominant training losses coexist and they include (1) reconstruction loss and (2) Kullback-Leibler (KL) divergence loss [90]. The reconstruction loss ensures that perovskite outputs from the decoding phase are in close similarity to their original encoded forms. On the other hand, the KL loss function measures the divergence (or appreciable distance) between two probability distributions (i.e. p_θ and q_ϕ) and is responsible for the smooth overlap of perovskite data points within the latent space. Through a sequence of back-propagation and stochastic gradient descent, the general VAE loss function $\mathcal{L}_{VAE}$, which is to be minimized, can be expressed using Equation 8:

$$\mathcal{L}_{VAE}(\phi,\theta) = KL\big[q_\phi(z|x)||p_\theta(z)\big] - \frac{1}{n}\sum_{i=1}^{n}[\log P_\theta(x|z)] \tag{8}$$

ϕ and θ are parameters corresponding to recognition and generative models, respectively. On averaging the distribution $\log P_\theta(x|z)$ over $i = 1,2,\dots,n$ entries, the reconstruction error of all perovskite feature embedding can be calculated, which is, in essence, equivalent to the Mean Squared Error (MSE).

The latent space of the traditional VAE model can be further organized on specific targets to produce a semi-supervisory DGM [91]. The present study implements such an approach by combining both the unsupervisory learning architecture from a VAE model and a supervisory learning arm from a feed-forward Neural Network (NN) (also known as Multi-layer Perceptron (MLP)) to produce a semi-supervisory variational autoencoder (SS-VAE) model. As a result, the training losses associated with Equation 8 are modified to incorporate the supervisory learning algorithm using Equation 9:

$$\mathcal{L}^* = \mathcal{L}_{VAE}(\phi,\theta) + \mathcal{L}_{MLP} \tag{9}$$

$\mathcal{L}^*$ denotes the contributing effect by the conventional VAE model and the effected MLP network for target-learning in hyperdimension (i.e. $\mathcal{L}_{MLP}$). Typical MLP losses for regression and binary classification are the Mean Squared Error (MSE) and binary cross-entropy, respectively. For multi-

classification (e.g. target-learning different crystal systems), the categorical cross-entropy is used instead of the binary cross-entropy.

3.2.3 Simulating the inverse design using generative adversarial networks

In general, Generative Adversarial Networks (GANs) autonomously learn hidden data regularities by optimizing the objective functions from two competing models, i.e. a generator model and a discriminator (or critic) model. The generator model comprehensively produces synthetic *de novo* data by assimilating the underlying variabilities within the perovskite data distribution. On the other hand, the discriminator model analyzes the data type by approximating the probability of the critiqued data to be either original or synthetic. The competing effect of generator and discriminator produces new data points that are unique but share common characteristics with the training data examples. In this research, the GAN model serves the primary purpose of generating synthetic latent space vectors (i.e. $\vec{z}^* \Rightarrow \mathbb{R}^{256}$) that are in similitude to their originally compressed perovskite representative forms ($\vec{z}$). As a result, given a random noise variable ($\vec{n} \Rightarrow \mathbb{R}^{100}$) that is provided as input to the generator, the value function $V(G, D)$ of a typical GAN model can be described using Equation 10 [72]:

$$\min_{G} \max_{D} V(G, D) = \mathbb{E}_{\vec{z} \sim p_{data}(\vec{z})}[\log D(\vec{z})] + \mathbb{E}_{\vec{n} \sim p_{\vec{n}}(\vec{n})}\left[\log\left(1 - D\left(G(\vec{n})\right)\right)\right] \quad (10)$$

G and D are differentiable modelling functions represented by generator and discriminator, respectively. $\mathbb{E}$ is the expected likelihood for maximizing and minimizing D and G, respectively. $p_{\vec{n}}(\vec{n})$ is the defined prior distribution on the input noise variable. Furthermore, similar to the VAE model, the conventional GAN can be optimized on target-specific properties to produce the semi-supervisory auxiliary (A-GAN) model. This thesis applies such an approach by multitasking the discriminator to query the authenticity of synthetic perovskite representations, while at the same time predicting their corresponding geometrical properties. As a result, the generator model of the A-GAN implicitly learns the geometrical constraints by producing novel perovskite candidates that are target-optimized on the predefined lattice constraints.

3.3 Spherical Linear Interpolation (SLERP) for sampling operation

In traditional DGMs such as the VAE and GAN, the sampling strategy used in exploring or exploiting the latent space is critical, given that the novel discovery of new materials will emerge from there. The latent space itself can be visualized as a hyperdimensional-encrypted form of the original data, whereby forensic investigations can be smartly conducted [96]. Several sampling strategies have been proposed in past studies such as random sampling, local perturbation, global perturbation, and spherical linear interpolation [14]. This thesis utilizes the Spherical Linear Interpolation (SLERP) [75] due to its preferred functionality for explorative perovskite search. Based on the theory of spherical quaternions, SLERP is used to create new data points by carrying out complex vector interpolations in conformity to the hyperdimensional geometry of the latent space. In the current research, the SLERP technique is specifically applied on inverse design models that are primarily developed using the SS-VAE model in order to generate statistical variants of their originally encoded version. Given the sampling of perovskite data points $\{z_i\}_{i=1}^{n} \subseteq Z \in \mathbb{R}^{Q}$ that are presently within the region of interest in the latent space, SLERP can be formulated using Equation 11:

$$\vec{Z}_{ij}(z_i, z_j; t) = z_i \frac{\sin(1-t)\theta}{\sin\theta} + z_j \frac{\sin t\theta}{\sin\theta} \tag{11}$$

$\vec{Z}_{ij} \Rightarrow (\mathbb{R}^{Q}: 1 \times Q)$ is the interpolated vector between two distinctive perovskite vector points z_i and z_j along finite length $t \in [0,1]$ (i.e. line space) in hyperdimensional Q space. θ is the angle between the two vectors. As a result, $\vec{Z}_{ij}$ contains intrinsic and beneficial properties hereditary to z_i and z_j perovskite data points. Figure 5 illustrates the displacement of encoded perovskite points within the latent space of a trained SS-VAE model prior to SLERP application. Due to the effect of the supervisory target-learning arm, the latent space can be seen to be target-optimized with respect to the organization/clustering of yellow and blue points, corresponding to interested and uninterested perovskite points, respectively. Moreover, Fig. 5(A) displays Principal Component Analysis (PCA) for capturing the nonlinear data structure of the latent space using the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm [97]. The t-SNE compresses the hyperdimensional $\{z_i\}_{i=1}^{n}$ vectors into two dimensions that produce the largest variance from the

data transformation process. For carrying out SLERP however, the projected t-SNE representations are not used due to the irreversible loss of information that accompanies the data transformation PCA process. As a result, the real latent space vectors are carefully scrutinized to identify the top two axes that best capture the displacement of interested versus uninterested data points. As such, Fig. 5(B) illustrates the real latent space representation and reveals the most preferred axes i.e. z^a and z^b where $[a, b]$ are subsets of hyperdimension-Q. Hence, the points that are within the encircled region are extracted and explored using the SLERP technique. Therefore, interpolating among all extracted points produces $(t_{max}/\Delta - 1) \times \mathbb{C}_2^Z$ unique and novel perovskite data points, where Δ is the number of finite spacings along $t \in [0,1]$.

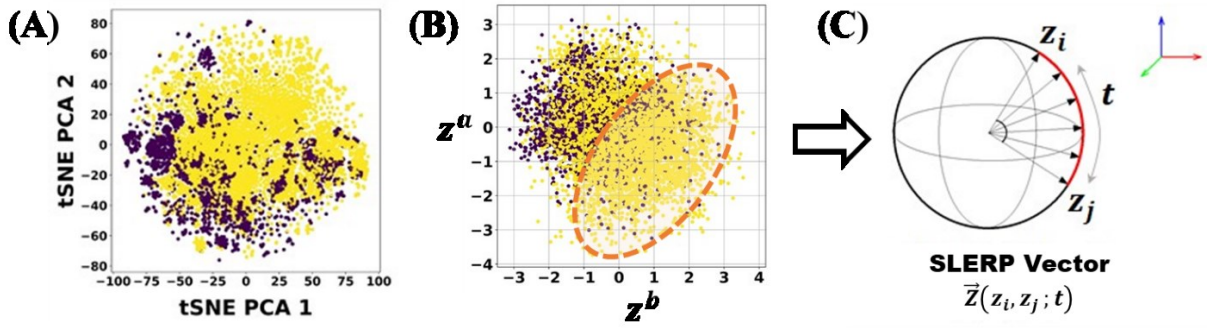

Figure 5: Interested region in the latent space for conducting SLERP sampling operation. (A) t-SNE PCA projection of the real latent space revealing the non-linear data structure; (B) Preferred 2D plane in the real latent space that best captures the displacement of interested versus uninterested data points; (C) SLERP visualization in hyperdimension.

3.4 Genetic algorithm for multi-objective target-search optimization

The Genetic Algorithm (GA) is inspired by the Charles Darwin theory on evolutionary biological selection (i.e. survival of the fittest) and is used to generate metaheuristic solutions for complex analytical problems. Their computational flexibility and non-derivative problem-solving capability makes them appealing for solving single and multi-objective optimization problems [76]. The core concept with GA involves the initialization of a batch population of potential

solutions, which are allowed to mate (i.e. crossover) and mutate in order to reproduce offspring that are considerably better than their parents, as monitored by a fitness function. In the present research, GA is used as a search algorithm for finding the most promising perovskite candidates emerging from the SLERP sampling operation (i.e. $H = g(\vec{Z}_{ij})$). The fitness function $g(.)$ is multi-objective and takes into account crucial stability and synthesizability target requirements in order to rank the most feasible candidates. In the present research, stability and synthesizability are measured using four important target properties. The first property is the formation energy (E_f), which quantifies the formability of a chemical structure from its disintegrated form and is necessary for developing phase diagrams [24-25]. Perovskites with negative E_f are generally considered to be stable. The second property is the energy above convex hull (E_{hull}) or stability energy (E_s), which indicates the thermodynamic state of a theoretical chemical compound with respect to decomposition [24]. As demonstrated in a study on sulfides and oxides, compounds with $E_{hull} \leq 0.08$ eV/atom were confirmed to be highly stable upon synthesization [104]. The third property is the *ICSD* label, which is selected due to its good correlation with the E_{hull} parameter. In general, materials that are decorated with descriptive labels from the Inorganic Crystal Structure Database (ICSD) are known to have been physically synthesized in experimental laboratories [28-29]. The final property is the dissimilarity factor ($\mathcal{F}$), which takes into consideration the geometry of a generated material with respect to proven standards. Perovskite candidates with lower $\mathcal{F}$ values are therefore characterized to be geometrically closer to proven standards, and hence, less likely to be prone to the detrimental concern of overlapping atomic coordination at the decoding phase of a generative model. As a result, the fitness function of the GA model discretely assimilates information derived from the aforementioned targets to highly rank new perovskite solutions that are favorably predicted to satisfy the predefined conditions. Therefore, all SLERP $\vec{Z}_{ij}$ vectors are comprehensively analyzed in batch populations using the GA model towards the reproduction of new vector solutions via a sequence of operations that involve crossover and mutation. Besides, the mutation process is designed to be quality-adaptive, by flipping the genes of low-quality solutions twice as much as high-quality solutions. Hence, given a set of $\vec{Z}_{ij}$ vectors, each scalar is represented as a gene, and a combination of genes belonging to a particular vector produces a parent chromosome, as illustrated in Fig. 6. Based on mathematical

schema theory, the probability (P_{GA}) of abolishing admirable genetic pattern in produced offspring by the crossover of parents can be expressed mathematically using Equation 12 [105, 120]:

$$P_{GA} = \frac{\vec{Z}_{ij} + 2\sqrt{G}}{3\vec{Z}_{ij}} \tag{12}$$

G is the number of generations/iterations allowed by the GA model. Thus, a lower P_{GA} value indicates that the excellent genetic traits of mating SLERP vectors are maintained or conserved among new generations. The GA model, as implemented in the present research, is based on the *PyGAD* module [106], which is an open-source Python library for performing evolutionary learning and optimizing ML algorithms.

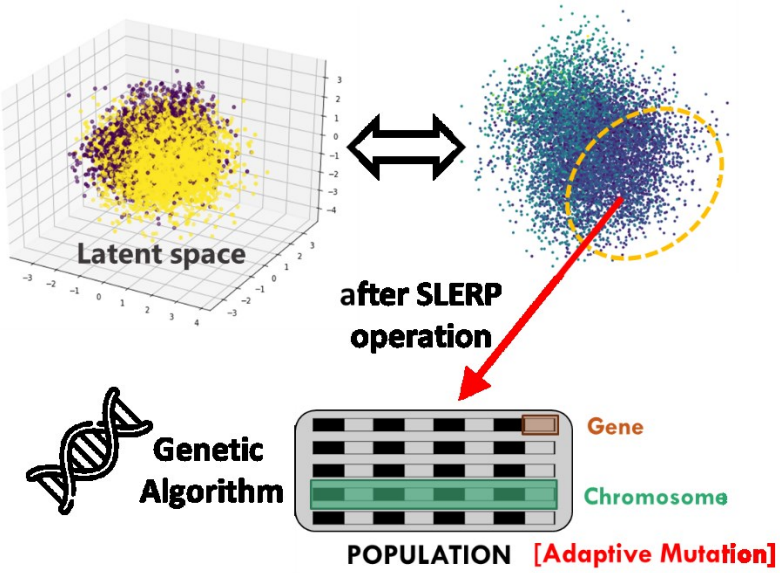

Figure 6: Integrated Genetic Algorithm (GA) for ranking high-quality perovskite candidates.

3.5 Feature engineering for multi-stoichiometrical perovskite representation

Unlike molecular and organic materials that have standard representative forms for feature engineering (e.g. Simplified Molecular Linear Input Specification (SMILES) representations [93] and graph-based methods [94]), crystalline materials lack absolute descriptor designs, which is a

consequence of material-inductive biases. In the context of such biases, modelling descriptor designs for crystalline materials will have to take into account several considerations that includes the crystal material class, the physicochemical state of the material, the stoichiometry, the periodic effect of the reciprocal lattice, and the developed ML framework. In materials informatics, moreover, the feature-engineering concept is expected to align to the mechanism of simulation, as it relates to the forward and inverse design. For instance, forward design descriptors are relatively straightforward, as they are not required to be recoverable or invertible after passing through a ML model. On the contrary, inverse design descriptors must be chemically and geometrically recoverable after assimilation by a model in order for a user to gain meaningful insight on the generated material and interested target property. As a result, the current thesis progressively designs four user-interpretable representations for feature-engineering/describing multi-stoichiometrical perovskites: (1) generalized compositional design; (2) Fourier-transformed design with feature extraction; (3) invertible image-based design; and (4) invertible mesh-grid design. The generalized compositional design is primarily used for simulating the forward design, as it lacks recoverability capabilities, which are necessary for invertible designs. The Fourier-transformed representation with feature extraction is designed to adequately model the periodicity of a perovskite crystal lattice, and is highly recommended for target property determination. For specifically simulating the inverse design, the invertible image-based and invertible mesh-grid design concepts are implemented in the deep generative modelling of novel perovskites. A full description of each descriptor concept, in addition to their applications, are made available in the integral journal and conference papers.

3.6 Data sources and preprocessing measures

The reliability of AI-based technologies often depends on the validity, clarity and acceptability of data to enable accurate modelling. Table 2 outlines some well-established crystal databases and their available sizes for materials informatics. The data platforms accommodate mostly inorganic compounds in their binary, ternary, quaternary, and hybrid forms. Databases such as the Materials Project (MP) [56-57], the Open Quantum Materials Database (OQMD) [25, 58], the Automatic Flow (AFLOW) [59] and Novel Materials Discovery (NOMAD) [60] are purely

open-sourced with little to no downloading restriction. In the present study, the MP and OQMD platforms were explored for extracting generic compounds that adopt the ABX_3, $A_2BB'X_6$ and $AA'BB'X_6$ perovskite stoichiometries. The datasets consist of a sizable portion of experimentally verified compounds as standardized by the inorganic crystal structure database (ICSD) [28-29], as well as hypothetical entries. For both MP and OQMD databases, the hypothetical and experimental materials are determined by performing DFT calculations using the Vienna Ab initio Simulation Package (VASP) [50] on a Projector Augmented Wave (PAW) pseudopotential, as parameterized by the PBE-GGA functional. Despite similar DFT implementations between MP and OQMD databases, key variances in outputs remain apparent, which is due to the different parametric choices effected in their respective calculations. A critical ML analysis, as carried out on the reproducibility and interoperability of crystal databases, establishes the need for thorough standardization, as it relates to DFT parameterization [95]. Due to these differences in standard calculations, the extracted and preprocessed perovskite samples from MP and OQMD are analyzed independently, and not collectively (i.e. samples are not mixed from both databases). In this thesis, data from MP were primarily used for modelling the inverse design due to their higher DFT-computed reliability, compared to OQMD. The better fidelity associated with MP's data is due to their rigorous DFT parametric choices and data warehousing approach for accurately reproducing real experimental target properties from the ICSD. For simulating the forward design, however, the OQMD dataset was used as the primary training dataset, given their much higher data availability compared to MP.

Table 2: Inorganic materials databases and their available data sizes (*K* and *M* equivalent to thousand and million, respectively).

Ref.	Materials database	Data size
[56, 57]	Materials Project (MP)	144 K
[25, 58]	Open Quantum Materials Database (OQMD)	816 K
[59]	Automatic Flow (AFLOW)	3.5 M
[60]	Novel Materials Discovery (NOMAD)	11 M
[61]	Materials Cloud (MCloud)	22 M
[62]	Crystallography Open Database (COD)	490 K

Prior to ML investigation, the extracted data from MP and OQMD are preprocessed/screened to ensure they satisfy certain modelling requirements. In general, the first screening measure involves the removal of data samples that have incomplete and/or incorrect entries. This includes samples with missing information on the interested targets, i.e. formation energy, bandgap, and stability energy. The next screening process removes certain entries that are considered *outliers*. Such samples are observed to be alienated from the targeted data distribution, and as such, possess the detrimental effect of obscuring the modelling accuracy during a simulation exercise. The threshold used to screen out outlier samples was set at formation energy and/or stability energy with values more than 5 eV/atom. Besides, such entries are highly unlikely to synthesize due to their critically unstable state. Overall, the data cleaning exercise from the OQMD extraction process resulted in 27,587 ABX_3 compounds, while data extraction and cleaning from MP resulted in 8,228 inorganic compounds comprising ABX_3, $A_2BB'X_6$ and $A_2BB'X_6$ perovskites for experimentation. For ML simulation, the cleaned data in training, testing and validation subsets are verified to contain both hypothetical and experimental entries in order to promote better generalization for effectively learning the correlations between perovskite structures and their corresponding targets.

3.7 Density Functional Theory (DFT) validation and Bayesian optimization

To validate the geometrical and chemical formability of novel perovskites emerging from the inversely designed AI pipelines, a first-principles technique based on the Density Functional Theory (DFT) is used. In principle, DFT techniques are quantum mechanical modelling methods that aim at solving the Schrödinger equation (i.e. $\widehat{H}\Psi_i(r) = E\Psi_i(r)$) using an electron density approach for obtaining ground-state electronic properties of many-body systems. In general, DFT techniques serve as the first major step towards physical or experimental synthesization. Without requiring higher-order parameters, DFT takes advantage of basic atomic quantities such as mass, charge, and Coulombic electron interaction for predicting material behaviours [1, 68]. As a function of energy density $n(r)$, the total DFT-calculated electronic energy of a perovskite chemical system can be expressed as in Equation 13:

$$E_{tot}[n(r)] = E_T[n(r)] + E_V[n(r)] + E_J[n(r)] + E_X[n(r)] + E_C[n(r)] \quad (13)$$

E_T is the total kinetic energy of electrons; E_V is the total potential energy of electrons due to Coulombic attraction to the nuclei centers; E_J is the total potential energy of electrons due to Coulombic repulsion between electron pairs; E_X is the total quantum mechanical exchange energy of electrons corrected for the strongly correlated motion of electrons of the same spin; and E_C is the total correlation energy of the electrons and accounts for the weakly correlated motion of electrons of the opposite spin.

This thesis applies standard practice by using the Quantum Espresso (QE) software package [33] for simulating DFT calculations. For all newly generated/discovered perovskite candidates, QE performs plane-wave spin-polarized DFT calculations on the variable-cell of a perovskite geometry (i.e. *vc-relax*) using the Perdew-Burke-Ernzerhof (PBE) functional of the Generalized Gradient Approximation (GGA) pseudopotential [77-78]. The choice of PBE-GGA is based on a large number of previous studies that have suggested their better computational efficiency for simulating bulk crystal materials over their Local Density Approximations (LDA) counterparts [1, 13, 14, 27, 58, 68]. Furthermore, by enabling spin-polarized calculations on all axes and angles of the variable-cell, the magnetic behaviour of new materials can be categorically determined (i.e. di-, para-, ferro- magnetism, etc.). The present research equally applies DFT for confirming crucial stability and electrical properties of some highly regarded materials, as it relates to relative energy (E_{rel}), energy bandgap (E_g), band-structure, and projected density of states (DOS). For assessing E_{rel}, a similar approach as previously demonstrated for hybrid organic-inorganic perovskites is applied in the current research [24, 79]. In essence, E_{rel} accounts for the marginal change in total DFT-computed energies (E_{tot}) by considering the energy differences between a bulk relaxed perovskite form and the sum of the isolated constitutive ions in the unit cell, at the same level of theory as the bulk material calculation. Mathematically, the E_{rel} can be expressed using Equation 14:

$$E_{rel} = E_{tot} - \sum_{i=1}^{x} E_{tot}(x) \tag{14}$$

E_{tot} is the total DFT-computed energy of the bulk system that is normalized by the constitutive number of atoms present in the system. E_{tot} is a function of x and represents the total DFT-computed energy for $i = 1,2, \dots, x$ isolated atoms that make up the system. Hence, E_{rel} is analogous to actual formation energy calculations and represents the required energy to form a perovskite material from its constitutive components x. As such, perovskites with negative E_{rel} are considered theoretically stable.

To ascertain the energy occupation at the fermi level and the electrical conductivity behaviour, band-structure and DOS analysis are implemented. In this thesis, band-structures are demonstrated along high-symmetry line segments in the irreducible Brillouin zone of their primitive crystal structures [98, 99]. Considering the sampling of the three-dimensional Brillouin zone of the reciprocal crystal lattice, fine *k*-points grid meshes (0.2 $Å^{-1}$) are used, as recommended by Materials Cloud [61]. In calculating DOS however, denser *k*-points are used (i.e. 0.1 $Å^{-1}$) for improved resolution. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) iterative algorithm [112] is applied for ionic and cell optimizations. Self-consistent field (SCF) electronic convergence is achieved by setting energy accuracy, force and pressure to 1.0×10^{-7} Rydberg, 1.0×10^{-3} Rydberg/Bohr, and 0.5 kbar, respectively. The energy cut-off threshold for charge density is set no less than nine times the corresponding value for wave function from the chemical element pseudopotential's condition [100]. To ensure that a smooth integration of electron occupation occurs across the fermi energy level, Gaussian-smearing technique with low broadening (0.01 Rydberg) is used.

This thesis demonstrates the applicability of the Bayesian Optimization (BO) algorithm for performing preliminary DFT relaxations. In principle, BOs are powerful active learning techniques that can be used to find the global solution of an unknown objective function via the approximation of surrogate Gaussian Process models [101]. They have been widely utilized in the field of computational materials science for solving several optimization problems [113-114]. In the present research, the BO algorithm is used to cheaply determine the pre-relaxed state of newly generated perovskite candidates that best minimizes the model-predicted E_{tot}. Given a generated perovskite unit cell $\hat{X}$, as characterized by their constitutive atoms ($\vec{R}_{(x,y,z)}$) and lattice geometry/parameters ($L_P \Rightarrow [a, b, c; \alpha, \beta, \gamma] \in \mathbb{R}^{(2\times3)}$), the goal of the BO algorithm is to

minimize the unknown/costly objective function (i.e. E_{tot}) while ensuring that predefined lattice constraints on the crystal system type are still maintained. Mathematically, the optimized lattice parameters ($L_{P|opt}$) can be expressed using Equation 15:

$$L_{P|opt} = \underset{L_P}{\operatorname{argmin}}\, E_{tot}(L_P) \tag{15}$$

Hence, in principle, the BO algorithm analogously performs pre-DFT relaxation, and as such, theoretically assists in conserving computational time for actual DFT simulations. Note that in the present research, the BO is demonstrated in the *Lattice-Constrained Materials Generative Model* (LCMGM) for perovskite discovery (**Section 5.2.3**). Perovskite materials that successfully undergo both preliminary BO and final DFT relaxation are therefore recommended for future in-depth analysis and potential synthesization.

CHAPTER 4: SCIENTIFIC PUBLICATIONS ON THE FORWARD DESIGN SIMULATION

4.1 Improving the prior art for higher target property prediction accuracy

The forward design approach can be formulated as follows: Given a perovskite material in training set X, find its target properties $y = f(X)$, where $f(.)$ is a function as determined by an AI algorithm, which learns the perovskite crystal structure and maps it to the predefined target. As such, the encoded perovskite is trained using a well-labelled dataset to output the desired properties that are consistent with the perovskite material. The desired target properties are normally DFT outputted or experimentally determined, such as the formation energy (E_f) and bandgap (E_g). As such, Table 3 outlines some recent benchmark evaluations on the prediction of E_f and E_g, based on different ML techniques and/or descriptors. For instance, the Crystal Graph Convolutional Neural Network (CGCNN) was developed to directly learn material properties from their interatomic connections in a crystal using a broad multigraph. Trained on about 28,000 general inorganic structures from the Materials Project (MP), the best Mean Absolute Error (MAE) scores realised by CGCNN on E_f and E_g were reported at 0.039 eV/atom and 0.388 eV/atom, respectively [63]. Furthermore, a progressive learning method was constructed to predict similar targets. Consisting of about 758 ABO_3 perovskite compounds from the MP and Open Quantum Materials Database (OQMD), the materials were described using 66 generalized input features, which included structural parameters from the octahedron BO_6 bond valence vector sum (BVVS). The best results obtained on the test set for E_f were reported at 0.087 eV/atom MAE, 0.126 eV/atom Root Mean Squared Error (RMSE), and 96.4% R^2 on Gradient Boosting Regression (GBR). For predicting E_g, the formation energy was included as an instrumental variable among the set of input features with the following results: 0.384 eV MAE, 0.521 eV RMSE, and 85.5% R^2 likewise on GBR [9]. Another investigation on target property prediction of crystal materials considered standardized periodic descriptors in the Coulomb matrix [69], Ewald-sum matrix [70] and Sine matrix [65, 92]. Using a model architecture that comprised a Laplacian kernel with Manhattan norm, their developed approach was demonstrated to predict E_f at 0.37 eV/atom MAE [65].

Table 3: Benchmark evaluation from past studies as related to formation energy and bandgap predictions for general inorganic crystalline structures.

Ref	Prediction technique	Accuracy evaluation	No. of training data
Formation Energy, E_f (eV/atom)			
[63]	Crystal Graph Convolutional Neural Networks (CGCNN)	0.039 MAE	28,046
[25]	Generalized Gradient Approximation (GGA+U) OQMD	0.081 - 0.136 MAE DFT	292,070
[9]	Bagging (Bond-valence sum)	0.087 MAE	606
[64]	Decision Forest	0.088 MAE	228,676
[65]	Kernel Ridge Regression (KRR) using periodic Sine-matrix descriptor	0.370 MAE	3,000
Energy Bandgap, E_g (eV)			
[66]	Random Forest	0.149 MAE	432
[67]	Materials Graph Network (MEGNet)	0.280 MAE	10,000
[9]	Gradient Boosting Regression (GBR)	0.384 MAE	606
[63]	Crystal Graph Convolutional Neural Networks (CGCNN)	0.388 MAE	16,458
[27]	Generalized Gradient Approximation (GGA+U)	0.6 MAE DFT	80,000

Despite the appreciable contribution by researchers in the field, there are some underlying concerns in their descriptor designs and/or AI/ML implementation, which are worth mentioning. For instance, CGCNN is demonstrated to work well on the prediction of E_f at 0.039 eV/atom MAE, but it fails to accurately predict E_g at 0.388 eV MAE. This may in part be due to two contributing factors. The first is related to the computational inconsistency with bandgap undervaluation as estimated by DFT [68]. The second may be due to the modelling limitation by their computational mechanism, which is inadequate for accurately determining E_g. Customarily, E_g can be derived from the electronic dispersion relation used in graphing the distinctive eigenstates of electrons within the conduction and valence band structures [32]. Because the dispersion relation equation (e.g. from the solution of a tight-binding model) depends on wavevector symmetrical variables in the Brillouin zones of the reciprocal lattice, a sensible

solution for improved E_g modelling may require similar or analogous input variables that are used in constructing the band structure. Previously, benchmark evaluations on E_g seem not to effectively consider such analogous features from the reciprocal space of a crystal lattice in their periodic modelling design. This highlights the importance of including Fourier-transforming representations in simulating the forward design for efficiently capturing the periodicity of a crystal material, as it relates to the reciprocal space of a crystal lattice. Hence, the present study advances benchmark evaluation from prior arts by developing a Fourier-transformed descriptor design with feature extraction for accurately modelling the energy bandgap, as applicable to ternary ABX_3 perovskites. The developed modelling design is inspired by the Fourier Transformed Crystal Property (FTCP) representation [14], but is modified to incorporate additional input features and a hybridized deep learning framework. The descriptor is feature engineered to capture the periodicity of a crystal based on an invertible image-based representation, which is demonstrated to appreciably improve the modelling accuracy, especially on E_g.

4.2 Investigations on target property prediction for perovskite structures

4.2.1 Journal publication 1 – Published in *Materials Today Communications*, 27, (2021) 102462. https://doi.org/10.1016/j.mtcomm.2021.102462:

Comparative Analysis of Machine Learning Approaches on the Prediction of the Electronic Properties of Perovskites: A Case Study of ABX_3 and $A_2BB'X_6$

Ericsson Tetteh Chenebuah, Michel Nganbe and Alain Beaudelaire Tchagang

The paper is visually summarised by the graphical abstract provided in Fig. 7. It presents a focused study on finite and infinite bandgap, and investigates a diverse set of perovskites, including oxides and halides occupying the X-anionic sites. Twelve ML techniques are described, implemented and compared against each other on the prediction of the formation energy and the energy bandgap of two distinctive crystal configurations: ABX_3 and $A_2BB'X_6$. The ML models

include Ada-Boost Regression (ABR), Bayesian Ridge Regression (BRR), Decision Tree Regression (DTR), Gradient Boosting Regression (GBR), Gaussian Process Regression (GPR), K-nearest Neighbor Regression (KNR), Kernel Ridge Regression (KRR), neural network Multi-layer Perceptron (MLP), Passive Aggressive Regression (PAR), Random Forest Regression (RFR), Stochastic Gradient Descent (SGD), and Support Vector Regression (SVR). The samples are extracted from the Materials Project (MP) database and are initially described using well developed and identical features. In addition, the effect of the Energy Above Hull feature (otherwise referred to as the stability energy) is systematically investigated among the set of initial features. As a result, the predictive performance of the formation energy is greatly improved. The Support Vector Regression (SVR) model is found to best predict the formation energy with error metrics at 0.055 eV/atom MAE, 0.096 eV/atom RMSE and 99% R^2 on the test set. Higher marginal errors are observed in the prediction of the energy bandgap, with SVR accuracy measurements evaluated at 0.462 eV MAE, 0.662 eV RMSE, and 85.18% R^2 on the test set. Despite the impressive effort by the SVR for best predicting targets, the study of the sample size effect shows that the Gradient Boosting Regression (GBR) and Random Forest Regression (RFR) models are better suited for energy bandgap prediction. Finally, feature importance is used to inspect the relative importance among all input features considered in the study. It was found that a strong relationship exists between the standard-deviated electronegativity and the formation energy. The study lays the foundation for ML application with respect to perovskite determination in the current thesis. All data and source codes implemented in this study are openly available at github.com/chenebuah/perovskite-ML.

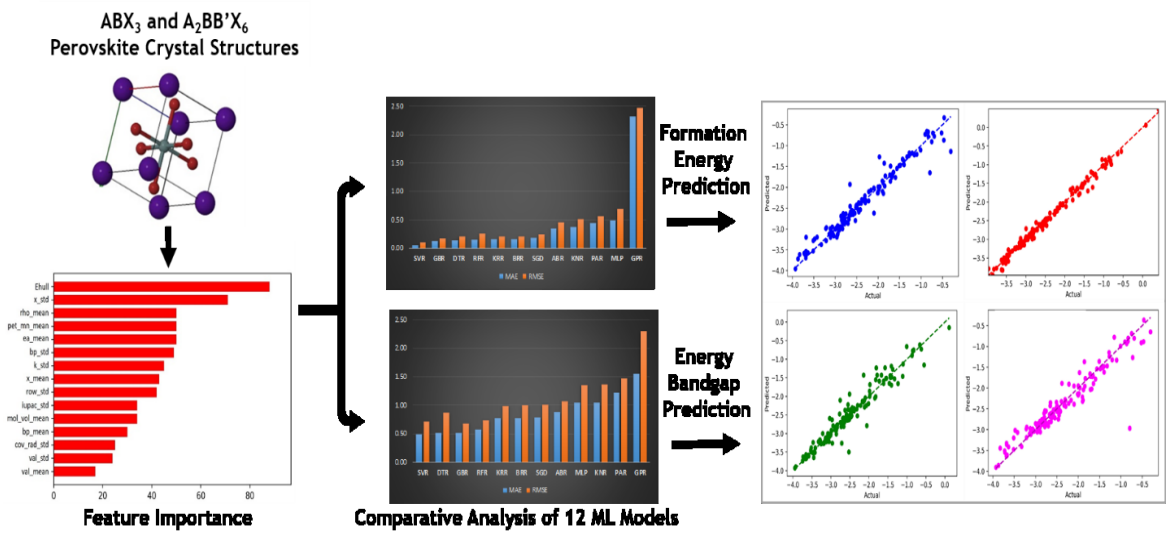

Figure 7: Graphical abstract on the comparative analysis of machine learning approaches on the prediction of the electronic properties of perovskites.

4.2.2 Journal publication 2 – Published in *Materials Research Express*, 10, (2023), 026301. https://doi.org/10.1088/2053-1591/acb683:

A Fourier-transformed feature engineering design for predicting ternary perovskite properties by coupling a two-dimensional convolutional neural network with a support vector machine (Conv2D-SVM)

Ericsson Tetteh Chenebuah, Michel Nganbe and Alain Beaudelaire Tchagang

The paper is graphically summarized in Fig. 8. The study proposes a new feature engineering approach that takes advantage of both the direct ionic features and the periodic Fourier-transformed reciprocal features of a three-dimensional perovskite polyhedral. The study is conducted on about 27,000 ABX_3 perovskite structures from the Open Quantum Materials Database (OQMD) with the stability energy, the formation energy, and the energy bandgap as

targets. For accurate modelling, a feature-extracting two-dimensional convolutional neural network (Conv2D) is coupled with a prediction-enhancing Support Vector Machine (SVM) to form a hybridized Conv2D-SVM architecture. A comparison with previous benchmark evaluations reveals appreciable improvements in modelling accuracy for all target properties, particularly for the energy bandgap, for which the feature extraction approach yields 0.105 eV MAE, 0.301 eV RMSE, and 93.48% R^2. Besides, the proposed design is further demonstrated to out-perform other similar periodic feature engineering approaches in the Coulomb matrix, Ewald-sum matrix, and Sine matrix, all in their absolute eigenvalue forms. The predictive accuracy with respect to the aforementioned standard periodic representations is improved by about 70%, 75% and 66% on the stability energy, formation energy and bandgap targets, respectively. All preprocessed data, source codes, and relevant sample calculations are made openly available at github.com/chenebuah/high_dim_descriptor.

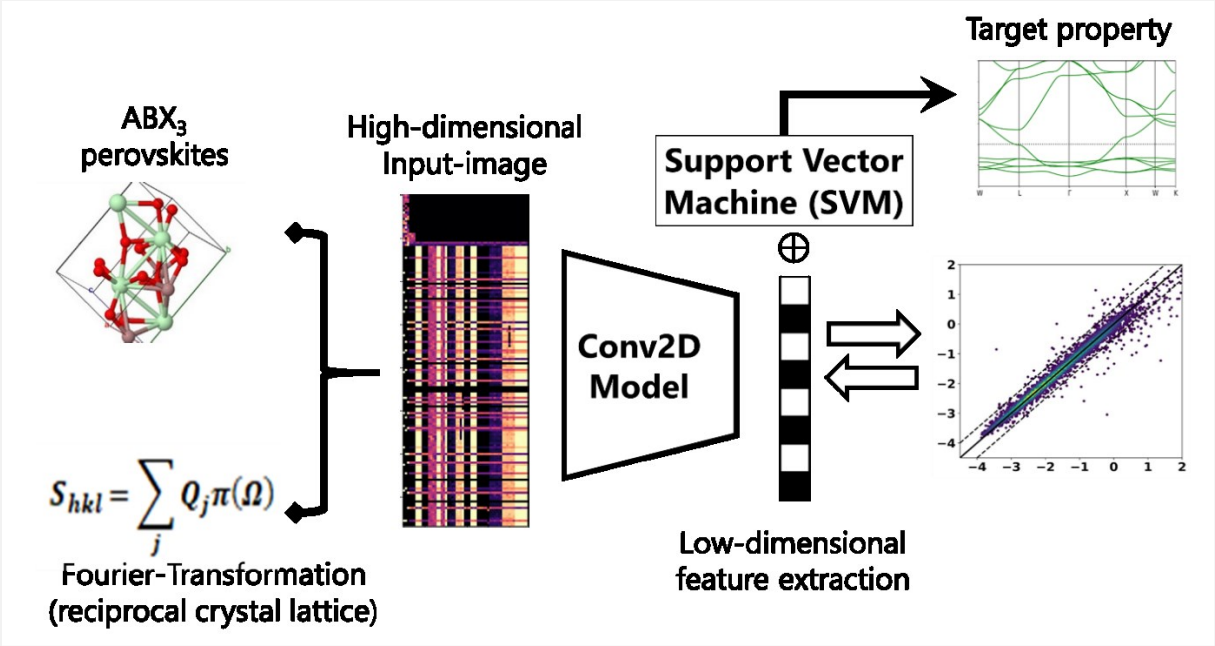

Figure 8: Graphical abstract on the Fourier-transformed feature engineering design for predicting ternary perovskite properties using the Conv2D-SVM setup.

CHAPTER 5: SCIENTIFIC PUBLICATIONS ON THE INVERSE DESIGN SIMULATION

5.1 Improving the prior art on generative inverse modelling for materials discovery

The inverse design approach can be formulated as follows: Given the desired properties y, find the crystal structure $\hat{X} = f'(y)$, where $f'(.)$ is the inverse function or sequence of reconstructive AI algorithms that regenerate the crystal structure (i.e. chemical composition and lattice geometry) from the provided target properties. Previous literature reports have contributed towards this novel approach of inversely designing crystals and molecules. The uniqueness of each design depends on several material inductive biases such as the chemical material of interest, the invertible descriptor design, the subjective target property in the latent space, and the inverse design AI pipeline. For instance, in discovering new metastable vanadium-oxide (V-O) materials, a variational autoencoder (iMatGen model) was trained on an initial dataset of 112 known V-O compounds using invertible mesh-grid descriptors that characterize atomic positions and lattice parameters. The investigation was able to produce over 20,000 polymorphic variants, all within metastability as determined by their formation energies, i.e. $E_f \leq 0.5$ eV/atom [13]. In a different study, a Fourier Transformed Crystal Properties (FTCP) image-based representation was proposed for discovering general inorganic crystals [14]. Using an initial dataset of over 34,000 samples from the Materials Project (MP), FTCP was able to successfully discover about 142 target crystals, which were all embedded within a target-structured latent space of a semi-supervisory VAE model. Moreover, other interesting designs have equally been adapted on different data structures and target-specific optimizations, such as representing crystals using the radial distribution function (RDF) for sampling low-energy materials [71], and using a constrained deep convolutional Generative Adversarial Network (GAN) [72] to produce new materials of the Bismuth Selenide (Bi-Se) type [15].

The limitations of previous literature are often related to the strategy used for latent space optimization, in addition to the geometrical and chemical quality of the generated material candidates. For instance, the iMatGen design is proven to optimize the latent space based only on

an integrated supervisory neural network for predicting the formation energy in hyperdimension. The iMatGen, however, does not theoretically address the synthesizability quality of generated candidates. In addition, the FTCP concept is demonstrated to generate highly stable materials but at the expense of geometrical quality due to significant reconstruction errors that occur from their generative design, which translates in the form of crystal asymmetrisation (i.e. low symmetrical properties) and overlapping atomic coordination. Such poor quality in the generated geometrical attributes produces anisotropic materials, which limits their usability in many engineering fields and applications. Moreover, there are few studies on deep generative modelling (DGM) designs that specifically provide discoverable pathways on the perovskite material class. Previous efforts for simulating perovskite discovery via AI have often been limited to straightforward tabular ML models and compositional phase-field representations [102-103].

As a result, the current thesis aims at contributing to the perovskite materials spectrum by innovatively developing efficient inverse design solutions via DGMs for discovering stable and synthesizable perovskites. By leveraging semi-supervisory forms of learning, the latent space is demonstrated to be organized based on predefined targets that define their stability and functionalization. The developed DGM architectures in this study are demonstrated to generate new perovskites with high chemical formability and geometrical stability. Some deterministic targets that are of particular interest in the modelling of the inverse design include the stability criterion in the formation energy, synthesizability index in the energy above convex hull parameter and ICSD labeling, and lattice symmetrisation properties in the primitive Bravais crystal systems and geometry.

5.2 Investigations on inverse design pipelines for novel perovskite discovery

This thesis develops three inverse design models for addressing specific challenges in novel perovskite discovery. The first model is called the *Target-Learning Variational Autoencoder* (TL-VAE) and is optimized for efficiently generating ABX_3 stable perovskite materials, as it relates to predefined target constraints on the formation energy. The second model is designed to enhance the TL-VAE by incorporating a genetically optimized algorithm, which assists in the high ranking of potentially stable and synthesizable perovskite candidates. This model is referred to as

the *Evolutionary Variational Autoencoder for Perovskite Discovery* (EVAPD) and is comprised of a semi-supervisory variational autoencoder (SS-VAE), an evolutionary-based genetic algorithm, and a one-to-one similarity analytical model. Unlike the TL-VAE that is strongly suited for generating ABX_3 compounds, the EVAPD is designed to be multi-stoichiometrical. The third model addresses the geometrical lapses from the EVAPD and is called the *Lattice-Constrained Materials Generative Model* (LCMGM). The developed LCMGM designs novel perovskites with multi-stoichiometries by ensuring that important geometrical constraints remain unchanged, as predefined at the encoding phase. The model is comprised of three distinctive modelling phases that consist of a semi-supervisory variational autoencoder (SS-VAE), an auxiliary generative adversarial network (A-GAN), and a Bayesian-induced model for geometrical optimization. This thesis makes openly available the preprocessed dataset used for deep machine learning, relevant source codes for developing the TL-VAE, EVAPD, and LCMGM models on GitHub (github.com/chenebuah).

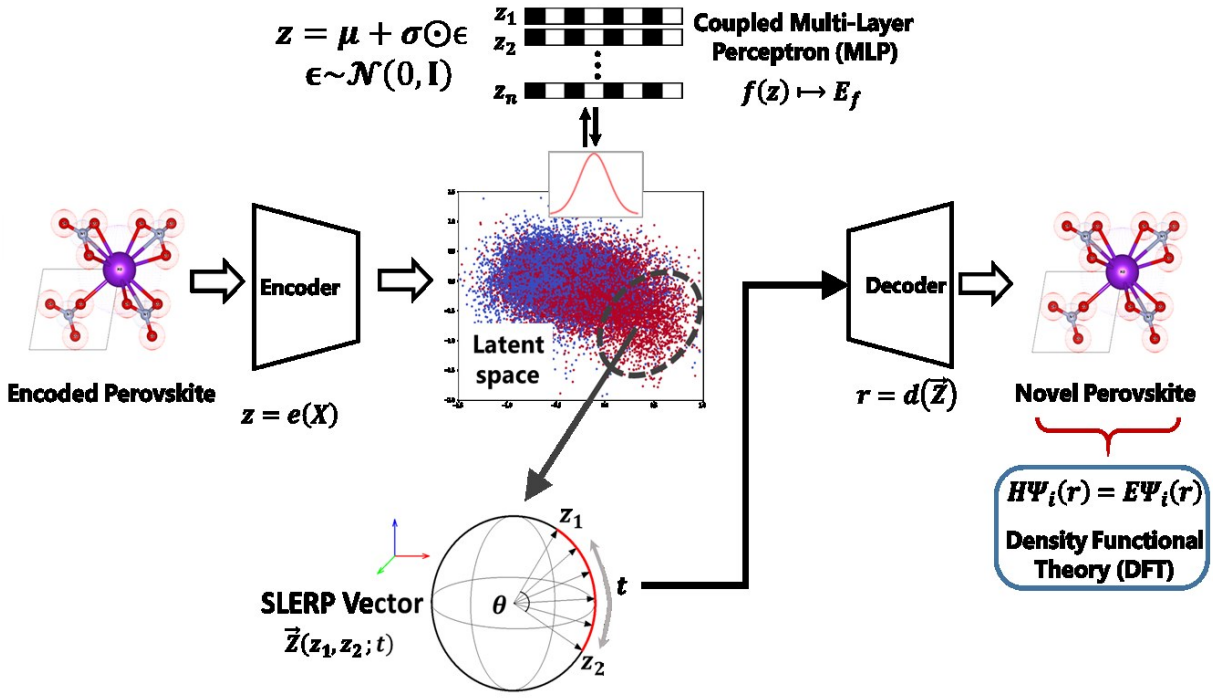

Figure 9: Modelling architecture of the Target-Learning Variational Autoencoder (TL-VAE) model.

5.2.1 Conference publication – Published in the *Proceedings of the 36th Canadian Conference on AI*, (2023). https://doi.org/10.21428/594757db.07402193:

Target-learning the latent space of a variational autoencoder model for the inverse design of stable perovskites.

Ericsson Tetteh Chenebuah, Michel Nganbe and Alain Beaudelaire Tchagang

The modelling architecture of the developed *Target-Learning Variational Autoencoder* (TL-VAE) model is visually illustrated in Fig. 9. The TL-VAE solves the inverse design challenge by primarily utilizing a semi-supervised VAE (SS-VAE) model towards the generation of novel perovskites. The modelling architecture is trained on about 27,000 ABX_3 perovskite chemical compounds from the Open Quantum Materials Database (OQMD) that are well represented using a Fourier-transforming image-based descriptor design. Given a known set of real perovskite samples in the training set (i.e. $\{x_i\} \subseteq X$), the TL-VAE model encodes all perovskite inputs into a latent space (i.e. $\{z_i\}_{i=1}^{n} \subseteq Z$) using a two-dimensional convolutional neural network (Conv2D). The latent space is optimized on distinguishable stability features, as it relates to the formation energy (E_f). The optimization is enabled by combining the unsupervised learning model from the traditional VAE with two target-learning Multi-layer Perceptron (MLP) models (i.e. feed-forward neural networks). The two interrelated MLP models are used to analyze the targeted-data distribution (i.e. formation energy), based on a dual regression and binary classification analysis (i.e. $f(z) \mapsto E_f$). For regressive analysis, the MLP's function $f(.)$ maps the known perovskite inputs to their continuous target variables, while in the case of classification, the function simply distinguishes inputs based on a predefined formation energy threshold. For sampling the latent space towards the generation of new perovskite data points, the Spherical Linear Interpolation (SLERP) technique is applied due to its extensive functionality for carrying out complex vector interpolations in conformity to the geometry of the latent space. The SLERP sampled points are decoded and compared with several proven standards in order to confirm their relative geometrical closeness. As a result, decoded perovskites with configurations that are farther away from proven standards are screened out and not considered for further Density Functional Theory (DFT)

analysis. As a proof of concept, the TL-VAE is demonstrated to generate four new materials that successfully underwent DFT validation. The new candidates include $AlPtS_3$, $AlPtO_3$, $GaOPd_3$, and $GaPdO_3$, and they were all predicted within metastability.

5.2.2 Journal publication 3 – Published in *Frontiers of Materials*, 10, (2023), 1233961. https://doi.org/10.3389/fmats.2023.1233961:

An Evolutionary Variational Autoencoder for Perovskite Discovery

Ericsson Tetteh Chenebuah, Michel Nganbe and Alain Beaudelaire Tchagang

The modelling architecture of the developed *Evolutionary Variational Autoencoder for Perovskite Discovery* (EVAPD) model is visually illustrated in Fig. 10. This research advances machine-learning capability by being one of the first contributions to demonstrate the applicability of deep evolutionary learning in the field of novel materials discovery. The EVAPD is used to screen stable and functional inorganic materials that adopt the complex $A_2BB'X_6$ and $AA'BB'X_6$ double perovskite stoichiometries. The EVAPD framework first begins by transforming multi-stoichiometrical perovskite samples from the Materials Project (MP) database into image-based representative forms (i.e. $\{x_i\} \subseteq X \in \mathbb{R}^R$). The transformed perovskite samples are trained using a SS-VAE model that dimensionally reduces all image-based inputs into hyperdimensional vectors in the latent space. The encoded latent space is pre-optimized on thermodynamic stability by conditioning a target-learning MLP model to regressively predict the formation energy. The target-learning operation assists in organizing the latent space by distinguishing stable versus unstable regions. For sampling new points within the interested stability region, a combined approach that involves the Spherical Linear Interpolation (SLERP) technique and an evolutionary learning technique, i.e. Genetic Algorithm (GA), are used. Through a sequence of biologically motivated crossover and adaptive mutation, the GA ranks the best solutions and recommends new perovskite candidates that are confirmed to satisfy the predefined

synthesizability conditions. The study presents 137 new perovskite materials comprising 114 $A_2BB'X_6$ and 23 $AA'BB'X_6$ generated compounds by the developed EVAPD model. Among them, 82 have not yet been reported in any known database model (i.e. novel and unique), and 17 are identified as potential candidates for photovoltaic and optoelectronic applications due to their DFT-determined energy bandgaps close to the ideal 1.3 eV value for high Power Conversion Efficiency (PCE) [115-116].

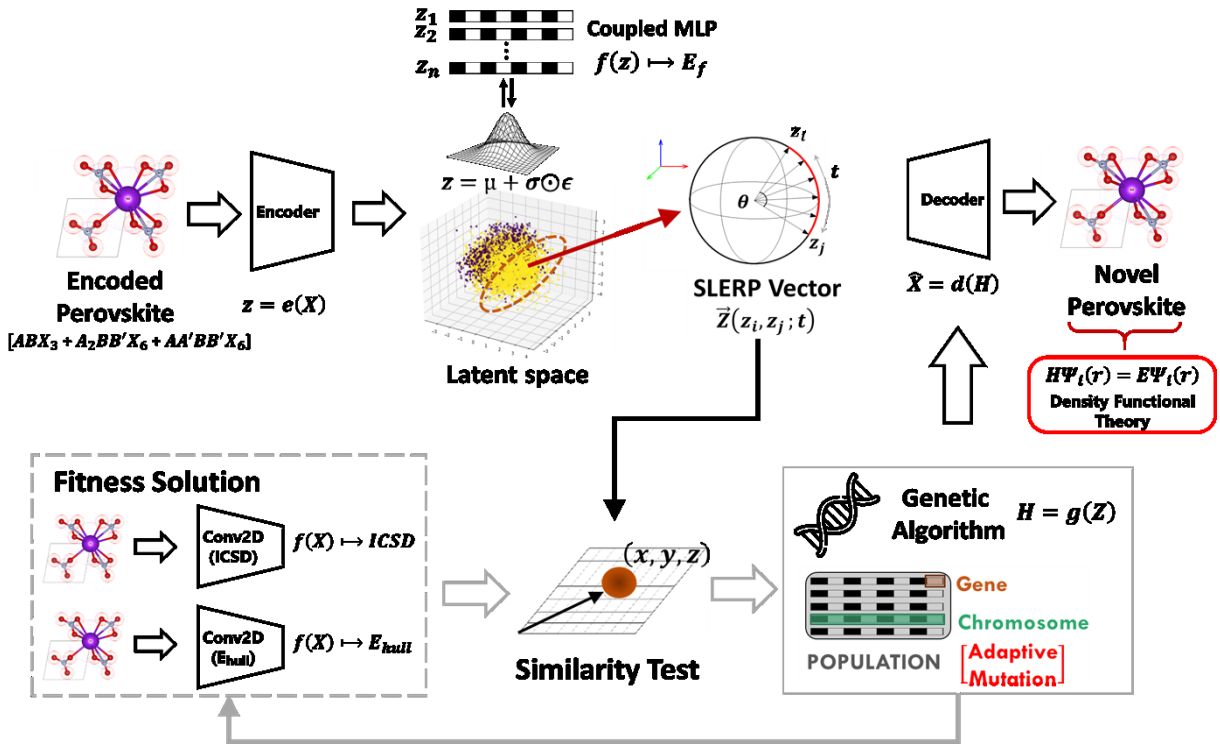

Figure 10: Modelling architecture of the Evolutionary Variational Autoencoder for Perovskite Discovery (EVAPD) model

5.2.3 Journal publication 4 – Article *under review for publication* (Mendeley data Repository at https://doi.org/10.17632/m262xxpgn2.1):

A Deep Generative Modelling Architecture for Designing Lattice-Constrained Perovskite Materials.

Ericsson Tetteh Chenebuah, Michel Nganbe and Alain Beaudelaire Tchagang

The modelling architecture of the developed *Lattice-Constrained Materials Generative Model* (LCMGM) is visually illustrated in Fig. 11. This research addresses a major challenge with deep generative modelling techniques, as it relates to lattice deconstruction at the decoding phase, which translates in the form of asymetrisation and geometrical inconsistencies of newly generated materials. The LCMGM is used to design perovskite materials with crystal lattice attributes that are consistent with predefined geometrical constraints at the encoding phase. The modelling architecture comprises three distinctive: (1) semi-supervised variational autoencoder (SS-VAE); (2) auxiliary generative adversarial network (A-GAN); and (3) geometrical optimization in Bayesian optimization (BO) and DFT. Using the SS-VAE model, mesh-grid represented perovskites of the ABX_3 and $A_2BB'X_6$ stoichiometries are projected into a target-learnable latent space. The latent space is preliminarily organized based on two target conditions, as it relates to the primitive Bravais crystal system and formation energy. For designing novel perovskites with specific geometrical attributes, perovskite data points of the intentional crystal system type are extracted and further analyzed in the next phase of modelling that involves the A-GAN model. The A-GAN comprehensively explores the chemical compositional design space of extracted perovskite data points by assimilating their respective geometrical constraints in the form of three-dimensional lattice parameters and atomic coordination. The synthetic vectors that are generated by the A-GAN are decoded and processed in a pre-validation Bayesian Optimization (BO) algorithm. The BO model analogously performs pre-DFT relaxation by finding the lowest DFT-predicted total energy, which best captures the optimized lattice configuration of the newly generated candidates. Emerging from the developed LCMGM inverse design pipeline are 124

newly designed perovskite materials with crystal lattice constraints that conform to cubic, monoclinic, orthorhombic, tetragonal and trigonal/rhombohedral systems.

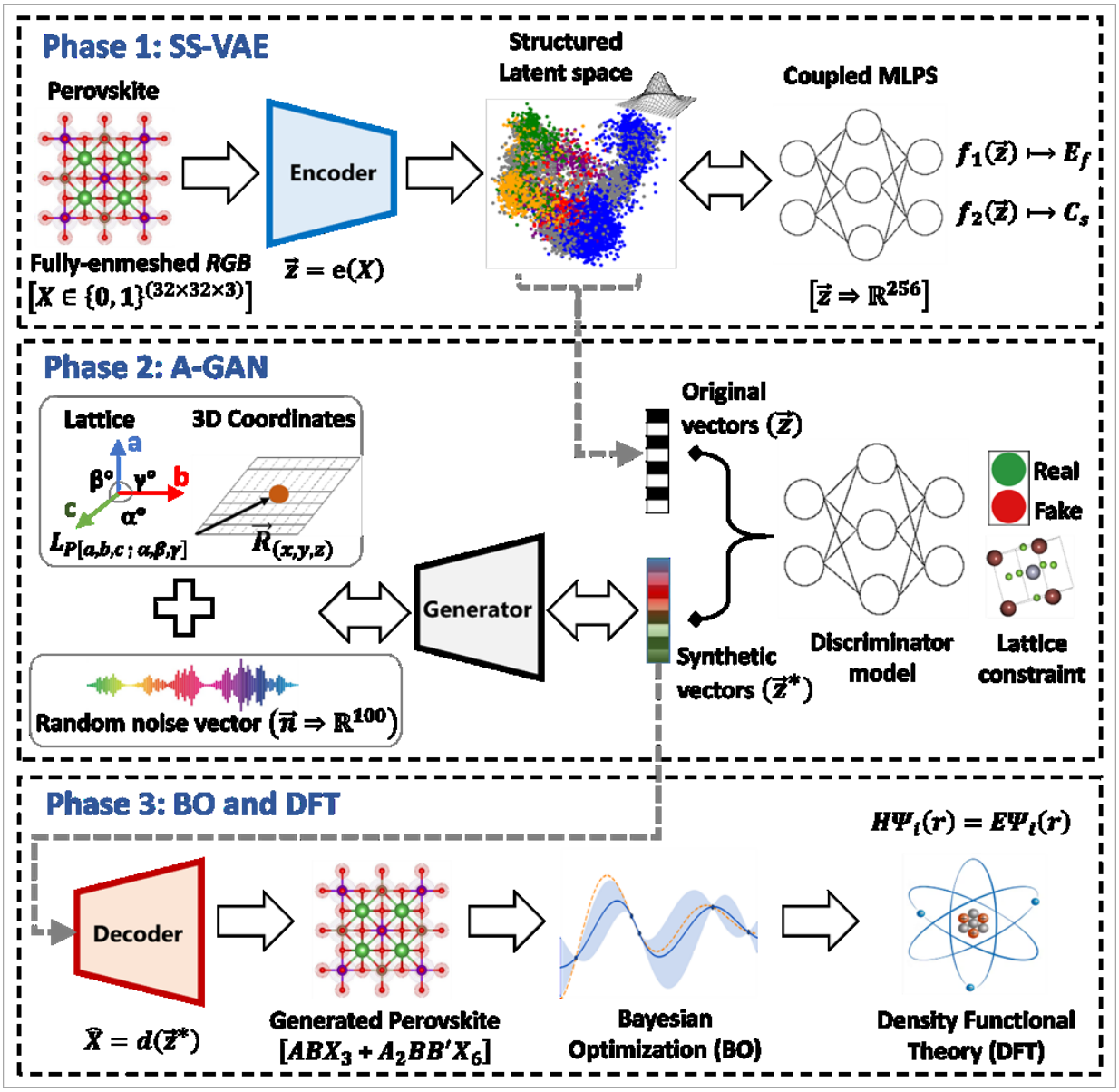

Figure 11: Modelling architecture of the Lattice-Constrained Materials Generative Model (LCMGM)

CHAPTER 6: OVERALL DISCUSSION

6.1 Newly discovered perovskites and property determination

Altogether, this thesis presents 265 new perovskite materials (69 ABX_3, 173 $A_2BB'X_6$ and 23 $AA'BB'X_6$ compounds) from the developed inverse design pipelines. All new materials were validated by DFT and were confirmed to converge upon spin-polarized relaxation calculations on all angles and axes of the variable-cell geometry, using predominantly PBE-GGA functionals. Specifically, 4 materials were generated by the TL-VAE, 137 by the EVAPD, and 124 by the LCMGM. Table 4 compares the three developed inverse design models, and outlines their respective functionalization for target-specific perovskite discovery. A full list of the materials, in addition to their determined properties, can be found in the appendix section of this thesis. Moreover, this research makes available the Crystallographic Information Files (CIF) of all generated materials on GitHub (github.com/chenebuah). The materials datasets are equally archived in the NOMAD and Mendeley data repositories, and are made openly available to interested researchers.

Table 4: Comparison of developed DGMs for perovskite discovery

DGM	Generative Algorithm	Optimization Technique	Perovskite Stoichiometry	Generated Materials	Specific Functionalization
TL-VAE	SS-VAE	Target learning	ABX_3	4	Good stability with deeper assimilation for underlying chemical laws
EVAPD	SS-VAE	Target learning, genetic algorithm and geometrical similarity learning	$A_2BB'X_6$ and $AA'BB'X_6$	137	High stability and synthesizability
LCMGM	SS-VAE and A-GAN	Target learning and Bayesian geometrical optimization	ABX_3 and $A_2BB'X_6$	124	Lattice-constrained with high symmetry and stability

6.1.1 New perovskites generated by the TL-VAE model

On examining the four materials generated by the TL-VAE model, Fig. 12 reveals ball-and-stick chemical images for $AlPtS_3$, $AlPtO_3$, $GaOPd_3$, and $GaPdO_3$ perovskite compounds. Upon crosschecking with the preprocessed OQMD dataset used in training the TL-VAE, the current research confirms $AlPtS_3$ and $GaOPd_3$ to be unique, whereas $AlPtO_3$ and $GaPdO_3$ are polymorphic duplicate variants. This demonstrates the dual functionality of the TL-VAE model for discovering both unique and polymorphic perovskite compounds, depending on the application or design consideration of interest. In addition, the TL-VAE model is able to autonomously learn uncommon forms of interesting ABX_3 compositions such as the antiperovskite $GaOPd_3$, even though the modeling framework does not directly include any premeditated screening on that specific material class. This is a result of the chemically diverse training dataset consisting of over 20,000 samples, which assists the TL-VAE model in learning underlying chemical laws. In general, anti- or inverse-perovskites are uncommon perovskite forms that are characterized by their ionic swapping nature, and are of profound interest to the materials science community due to their eccentric properties for diverse engineering applications, especially in the field of superionic conductivity [38, 107]. Moreover, it can be observed that the four new perovskite candidates are seen to coincide with A-site Aluminum (Al) and Gallium (Ga), and B-site Platinum (Pt) and Palladium (Pd). This is a consequence of the targeted sampling region for carrying out SLERP, as most perovskite structures that are encompassed within this region are predominately of a similar chemical composition. As a result, expanding to accommodate newer chemistries will necessitate intelligent search mechanisms for critically exploring the targeted sampling region within the latent space. Furthermore, by observing a common trend in the valence and periodic table properties with the new candidates, the TL-VAE model displays its ability to unsupervisorily learn chemical laws. For instance, Al and Ga elements have the same group numbers (i.e. group number 13) and the same number of valence electrons in their outermost shells (+3 ionic charge). Likewise, Pt and Pd share the same group number on the periodic table of elements (i.e. group number 10). In retrospect, both the valence electrons and group numbers were used in feature engineering the discretized thermochemistry section of the invertible image-based descriptor design (i.e. Fig. 7). Therefore, the generated results, as characterized by their similar chemical attributes, suggest that the TL-

VAE model intuitively designs new materials with similar chemical traits and does not just perform random chemical or ionic substitutions.

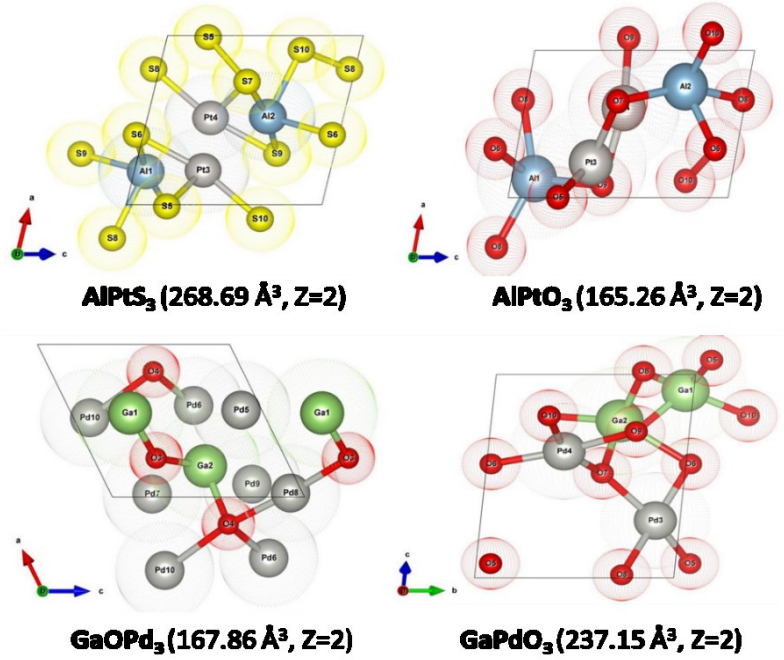

Figure 12: New perovskite candidates from the TL-VAE inverse design pipeline. In brackets are the unit cell volume (in cubic-Angstroms, $Å^3$) and the number of formula units, Z.

6.1.2 New perovskites generated by the EVAPD model

By incorporating the genetic algorithm, the EVAPD is able to produce high-quality candidates that are predicted to be stable and synthesizable. In the process of screening new materials from the EVAPD, high priority is given to chemical compositions that are unique and not in the training dataset. Screening for unique compositions and not polymorphs is motivated by the general search among researchers in the field to understand newer chemistry behaviours due to the quasi-unlimited number of unknown compositions from the perovskites' large design space. On inspecting the 137 new perovskite materials (114 $A_2BB'X_6$ and 23 $AA'BB'X_6$) from the EVAPD pipeline, 82 are confirmed to be unique and novel, as they cannot be found in any known database including MP, OQMD, and NOMAD. In addition, all new materials are predicted to be stable with negative formation energies and 73% are predicted to meet initially defined

synthesizability requirements (i.e. $E_{hull} \leq 0.08$ eV/atom), as demonstrated in a previous study on oxides and sulfides [104]. The most stable candidates with respect to $A_2BB'X_6$ and $AA'BB'X_6$ stoiciometries are identified to be La_2MgUO_6 and $SrLaTaWO_6$ compounds with predicted formation energy (E_f) at -3.3863 eV/atom and -3.3072 eV/atom, respectively. Finally, to illustrate the effect of DFT for finding the best optimized/relaxed geometry, Fig. 13 visualizes the difference in atoms and bonds coordination between a newly generated Ca_2YOsO_6 (*CIF ID: 001*) by the EVAPD, and the final relaxed state after DFT relaxation. Unlike the crude form that was generated by the model, the relaxed geometry is seen to be optimized with better geometrical coordination, as all constitutive atoms are better positioned to occupy vacant spaces within the unit cell lattice.

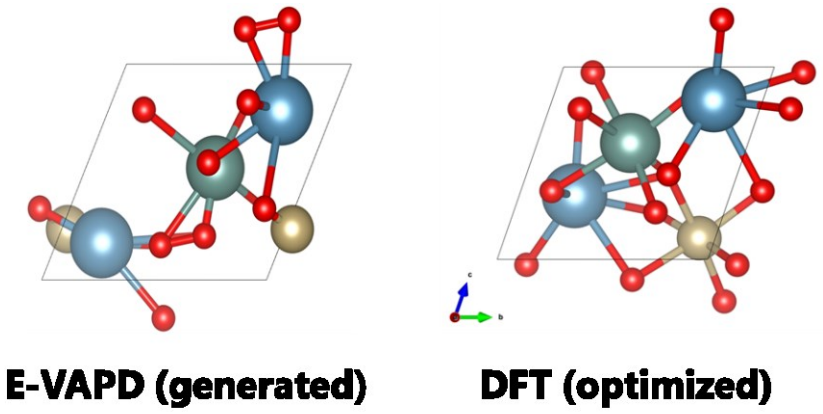

Figure 13: Comparison of geometrical coordination between a newly generated Ca_2YOsO_6 perovskite candidate by the EVAPD model and subsequent DFT optimization/relaxation.

6.1.3 New perovskites generated by the LCMGM model

The LCMGM is designed to generate lattice-constrained perovskite materials with multi-stoichiometry, as characterized by their crystal systems. The ability of the LCMGM to generate such materials with crystal lattice conformities is due to the strict imposition of geometrical constraints at the encoding phase. In the current implemented approach, the LCMGM is demonstrated to design materials that conform to either one of the following primitive Bravais

systems: cubic, monoclinic, orthorhombic, tetragonal, and trigonal crystal systems. On examining the 124 newly designed perovskite candidates by the LCMGM, 65 are of the ABX_3 stoichiometry while 59 are of the $A_2BB'X_6$ stoichiometry. On accounting for crystal type, 23 are cubic, 18 are monoclinic, 24 are orthorhombic, 25 are tetragonal and 34 are trigonal structures. The crystal systems of the new materials were found to be consistent with the predefined constraints at the encoding phase. Moreover, on searching well-established databases such as MP, OQMD and NOMAD, 72 of the new materials are confirmed to be unique and novel (i.e. new compositions), while the other 52 are polymorphic variants that have been discovered in past studies. On further investigating their properties, all new materials are predicted to be stable with negative formation energies and about 81% are metallic with infinite energy bandgaps (i.e. $E_g = 0$ eV). Figure 14 displays the most stable compounds with respect to the considered crystal systems and stoichiometries. It should be noted that the LCMGM specifically focusses on screening perovskites that are neutrally charged. In general, electroneutral compounds are more likely to be stable due to their minimized Coulombic interaction of constitutive ions at the atomic level [108].

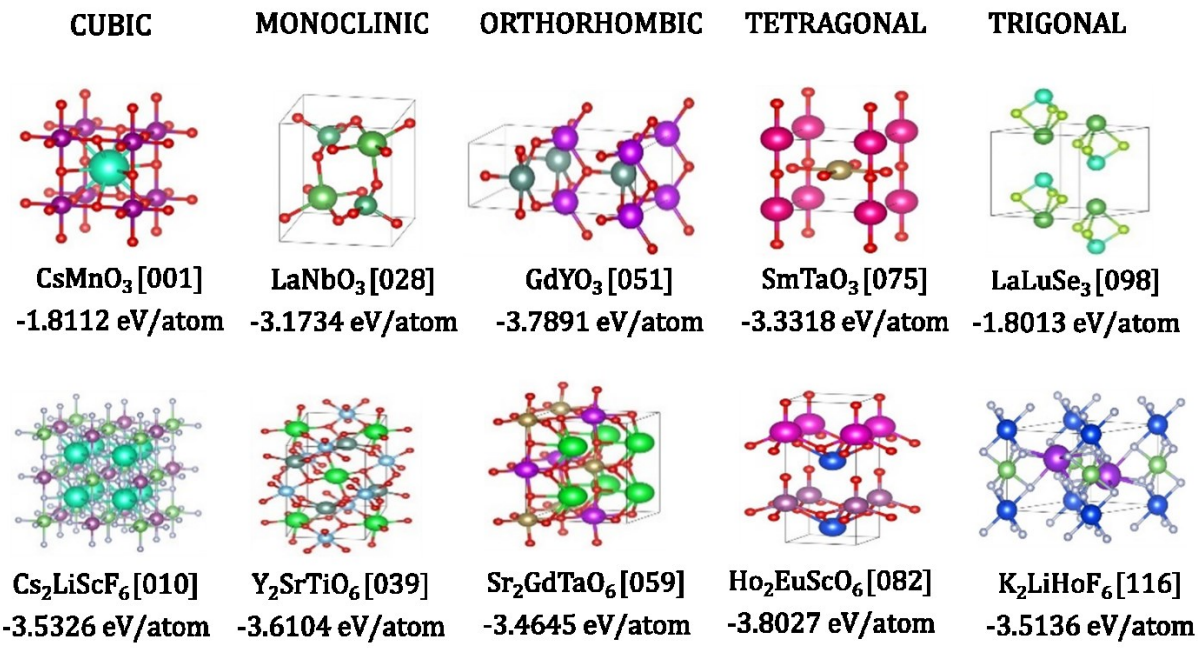

Figure 14: LCMGM designed perovskites with the best model-predicted formation energy (eV/atom) with respect to the considered crystal systems and stoichiometries.

6.2 Application in photovoltaics and optoelectronics

Perovskites are widely known to compete and advance Power Conversion Efficiency (PCE) levels in the next generation of solar cell materials due to several advantageous qualities that extend to tunable bandgaps, higher absorptivity, tandem systems, ease of fabrication, flexibility, thin-film technology, and low cost [22, 37, 117, 118]. To further demonstrate the experimental impact of this thesis, the new materials discovered in the study are investigated to identify some promising candidates with potential serviceability as host perovskites in light-harvesting applications. In classical semiconducting physics, materials with energy bandgap within the near infrared to ultraviolet region (i.e. NIR-VIS-UV) of the electromagnetic spectrum are known to quantum mechanically interact with visible light. As a result, the 265 new perovskites discovered in the study are inspected to identify some interesting candidates that have semiconducting bandgaps within the prescribed range for optical and electronical applications (i.e. 1.0 – 3.1 eV). Upon inspection, 47 materials from EVAPD and 6 materials from LCMGM are suggested to be potentially resourceful in light-harvesting energy applications. Among them, 22 candidates are seen to be specifically suitable as host materials for single junction solar cells. This special class of materials for solar cell applications possess bandgaps that are within the Shockley-Quessier ideal limits of 1-1.7 eV, as it relates to PCEs in excess of 30% [115, 116]. As such, Table 5 outlines the 22 special candidates and reports their DFT-evaluated energy bandgaps, in addition to the exchange correlation functional used for estimation. To confirm the formability/synthesizability of these special candidates, their model-predicted formation energies (E_f) are compared to another stability indicator, which in this context is the standard DFT relative energy (E_{rel}), as described previously using Equation 14. To further elucidate on the electrical behaviour of the special candidates, band-structure and Projected Density of States (P-DOS) investigations are used to categorize the bandgap property (i.e. direct or indirect). As such, Fig. 15 displays band-structure and P-DOS plots for three highly regarded materials from the LCMGM pipeline with ideal bandgaps. The materials include $TaAlO_3$ (CIF ID: *026*), Sr_2CWO_6 (CIF ID: *019*), and $RbZnH_3$ (CIF ID: *047*), and they are all defined by their bandgaps relatively close to the ideal 1.3 eV for maximum PCE. Note that for demonstrating band-structures, the *k*-points are estimated along high-symmetry line segments in the irreducible Brillouin zone of their primitive crystal structures [98, 99]. The three materials are seen to be indirect due to their valence band

maximum (VBM) and conduction band minimum (CBM) occurring at different points in the *k*-momentum space.

Table 5: Special class of materials with potential serviceability as host perovskites for single-junction solar cells.

CIF ID	Host perovskite	Crystal system	DFT-determined E_g (eV)	Model-predicted E_f (eV/atom)	DFT-determined E_{rel} (eV/atom)	GGA DFT functional
EVAPD						
003	In_2YSbO_6	Triclinic	1.3173	-2.6351	-3.0301	PBE
017	K_2UVO_6	Triclinic	1.0767	-3.2084	-4.4463	PBE
030	Na_2BiAlH_6	Triclinic	1.6872	-0.7374	-0.8165	PBE
031	Na_2BiAlO_6	Triclinic	1.7100	-2.6915	-2.4504	PBE
039	Na_2LiAlO_6	Triclinic	1.0735	-3.0514	-2.6422	PBE
056	Sr_2CaReO_6	Triclinic	1.6654	-2.8774	-4.0852	PBE
064	Sr_2LiAlH_6	Triclinic	1.2878	-0.2312	-0.9236	PBE
072	Sr_2MgMoO_6	Triclinic	1.5866	-2.8090	-4.2076	PBE
075	Sr_2MgReO_6	Triclinic	1.6831	-2.7664	-4.2708	PBE
085	Sr_2TaNbO_6	Triclinic	1.2333	-3.2132	-5.0745	PBE
099	Sr_2YUO_6	Triclinic	1.4735	-3.5173	-4.9312	PBE
113	Ti_2YBiO_6	Triclinic	1.0625	-2.8358	-4.8943	PBE
126	$SrLaBiSbO_6$	Triclinic	1.1895	-2.4481	-3.6149	PBE
131	$SrLiBiTeO_6$	Triclinic	1.2793	-2.2757	-2.3127	PBE
132	$SrLiWTeO_6$	Triclinic	1.3587	-2.4863	-3.6580	PBE
133	$SrScIrTeO_6$	Triclinic	1.2786	-2.3979	-3.8637	PBE
136	$TiLaBiTeO_6$	Triclinic	1.3341	-2.4840	-3.9605	PBE
LCMGM						
019	Sr_2CWO_6	Cubic	1.4636	-2.7890	-3.6199	PBEsol
026	$TaAlO_3$	Monoclinic	1.3661	-2.9427	-4.4095	PBEsol

047	$RbZnH_3$	Orthorhombic	1.4113	-0.3451	-0.5400	PBEsol
113	Cd_2LuClF_6	Trigonal	1.0434	-2.5610	-2.9079	PBEsol
121	Rb_2AuLuF_6	Trigonal	1.5232	-2.9630	-3.4168	PBEsol

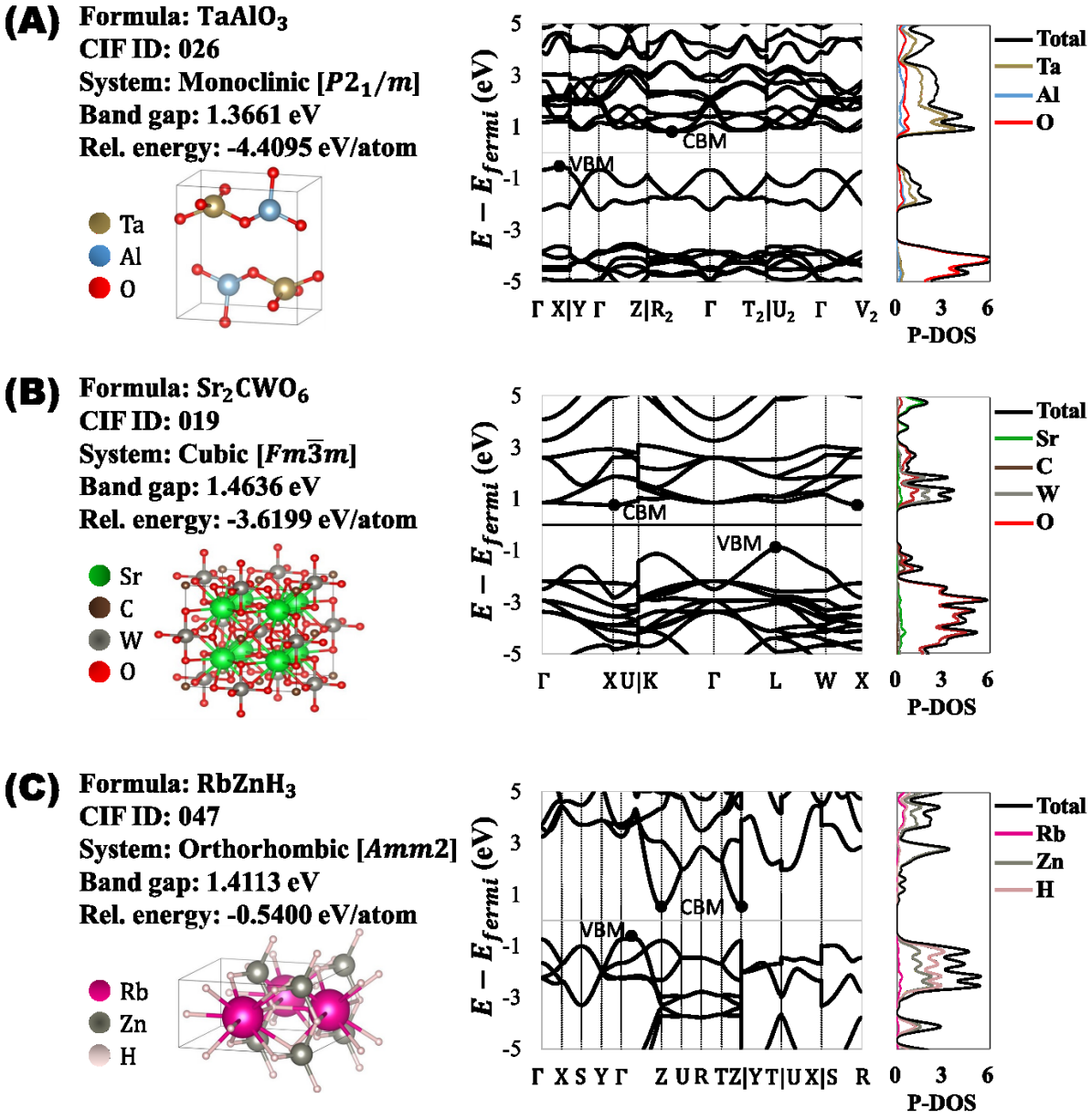

Figure 15: Band-structure and Projected Density of States (P-DOS) for three LCMGM-designed perovskites with potential serviceability as host materials in solar cells. The electronic behaviour of these materials properties was DFT-determined using PBEsol-GGA functionals.

CHAPTER 7: CONCLUSION AND FUTURE RESEARCH

7.1 Conclusion

This research work demonstrates the applicability of the data-driven Artificial Intelligence (AI) approach towards the accelerated discovery of novel energy materials that possess target specific qualities. The main contributions of this thesis are summarized as follows:

1. In the forward design simulation, the Support Vector Machine (SVM) is confirmed as the best tabular dataset model for predicting deterministic perovskite's target properties that define stability and functionality, such as stability energy, formation energy, and bandgap. Using generalized input features for describing perovskites in the training set, the Mean Absolute Error (MAE) predictive results with SVM were obtained at 0.055 eV/atom and 0.462 eV for formation energy and bandgap, respectively.
2. By hybridizing a two-dimensional convolutional neural network (Conv2D) to the SVM (i.e. Conv2D-SVM model), the target predictive result can be significantly improved, especially on the energy bandgap. The Conv2D-SVM architecture is well suited for analyzing high-dimensional data structures in the form of image-based perovskite representations. The image-based descriptor is conceptualized to include direct/real ionic input features and Fourier-transformed input features for analogously modelling the periodic effect of the reciprocal crystal lattice. The MAE results from the prediction exercise were obtained at 0.050 eV/atom, 0.058 eV/atom, and 0.105 eV for the stability energy, the formation energy and the bandgap, respectively. Besides, the achieved predictive accuracy out-performs periodic benchmark representations in the Coulomb matrix, Ewald-sum matrix, and Sine matrix by about 70%, 75% and 66%, on the same targets, respectively.
3. For the inverse design simulation, this research progressively develops three Deep Generative Models (DGM) for addressing specific challenges in novel perovskite discovery, namely Target-Learning Variational Autoencoder (TL-VAE) model, Evolutionary Variational Autoencoder for Perovskite Discovery (EVAPD), and Lattice-Constrained Materials Generative Model (LCMGM).
4. The TL-VAE model is specifically optimized on generating stable perovskites that adopt the ABX_3 stoichiometry. The TL-VAE is robustly trained on a relatively larger dataset, and as

such, can more efficiently assimilate underlying chemical laws towards the generation of uncommon forms of perovskite chemical compositions, such as antiperovskites. As a proof of concept, the TL-VAE is demonstrated to discover four new compounds. Among the new compounds is the unique $GaOPd_3$ antiperovskite structure, which is an eccentric form of ionic swapping perovskites with promising potential as superionic conductors.

5. The EVAPD model is designed to improve the TL-VAE by extending the search to other perovskite stoichiometries, in addition to incorporating the genetic algorithm (GA) for ranking quality perovskite candidates that are highly stable and synthesizable. The EVAPD led to the discovery of 137 new highly stable perovskite materials (114 $A_2BB'X_6$ and 23 $AA'BB'X_6$), of which 82 are unique and novel.
6. The LCMGM model is designed to address the challenge of geometrical inconsistencies in the form of lattice asymmetrisation that is persistent with EVAPD. The LCMGM is comprised of three distinctive modelling phases that include a semi-supervised VAE (SS-VAE) model, an Auxiliary Generative Adversarial Network (A-GAN) model, and geometrical optimization in Bayesian Optimization (BO) algorithm. The LCMGM led to the discovery of 124 new perovskite materials (65 ABX_3 and 59 $A_2BB'X_6$), of which 23 are cubic, 18 are monoclinic, 24 are orthorhombic, 25 are tetragonal and 34 are trigonal structures. The crystal systems of the new materials were found to be consistent with the predefined geometrical lattice constraints at the encoding phase.
7. Furthermore, to highlight the experimental relevance of this thesis for energy applications, the 265 newly discovered materials were investigated to identify promising candidates with potential serviceability in light-harvesting applications. Among them, 53 are suggested to be reliable as host perovskite candidates in photovoltaic and/or optoelectronic applications. Specifically, 22 new materials are highly regarded for single junction solar cells due to their bandgaps within Shockley-Quessier ideal limits for optimized Power Conversion Efficiency (PCE) in excess of 30%. Some prominent examples with bandgaps close to the ideal 1.3 eV include $TaAlO_3$, Sr_2CWO_6, and $RbZnH_3$.

Finally, the total of 265 new materials generated by the TL-VAE, EVAPD and LCMGM inverse design pipelines were validated via Density Functional Theory (DFT) and are openly made available in the Novel Materials Discovery (NOMAD) and Mendeley data repositories. The

preprocessed dataset used for machine and deep learning, relevant source codes for designing the forward and inverse design pipelines, and Crystallographic Information Files (CIF) of newly generated materials are openly made available on GitHub (github.com/chenebuah).

7.2 Future research work

The developed forward and inverse design models highlight the important role AI can play in accelerated materials discovery. However, areas of potential improvements remain and can be addressed in future studies. Some identified research opportunities are therefore discussed as follows.

1. The present thesis uses the theoretical first-principles approach via DFT technique to validate the 265 new materials emerging from the inverse design pipelines. In essence, DFT validation assists in limiting laboratory experimentation to the most promising candidates. Hence, for unambiguity and final verification on formability and functionality, the materials should undergo laboratory synthesization in order to ascertain the modelling efficiencies and accuracies of the developed AI algorithms in the present study. In the synthesization process, high priority should be given to materials that were predicted to be highly stable, as well as those identified as potential host perovskites for light-harvesting applications. Moreover, AI can equally be used in the synthesization phase for optimizing manufacturing process parameters, as demonstrated in previous studies [122-123].
2. The inferences obtained from the synthesization process can be used adaptively to improve the modelling frameworks of the forward and inverse design models. By integrating *on-the-fly* DFT and experimental feedback algorithms into the overall modelling framework, the success rate of generating newer perovskite chemistries can be further optimized.
3. In the present research, the inverse design models were optimized on target properties that majorly define stability and formability. However, to develop definitive frameworks that equally address specific engineering applications, the targets should be expanded to other functional properties. For instance, some functional targets that are of higher interest to materials scientists and engineers include energy bandgap, ionic conductivity, Curie temperature, dielectric susceptibility, etc. The development of application-specific AI models

will first necessitate targeted-data extraction from reliable materials database platforms prior to forward design simulation for accurate target prediction. The results and inferences obtained from the forward design simulation can serve as a basis for designing efficient semi-supervisory models for addressing the inverse design simulation.

4. The invertible image-based descriptor concept used for feature engineering of perovskites can substantially be modified to capture other complex forms of perovskites that were not addressed in the present research, such as $A_3BB'X_9$, $A'_2A_{n-1}B_nX_{3n+1}$ (i.e. Ruddlesden-Popper structure), Hybrid Organic-Inorganic Perovskite (HOIP), and doped forms. Moreover, the descriptor and modelling frameworks could be broadened to accommodate material classes other than perovskites, such as high entropy alloys that are equally complex and very promising multi-element materials, in order to generalize the simulation process to a variety of unknown compounds and/or stoichiometries.
5. The present study predominantly applies variational autoencoders (VAE), generative adversarial networks (GAN) and spherical linear interpolation (SLERP) models in designing the inverse design pipeline. Other interesting architectures that could equally be explored in future studies include Graph Neural Networks (GNN) and thermodynamic diffusion models. GNNs are well utilized in novel drug discovery [2], and as such, could equally be applied to solve materials challenge. Moreover, diffusion models are gaining significant interest among researchers for solving general challenges in computer vision and imaging, and hence, can be used in the deep generative modelling of new materials [124].

REFERENCES

1. A.J. Cohen, P. Mori-Sánchez, W. Yang, Challenges for Density Functional Theory, Chem. Rev., 112(1), (2012), 289-320. https://doi.org/10.1021/cr200107z

2. M. Mukaidaisi, A. Vu, K. Grantham, A. Tchagang, Y. Li, Multi-Objective Drug Design Based on Graph-Fragment Molecular Representation and Deep Evolutionary Learning, Front. Pharmacol., 13, 920747, (2022). https://doi.org/10.3389/fphar.2022.920747

3. M. Cunneen, M. Mullins, F. Murphy, Autonomous Vehicles and Embedded Artificial Intelligence: The Challenges of Framing Machine Driving Decisions, Appl. Artif. Intell., 33(8), (2019), 706-731. https://doi.org/10.1080/08839514.2019.1600301

4. J. Andreu-Perez, F. Deligianni, D. Ravi, G-Z. Yang, Artificial Intelligence and Robotics, arXiv:1803.10813v1 [cs.AI], (2018). https://doi.org/10.48550/arXiv.1803.10813

5. R. Chen, M. Wang, Y. Lai, Analysis of the role and robustness of artificial intelligence in commodity image recognition under deep learning neural network, Plos One, 15(7), e0235783, (2020). https://doi.org/10.1371/journal.pone.0235783

6. M. Alam, M.D. Samad, L. Vidyaratne, A. Glandon, K.M. Iftekharuddin, Survey on Deep Neural Networks in Speech and Vision Systems, arXiv:1908.07656v2 [cs.CV], (2019). https://doi.org/10.48550/arXiv.1908.07656

7. R.B. Asha, K.R.S. Kumar, Credit card fraud detection using artificial neural network, Glob. Trans. Proceed., 2(1), 35-41, (2021). https://doi.org/10.1016/j.gltp.2021.01.006

8. W. Li, R. Jacobs, D. Morgan, Predicting the thermodynamic stability of perovskite oxides using machine learning models, Comput. Mater. Sci, 150, (2018), 454-463. https://doi.org/10.1016/j.commatsci.2018.04.033

9. C. Li, H. Hao, B. Xu, G. Zhao, L. Chen, S. Zhang, H. Liu, A progressive learning method for predicting the band gap of ABO3 perovskites using an instrumental variable, J. Mater. Chem. C, 8, (2020), 3127-3136. https://doi.org/10.1039/C9TC06632B

10. W.B. Park, J. Chung, J. Jung, K. Sohn, S.P. Singh, M. Pyo, N. Shin, K-S Sohn, Classification of crystal structure using a convolutional network, IUCrJ, 4(4), (2017), 486-494. https://doi.org/10.1107/S205225251700714X

11. Y. Zhao, Y. Cui, Z. Xiong, J. Jin, Z. Liu, R. Dong, J. Hu, Machine Learning-Based Prediction of Crystal Systems and Space Groups from Inorganic Materials Compositions, ACS Omega, 5(7), (2020), 3596-3606. https://doi.org/10.1021/acsomega.9b04012

12. T. Mueller, A. Hernandez, C. Wang, Machine learning for interatomic potential models, J. Chem. Phys. 152, 050902, (2020). https://doi.org/10.1063/1.5126336

13. J. Noh, J. Kim, H.S. Stein, B. Sanchez-Lengeling, J.M. Gregoire, A. Aspuru-Guzik, Y. Jung, Inverse Design of Solid-State Materials via a Continuous Representation, Matter, 1(5), (2019), 1370-1384. https://doi.org/10.1016/j.matt.2019.08.017

14. Z. Ren, et al., An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties, Matter, 5(1), (2022), 314-335. https://doi.org/10.1016/j.matt.2021.11.032

15. T. Long, N.M. Fortunato, I. Opahle, Y. Zhang, I. Samathrakis, C. Shen, O. Gutfleisch, H. Zhang, Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures, npj Comput. Mater., 7, 66, (2021), https://doi.org/10.1038/s41524-021-00526-4

16. A. Walsh, The quest for new functionality, Nature Chem., 7, (2015), 274–275. https://doi.org/10.1038/nchem.2213

17. S. Choi, S. Johnston, W-J. Jang, K. Koepernik, K. Nakatsukasa, J.M. Ok, H-J. Lee, H.W. Choi, A.T. Lee, A. Akbari, Y.K. Semertzidis, Y. Bang, J.S. Kim, J. Lee, Correlation of Fe-Based Superconductivity and Electron-Phonon Coupling in an FeAs/Oxide Heterostructure, Phys. Rev. Lett., 119(10), (2017), 107003. https://doi.org/10.1103/PhysRevLett.119.107003

18. G.J. La O', S.J. Ahn, E. Crumlin, Y. Orikasa, M.D. Biegalski, H.M. Christen, Y. Shao-Horn, Catalytic activity enhancement for oxygen reduction on epitaxial perovskite thin films for solid-oxide fuel cells, Angew. Chem., Int. Ed., 49(31), (2010), 5344–5347. https://doi.org/10.1002/anie.201001922

19. Z. Chen, X. Zou, W. Ren, L. You, C. Huang, Y. Yang, P. Yang, J. Wang, T. Sritharan, L. Bellaiche, L. Chen, Study of strain effect on in-plane polarization in epitaxial BiFeO3 thin films using planar electrodes, Phys. Rev. B, 86(23), (2012), 235125. https://doi.org/10.1103/PhysRevB.86.235125

20. X-H Du, U. Belegundu, K. Uchino, Crystal Orientation Dependence of Piezoelectric Properties in Lead Zirconate Titanate: Theoretical Expectation for Thin Films, Jpn. J. Appl. Phys., 36, (1997), 5580-5587. https://doi.org/10.1143/jjap.36.5580.

21. Y. Liu, H. Ye, Y. Zhang, K. Zhao, Z. Yang, Y. Yuan, H. Wu, G. Zhao, Z. Yang, J. Tang, Z. Xu, S. Liu, Surface-Tension-Controlled Crystallization for High-Quality 2D Perovskite Single Crystals for Ultrahigh Photodetection, Matter, 1(2), (2019), 465-480. https://doi.org/10.1016/j.matt.2019.04.002

22. J.H. Noh, S.H. Im, J.H. Heo, T.N. Mandal, S. Seok, Chemical Management for Colorful, Efficient, and Stable Inorganic–Organic Hybrid Nanostructured Solar Cells, Nano Lett., 13(4), (2013), 1764-1769. https://doi.org/10.1021/nl400349b

23. D. Jia, M. Xu, S. Mu, W. Ren, C. Liu, Recent Progress of Perovskite Nanocrystals in Chem/Bio Sensing, Biosensors, 12(9), 754, (2022). https://doi.org/10.3390/bios12090754

24. A. Emery, C. Wolverton, High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO_3 perovskites. Sci. Data, 4, 170153 (2017). https://doi.org/10.1038/sdata.2017.153

25. S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, npj Comput. Mater., 1, 15010, (2015), https://doi.org/10.1038/npjcompumats.2015.10

26. Z. Hu, Z. Lin, J. Su, J. Zhang, J. Chang, Y. Hao, A Review on Energy Band-Gap Engineering for Perovskite Photovoltaics, Sol. RRL, 3(12), (2019), 1900304. https://doi.org/10.1002/solr.201900304

27. A. Jain, G. Hautier, C. J. Moore, S.P. Ong, C.C. Fischer, T. Mueller, K.A. Persson, G. Ceder, A high-throughput infrastructure for density functional theory calculations, Comput. Mater. Sci., 50(8), (2011), 2295–2310. https://doi.org/10.1016/j.commatsci.2011.02.023

28. A. Belsky, M. Hellenbrandt, V.L. Karen, P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design, Acta Cryst., B58, (2002), 364-369. https://doi.org/10.1107/S0108768102006948

29. R. Allmann, R. Hinek, The introduction of structure types into the Inorganic Crystal Structure Database ICSD, Acta Cryst., A63, (2007), 412-417. https://doi.org/10.1107/S0108767307038081

30. C.J. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res., 30, (2005), 79–82, https://doi.org/10.3354/cr030079

31. K.P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, Cambridge, (2012).

32. S.H. Simon, The Oxford Solid State Basics, Oxford University Press, Oxford, (2013).

33. P. Giannozzi, et al., QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, J. Phys.:Condens. Matter, 21(39), 395502, (2009). https://doi.org/10.1088/0953-8984/21/39/395502

34. J.P. Attfield, P. Lightfoot, R.E. Morris, Perovskites, Dalton Trans., 44(23), (2015), 10541-10542. http://dx.doi.org/10.1039/C5DT90083B

35. M. Johnsson, P. Lemmens, Crystallography and Chemistry of Perovskites. In Handbook of Magnetism and Advanced Magnetic Materials (eds H. Kronmüller, S. Parkin, M. Coey, A. Inoue and H. Kronmüller), (2007). https://doi.org/10.1002/9780470022184.hmm411

36. T. Saha-Dasgupta, Double perovskites with 3d and 4d/5d transition metals: compounds with promises, Mater. Res. Express, 7(1), (2020), 014003. https://doi.org/10.1088/2053-1591/ab6293

37. D.A. Egger, A.M. Rappe, L. Kronik, Hybrid Organic–Inorganic Perovskites on the Move, Acc. Chem. Res., 49(3), (2016), 573-581. https://doi.org/10.1021/acs.accounts.5b00540

38. Y. Wang, H. Zhang, J. Zhu, X. Lü, S. Li, R. Zou, Y. Zhao, Antiperovskites with Exceptional Functionalities, Adv. Mater., 32, (2020), 1905007. https://doi.org/10.1002/adma.201905007

39. M.W. Lufaso, P.M. Woodward, Jahn–Teller distortions, cation ordering and octahedral tilting in perovskites. Acta Cryst. B, 60, (2004), 10-20. https://doi.org/10.1107/S0108768103026661

40. P.M. Woodward, Octahedral Tilting in Perovskites. I. Geometrical Considerations, Acta Cryst. B, (1997), 32-43. https://doi.org/10.1107/S0108768196010713

41. P.M. Woodward, Octahedral Tilting in Perovskites. II. Structure Stabilizing Forces, Acta Cryst. B, (1997), 44-66. https://doi.org/10.1107/S0108768196012050

42. M.C. Knapp, P.M. Woodward, A-site cation ordering in $AA'BB'O_6$ perovskites, J. Solid State Chem., 179(4), (2006), 1076-1085. https://doi.org/10.1016/j.jssc.2006.01.005

43. R. Mitchell, M. Welch, A. Chakhmouradian, Nomenclature of the perovskite supergroup: A hierarchical system of classification based on crystal structure and composition. Mineral Mag., 81(3), (2017), 411-461. https://doi.org/10.1180/minmag.2016.080.156

44. S. Behara, T. Poonawala, T. Thomas, Crystal structure classification in ABO_3 perovskites via machine learning, Comput. Mater Sci., 188, 110191, (2021). https://doi.org/10.1016/j.commatsci.2020.110191

45. L. Fu, B. Li, S. Li, L. Yin, Magnetic, Electronic, and Optical Properties of Perovskite Materials. In: Arul, N., Nithya, V. (eds) Revolution of Perovskite. Materials Horizons: From Nature to Nanomaterials. Springer, Singapore. (2020). https://doi.org/10.1007/978-981-15-1267-4_2

46. V.M. Goldschmidt, Die gesetze der krystallochemie, Naturwissenschaften, 14, (1926), 477-485. https://doi.org/10.1007/BF01507527

47. S.A. Hollingsworth, R.O. Dror, Molecular Dynamics Simulation for All, Neuron, 99(6), (2018), 1129-1143. https://doi.org/10.1016/j.neuron.2018.08.011

48. J.M. Sellier, M.Nedjalkov, I. Dimov, An introduction to applied quantum mechanics in the Wigner Monte Carlo formalism, Phys. Rep., 577, (2015), 1-34. https://doi.org/10.1016/j.physrep.2015.03.001

49. K.D. Jayan, V. Sebastian, Ab initio DFT determination of structural, mechanical, optoelectronic, thermoelectric and thermodynamic properties of $RbGeI_3$ inorganic perovskite for different exchange-correlation functionals, Mater. Today Commun., 28, (2021), 102650. https://doi.org/10.1016/j.mtcomm.2021.102650

50. G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci., 6(1), (1996), 15-50. https://doi.org/10.1016/0927-0256(96)00008-0

51. K. O'Shea, R. Nash, An Introduction to Convolutional Neural Networks, arXiv:1511.08458v2 [cs.NE], (2015). https://doi.org/10.48550/arXiv.1511.08458

52. H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, In: NIPS'96. Denver, MIT Press, (1996), 155–161.

53. S. Haykin, Neural networks: a comprehensive foundation, Prentice Hall PTR, (1994).

54. B. Ghojogh, M. Crowley, Unsupervised and Supervised Principal Component Analysis: Tutorial, arXiv:1906.03148 [stat.ML], (2019). https://doi.org/10.48550/arXiv.1906.03148

55. D.P. Kingma, M. Welling, An Introduction to Variational Autoencoders, arXiv:1906.02691v3 [cs.LG], (2019). https://doi.org/10.48550/arXiv.1906.02691

56. A. Jain, S.P. et al., The Materials Project: A materials genome approach to accelerating materials innovation, APL Materials, 1(1), (2013), 011002. https://doi.org/10.1063/1.4812323

57. S. P. Ong, et al., Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci., 68, (2013), 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028

58. J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD), JOM, 65, (2013), 1501-1509. https://doi.org/10.1007/s11837-013-0755-4

59. S. Curtarolo, W. Setyawan, et al., Aflow: An automatic framework for high-throughput materials discovery. Comput. Mater. Sci., 58, (2012), 218-226. https://doi.org/10.1016/j.commatsci.2012.02.005

60. C. Draxl, M. Scheffler, Nomad: The fair concept for big data-driven materials science. MRS Bull., 43, (2018), 676-682. https://doi.org/10.1557/mrs.2018.208

61. L. Talirz, S. Kumbhar, E. Passaro, et al. Materials cloud, a platform for open computational science. Sci. Data, 7, (2020), 1–12. https://doi.org/10.1038/s41597-020-00637-5

62. A. Vaitkus, A. Merkys, S. Gražulis, Validation of the Crystallography Open Database using the Crystallographic Information Framework, J. Appl. Cryst., 54(2), 661-672. https://doi.org/10.1107/S1600576720016532

63. T. Xie and J.C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, Phys. Rev. Lett., 120(14), (2018), 145301. https://doi.org/10.1103/physrevlett.120.145301

64. L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, npj Comput. Mater., 2, (2016), 16028. https://doi.org/10.1038/npjcompumats.2016.28

65. F. Faber, A. Lindmaa, O.A. von Lilienfeld, R. Armiento, Crystal structure representations for machine learning models of formation energies, IJQC, 115, (2015), 1094-1101. https://doi.org/10.1002/qua.24917

66. Z. Guo, B. Lin, Machine learning stability and band gap of lead-free halide double perovskite materials for perovskite solar cells, Solar Energy, 228, (2021), 689-699. https://doi.org/10.1016/j.solener.2021.09.030

67. P. Omprakash, B. Manikandan, A. Sandeep, R. Shrivastava, V. P., D.B. Panemangalore, Graph representational learning for bandgap prediction in varied perovskite crystals, Comput. Mater. Sci., 196, (2021), 110530. https://doi.org/10.1016/j.commatsci.2021.110530

68. J.P. Perdew, Density functional theory and the band gap problem, Int. J. Quantum Chem., 28, (1985), 497–523. https://doi.org/10.1002/qua.560280846

69. M. Rupp, A. Tkatchenko, K-R. Müller, O.A. von Lilienfeld, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, Phys. Rev. Lett., 108(5), (2012). https://doi.org/10.1103/PhysRevLett.108.058301

70. P.P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, Ann. Phys., 369, (1921), 253-287. https://doi.org/10.1002/andp.19213690304

71. I-H Lee, K.J. Chang, Crystal structure prediction in a continuous representative space, Comput. Mater. Sci., 194, 110436, (2021). https://doi.org/10.1016/j.commatsci.2021.110436

72. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks, arXiv:1406.2661v1 [stat.ML], (2014). https://doi.org/10.48550/arXiv.1406.2661

73. C. Suryanarayana, Structure and properties of nanocrystalline materials, Bull. Mater. Sci., 17, (1994), 307–346. https://doi.org/10.1007/BF02745220

74. D.P. Kingma, N. Welling, Auto-Encoding Variational Bayes, arXiv:1312.6114v11 [stat.ML], (2013). https://doi.org/10.48550/arXiv.1312.6114

75. K. Shoemake, Animating rotation with quaternion curves, SIGGRAPH Comput. Graph, 19(3), (1985), 245–254. https://doi.org/10.1145/325165.325242

76. Z. Michalewicz, M. Schoenauer, Evolutionary algorithms for constrained parameter optimization problems, Evol. Comput., 4(1), (1996), 1–32. https://doi.org/10.1162/evco.1996.4.1.1

77. J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple. Phys. Rev. Lett., 77(18), (1996), 3865-3868. https://doi.org/10.1103/PhysRevLett.77.3865

78. P.E. Blöchl, Projector augmented-wave method. Phys. Rev. B, 50(24), (1994), 17953-17979. https://doi.org/10.1103/PhysRevB.50.17953

79. C. Kim, T. Huan, S. Krishnan, R. Ramprasad, A hybrid organic-inorganic perovskite dataset, Sci. Data 4, (2017), 170057. https://doi.org/10.1038/sdata.2017.57

80. J. Heyd, G.E. Scuseria, M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, J. Chem. Phys., 118(18), (2003), 8207–8215. https://doi.org/10.1063/1.1564060

81. S. Nieto, R. Polanco, R. Roque-Malherbe, Absorption kinetics of hydrogen in nanocrystals of $BaCe_{0.95}Yb_{0.05}O_{3-\delta}$ proton conducting perovskite, J Phys. Chem. C., 111(6), (2007), 2809–2818. https://doi.org/10.1021/jp067389i

82. N.F. Atta, A. Galal, E.H. El-Ads, Perovskite Nanomaterials – Synthesis, Characterization, and Applications, In L. Pan, & G. Zhu (Eds.), Perovskite Materials - Synthesis, Characterisation, Properties, and Applications, (2016). https://doi.org/10.5772/61280

83. E.A.R. Assirey, Perovskite synthesis, properties and their related biochemical and industrial application, SPJ, 27(6), (2019), 817-829. https://doi.org/10.1016/j.jsps.2019.05.003

84. S.M. Selbach, M.A., Einarsrud, T. Tybell, T. Grande, Synthesis of $BiFeO_3$ by Wet Chemical Methods, J. Am. Ceram. Soc., 90, (2007), 3430-3434. https://doi.org/10.1111/j.1551-2916.2007.01937.x

85. P.M. Mankiewich, J.H. Scofield, W.J. Skocpol, R.E. Howard, A.H. Dayem, E. Good, Reproducible technique for fabrication of thin films of high transition temperature superconductors, Appl. Phys. Lett., 51(21), (1987), 1753. https://doi.org/10.1063/1.98513

86. A. Galal, N.F. Atta, S.M.Ali, Investigation of the catalytic activity of $LaBO_3$ (B = Ni, Co, Fe or Mn) prepared by the microwave-assisted method for hydrogen evolution in acidic medium, Electrochim. Acta., 56(16), (2011), 5722-5730. https://doi.org/10.1016/j.electacta.2011.04.045

87. L. Alzubaidi, J. Zhang, A.J. Humaidi, et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data, 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

88. C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation Functions: Comparison of trends in Practice and Research for Deep Learning, arXiv:1811.03378v1 [cs.LG], (2018). https://doi.org/10.48550/arXiv.1811.03378

89. V.N. Vapnik, Statistical Learning Theory, Wiley, New York, (1998).

90. S. Kullback, R.A. Leibler, On Information and Sufficiency. The Annals of Mathematical Statistics. JSTOR. 22(1), (1951), 79-86. https://doi.org/10.1214/aoms/1177729694

91. D.P. Kingma, D.J. Rezende, S. Mohamed, M. Welling, M, Semi-Supervised Learning with Deep Generative Models, arXiv:1406.5298v2 [cs.LG], (2014). https://doi.org/10.48550/arXiv.1406.5298

92. L. Himanen, M.O.J. Jäger, E.V. Morooka, et al., DScribe: Library of descriptors for machine learning in materials science, Comput. Phys. Commun., 247, (2020), 106949. https://doi.org/10.1016/j.cpc.2019.106949

93. D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci., 28(1), (1988), 31–36. https://doi.org/10.1021/ci00057a005

94. E. Mansimov, O. Mahmood, S. Kang, S., K. Cho, Molecular Geometry Prediction using a Deep Generative Graph Neural Network. Sci. Rep., 9, 20381, (2019). https://doi.org/10.1038/s41598-019-56773-5

95. V.I. Hedge, C.K.H. Borg, Z. del Rosario et al., Reproducibility in high-throughput density functional theory: a comparison of AFLOW, Materials Project, and OQMD, arXiv:2007.01988v1, (2020). https://doi.org/10.48550/arXiv.2007.01988

96. K. Kamnitsas, D.C. Castro, L. Le Folgoc, I. Walker, R. Tanno, et al., Semi-supervised learning via compact latent space clustering, arXiv:1806.02679v2 [cs.LG], (2018). https://doi.org/10.48550/arXiv.1806.02679

97. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE, JMLR, 9 (86), (2008). 2579-2605.

98. W. Setyawan, S. Curtarolo, High-throughput electronic band structure calculations: Challenges and tools, Comput. Mater. Sci., 49(2), (2010), 299-312. https://doi.org/10.1016/j.commatsci.2010.05.010

99. Y. Hinuma, G. Pizzi, Y. Kumagai, F. Oba, I. Tanaka, Band structure diagram paths based on crystallography, Comp. Mater. Sci., 128, 140 (2017). https://doi.org/10.1016/j.commatsci.2016.10.015

100. G. Prandini, A. Marrazzo, I.E. Castelli, N. Mounet, N. Marzari, Precision and efficiency in solid-state pseudopotential calculations, npj Comput. Mater., 4, 72, (2018). https://doi.org/10.1038/s41524-018-0127-2

101. J. Snoek, H. Larochelle, R.P. Adams, Practical Bayesian optimization of machine learning algorithms, arXiv:1206.2944v2 [stat.ML], (2012). https://doi.org/10.48550/arXiv.1206.2944

102. G. Pilania, P.V. Balachandran, C. Kim, T. Lookman, Finding new perovskite halides via machine learning, Front. Mater., 3(19), (2016). https://doi.org/10.3389/fmats.2016.00019

103. S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning, Nat. Comm., 9(1) (2018), 3405. https://doi.org/10.1038/s41467-018-05761-w

104. A.K. Singh, J.H. Montoya, J.M. Gregoire, K.A. Persson, Robust and synthesizable photocatalysts for CO_2 reduction: a data-driven materials discovery, Nat. Commun., 10, 443, (2019). https://doi.org/10.1038/s41467-019-08356-1

105. D. Liu, Mathematical modeling analysis of genetic algorithms under schema theorem, JCMSE, 19(21), (2019), 131–137. https://doi.org/10.3233/JCM-191019

106. A.F. Gad, PyGAD: An Intuitive Genetic Algorithm Python Library, arXiv:2106.06158v1 [cs.NE], (2021). https://doi.org/10.48550/arXiv.2106.06158

107. J. Zheng, B. Perry, Y. Wu, Antiperovskite superionic conductors: a critical review, ACS Mater. Au., 1(2), (2021), 92-106. https://doi.org/10.1021/acsmaterialsau.1c00026

108. R.D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, Acta Cryst., A32, (1976), 751-767. https://doi.org/10.1107/S0567739476001551

109. M. Abadi, P. Barham, J. Chen, et al. TensorFlow: a system for large-scale machine learning, arXiv:1605.08695v2 [cs.DC], (2016). https://doi.org/10.48550/arXiv.1605.08695

110. F. Tran. P. Blaha, Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential, Phys. Rev. Lett., 102(22), (2009), 226401. https://doi.org/10.1103/PhysRevLett.102.226401

111. X-G, Zhao, G.M. Dalpian, A. Wang, A. Zunger, Polymorphous nature of cubic halide perovskites, Phys. Rev. B., 101, (2020), 155137. https://doi.org/10.1103/PhysRevB.101.155137

112. R. Fletcher, C.M. Reeves, Function minimization by conjugate gradients. Comput. J., 7(2), (1964), 149-154. https://doi.org/10.1093/comjnl/7.2.149

113. Y. Zuo, M. Qin, C. Chen, W. Ye, X. Li, J. Luo, S.P. Ong, Accelerating materials discovery with Bayesian optimization and graph deep learning, Mater. Today, 51, (2021), 126-135. https://doi.org/10.1016/j.mattod.2021.08.012

114. D. Packwood, Bayesian optimization for materials science, Springer Singapore, (2017). https://doi.org/10.1007/978-981-10-6781-5

115. W. Shockley, H.J. Queisser, Detailed balance limit of efficiency of p-n junction solar cells, J. Appl. Phys., 32(3), (1961), 510–519. https://doi.org/10.1063/1.1736034

116. S. Rühle, Tabulated values of the Shockley–Queisser limit for single junction solar cells, Sol. Energy, 130, (2016), 139-147. https://doi.org/10.1016/j.solener.2016.02.015

117. N.S. Kumar, K.C.B. Naidu, A review on perovskite solar cells (PSCs), materials and applications, J. Mater, 7(5), (2021), 940-956. https://doi.org/10.1016/j.jmat.2021.04.002

118. T. Wu, Z. Qin, Y. Wang, et al., The main progress of perovskite solar cells in 2020–2021, Nano-Micro Lett., 13, 152, (2021). https://doi.org/10.1007/s40820-021-00672-w

119. L. Wang, J. Wu, S. Wang, H. Liu, Y. Wang, D. Wang, The reformation of catalyst: From a trial-and-error synthesis to rational design, Nano Res., 17(4), (2024), 3261-3301. https://doi.org/10.1007/s12274-023-6037-8

120. S. Katoch, S.S. Chauhan, V. Kumar, A review on genetic algorithm: past, present, and future, Multimed Tools Appl 80, (2021), 8091–8126. https://doi.org/10.1007/s11042-020-10139-6

121. M. Bogojeski, L. Vogt-Maranto, M.E. Tuckerman, et al., Quantum chemical accuracy from density functional approximations via machine learning, Nat Commun., 11, (2020), 5223. https://doi.org/10.1038/s41467-020-19093-1

122. D. Weichert, P. Link, A. Stoll, et al., A review of machine learning for the optimization of production processes, Int J Adv Manuf Technol 104, (2019), 1889–1902. https://doi.org/10.1007/s00170-019-03988-5

123. H.S. Park, D.S. Nguyen, T. Le-Hong, et al., Machine learning-based optimization of process parameters in selective laser melting for biomedical applications, J Intell. Manuf. 33, (2022), 1843–1858. https://doi.org/10.1007/s10845-021-01773-4

124. P. Lyngby, K.S. Thygesen, Data-driven discovery of 2D materials by deep generative models. npj Comput. Mater., 8, 232, (2022). https://doi.org/10.1038/s41524-022-00923-3

Appendix

Altogether, this thesis presents 265 new perovskite materials (69 ABX_3, 173 $A_2BB'X_6$ and 23 $AA'BB'X_6$ compounds) from the developed inverse design pipelines. All new materials were validated by DFT and were confirmed to converge upon spin-polarized relaxation calculations on all angles and axes of the variable-cell geometry, using predominantly PBE-GGA functionals. A detailed summary of the newly discovered materials, in addition to their determined properties, are listed as follows.

Table A1: Crystallographic features and property determination of new perovskites from the TL-VAE model. All DFT-determined properties and structural relaxations were confirmed using the PBE-GGA functional.

Perovskite	A (Å)	B (Å)	C (Å)	Alpha (°)	Beta (°)	Gamma (°)	DFT-determined E_g (eV)	Model predicted E_f (eV/atom)
$AlPtS_3$	5.706	5.879	8.010	132.57	72.13	102.45	0.000	0.009
$AlPtO_3$	3.653	6.749	6.703	124.72	74.93	101.10	0.304	-0.607
$GaOPd_3$	4.906	6.333	5.402	80.56	117.18	103.55	0.000	-0.575
$GaPdO_3$	5.178	5.635	8.129	68.15	134.30	116.48	0.106	-0.854

Table A2: Fully-relaxed new perovskites from the EVAPD model. Unique and novel compositions are marked in orange background, and are 82 in total. The remainders are polymorphs of previously discovered chemical compositions.

CIF ID	New perovskites	Model-predicted E_{hull} (eV/atom)	Model-predicted E_f (eV/atom)	DFT-determined E_g (eV)	DFT-determined magnetization (Bohr Mag/cell)	DFT-relaxed unit cell volume (Å^3)
$A_2BB'X_6$ compounds						
1	Ca_2YOsO_6	0.0856	-2.7913	0	3	187.9863
2	In_2YOsO_6	0.048	-2.4760	0	1.30	186.2338
3	In_2YSbO_6	0.0443	-2.6351	1.3173	0	218.1894
4	K_2LiAlF_6	0.0083	-3.3828	7.4390	0	180.0827
5	K_2LuSbO_6	0.0812	-2.9783	2.6903	0	246.4022
6	K_2LuTaO_6	0.0963	-3.3696	2.9115	0	234.0484
7	K_2MgVO_6	0.0075	-2.9025	0	1	199.1696
8	K_2MgWO_6	0.0261	-2.7115	2.0025	0	208.3489
9	K_2NaAlF_6	0.0228	-3.3659	6.8882	0	198.0626
10	K_2NaVO_6	0.0397	-2.4516	0	2	222.1200
11	K_2NaWO_6	0.0598	-2.6772	0	1	237.0695
12	K_2SmVO_6	0.0283	-3.0930	0	3	214.8249
13	K_2TaInO_6	0.0615	-2.8902	2.4008	0	228.2666
14	K_2TaPdO_6	0.0855	-2.5618	0.0852	1	285.8298
15	K_2TaSbO_6	0.1028	-2.9635	2.2023	0	184.0429
16	K_2TaVO_6	0.0588	-3.1233	0.7552	0	208.1623
17	K_2UVO_6	0.0268	-3.2084	1.0767	1	193.6815
18	K_2UZnO_6	0.0372	-3.0309	1.8107	0	254.7339
19	La_2CaOsO_6	0.0887	-2.7011	0	2	204.5027
20	La_2MgIO_6	0.2091	-2.8906	0.1673	0.97	207.4252
21	La_2MgSnO_6	0.0440	-3.0197	3.9553	0	183.9209
22	La_2MgUO_6	0.0720	-3.3863	0.1783	2	200.8788
23	La_2MgZrO_6	0.1016	-3.3645	4.0578	0	183.2739

24	La_2NaSnO_6	0.0364	-2.7480	0	0.89	215.4072
25	La_2NbZnO_6	0.0800	-2.9022	2.1099	0.99	143.4851
26	La_2SrUO_6	0.0762	-3.2995	0	2	240.0037
27	La_2SrWO_6	0.0886	-2.9263	0.3512	0	223.8164
28	La_2TaInO_6	0.0579	-3.0543	1.9279	0	227.6226
29	La_2TaNbO_6	0.0904	-3.2607	0	0.02	167.6303
30	Na_2BiAlH_6	0.0637	-0.7374	1.6872	0	230.9296
31	Na_2BiAlO_6	0.0278	-2.6915	1.7100	0	189.1179
32	Na_2BiIrH_6	0.0604	-0.7101	0	0	232.7939
33	Na_2BiIrO_6	0.0193	-1.8546	0.0350	0	236.5911
34	Na_2CaAlO_6	0.0611	-3.1006	0	1	191.4990
35	Na_2CaMoO_6	0.0604	-2.6822	2.5511	0	264.3304
36	Na_2CaOsO_6	0.0565	-2.4036	0.28	0	197.6243
37	Na_2LiAlF_6	0.0326	-3.3884	4.8759	0	204.3696
38	Na_2LiAlH_6	0.0218	-0.2936	0	0	162.0319
39	Na_2LiAlO_6	0.0152	-3.0514	1.0735	0	204.0824
40	Na_2LiIrH_6	0.0297	-0.6518	3.1077	0	136.3614
41	Na_2LiIrO_6	0.0162	-2.1437	0	1.31	182.7216
42	Na_2LiMoO_6	0.0183	-2.1355	0	0	226.0315
43	Na_2LiOsO_6	0.0393	-2.3692	0	1	182.7384
44	Na_2LiReO_6	0.0689	-2.4378	2.0998	0	196.2724
45	Na_2LiSbO_6	0.0604	-2.6087	0	0.32	185.6615
46	Na_2LiWO_6	0.0590	-2.6704	0	1	245.8288
47	Na_2LuSbO_6	0.0716	-2.7760	2.1790	0	197.1082
48	$Na_2LiAlCl_6$	0.1071	-2.0026	4.3836	0	333.5019
49	Na_2PbIO_6	0.0190	-1.4003	0	0.01	264.0148
50	Na_2SrAlO_6	0.0495	-3.1057	0	0.99	188.1454
51	Na_2SrWO_6	0.0817	-2.6955	2.8114	0	210.0640
52	Na_2YOsO_6	0.1174	-2.7037	0.6172	0	186.0850

53	Rb_2YCrO_6	0.0588	-2.9201	0	1	240.2418
54	Sr_2CaCrO_6	0.1076	-2.7039	0.5864	0	191.7340
55	Sr_2CaOsO_6	0.0304	-2.6552	0	2	197.3510
56	Sr_2CaReO_6	0.0175	-2.8774	1.6654	1	271.3090
57	Sr_2CaWO_6	0.0240	-2.9441	3.3256	0	293.9386
58	Sr_2LaBiO_6	0.0909	-2.8449	1.8002	0	232.8416
59	Sr_2LaIO_6	0.0958	-2.5272	2.8215	0	238.5288
60	Sr_2LaOsO_6	0.0195	-2.7273	0	3	216.5066
61	Sr_2LaSbO_6	0.0960	-2.9495	3.6379	0	223.6365
62	Sr_2LaTaO_6	0.0818	-3.3471	3.6806	0	223.7284
63	Sr_2LaWO_6	0.1108	-2.9554	2.6374	0.91	214.7024
64	Sr_2LiAlH_6	0.0389	-0.2312	1.2878	0	219.3971
65	Sr_2LiAlO_6	0.0284	-3.1793	3.4147	0	224.3396
66	Sr_2LuCrO_6	0.0500	-2.8415	0.3637	1	185.8248
67	Sr_2LuReO_6	0.0275	-3.0234	0	2	197.3544
68	Sr_2LuSbO_6	0.0231	-3.1057	3.2007	0	200.4097
69	Sr_2LuTaO_6	0.0205	-3.4885	3.6944	0	199.7342
70	Sr_2MgCrO_6	0.0363	-2.7345	0.3661	0	167.0767
71	Sr_2MgIrO_6	0.0033	-2.4959	0	2.61	177.9777
72	Sr_2MgMoO_6	0.0106	-2.8090	1.5866	0	129.7052
73	Sr_2MgOsF_6	0.0154	-2.8075	0	2	272.8870
74	Sr_2MgOsO_6	0.0110	-2.5261	0	1.99	177.7485
75	Sr_2MgReO_6	0.0142	-2.7664	1.6831	0.98	178.3182
76	Sr_2MgRuF_6	0.0096	-2.7924	0	2	247.8723
77	Sr_2MgRuO_6	0.0132	-2.5165	0	2	221.7697
78	Sr_2MgWO_6	0.0186	-2.8973	3.0590	0	180.6957
79	Sr_2MgZnO_6	0.0165	-2.8075	2.3762	0	202.8798
80	Sr_2NaOsO_6	0.0208	-2.4579	0.0663	1	253.7158
81	Sr_2SmCrO_6	0.0335	-2.8132	0	6	192.3164

82	Sr_2TaBiO_6	0.0430	-2.9384	2.4783	0	209.4533
83	Sr_2TaCrO_6	0.0229	-3.1039	0	3	176.6377
84	Sr_2TaInO_6	0.0209	-2.9940	3.8795	0	192.2006
85	Sr_2TaNbO_6	0.0201	-3.2132	1.2333	0	229.4728
86	Sr_2TaReO_6	0.0176	-2.9210	0	1.97	182.4202
87	Sr_2TaSbO_6	0.0212	-3.0180	2.3901	0	204.7348
88	Sr_2TaTlO_6	0.0419	-2.9036	3.0273	0	242.0947
89	Sr_2UOsO_6	0.0287	-2.7800	0	-	203.1170
90	Sr_2UReO_6	0.0394	-2.9491	0	3	206.2281
91	Sr_2UZnO_6	0.0427	-3.0953	1.7157	0	200.5956
92	Sr_2YAsO_6	0.0586	-2.6479	3.3749	0	187.9900
93	Sr_2YCrO_6	0.1250	-2.7500	0.3655	1	187.4396
94	Sr_2YIrO_6	0.0433	-2.7223	0	2	199.0370
95	Sr_2YNbO_6	0.0521	-3.3244	3.3429	0	205.8638
96	Sr_2YOsO_6	0.0724	-2.7742	0	2.98	201.3224
97	Sr_2YSbO_6	0.0636	-2.9898	3.6141	0	206.4964
98	Sr_2YTaO_6	0.0592	-3.4184	3.7983	0	205.4073
99	Sr_2YUO_6	0.1066	-3.5173	1.4735	1	247.2421
100	Sr_2ZnInO_6	0.0273	-2.4509	0	1	238.7750
101	Sr_2ZnOsO_6	0.0775	-2.2847	0	2	247.2131
102	Sr_2ZnWO_6	0.0804	-2.5807	2.9726	0	248.8888
103	Ti_2AgBiO_6	0.0737	-2.4187	2.1840	0	154.9840
104	Ti_2AgCrO_6	0.0559	-2.4021	0	3	189.4848
105	Ti_2AgIO_6	0.0251	-1.8277	1.9355	0	191.4311
106	Ti_2AgOsO_6	0.0812	-2.0162	0	1	187.2507
107	Ti_2AgSbO_6	0.0903	-2.0542	2.6975	0	185.4760
108	Ti_2CaBiO_6	0.0429	-2.6432	2.0389	0.58	174.0697
109	Ti_2CaOsO_6	0.0350	-2.6602	0.8013	0	143.2558
110	Ti_2LaCrO_6	0.0683	-3.0520	0.0192	2.99	210.7606

111	Ti_2SrUO_6	0.1124	-3.3210	0	2.06	166.8265
112	Ti_2UBiO_6	0.0308	-2.8286	0.3062	1	189.8967
113	Ti_2YBiO_6	0.0930	-2.8358	1.0625	0	178.6732
114	Ti_2YOsO_6	0.0771	-2.8459	0.0358	1	146.7013
$AA'BB'X_6$ compounds						
115	$NaLaBiTeO_6$	0.0325	-2.4001	0	0	203.6228
116	$NaLaIrTeO_6$	0.0802	-2.3650	0	0.16	267.4710
117	$SrLaMgBiO_6$	0.0643	-2.5990	2.6394	0	221.4500
118	$NaLaSbTeO_6$	0.0127	-2.4894	0.2000	1	231.8197
119	$NaScIrTeO_6$	0.0646	-2.3068	0	1.21	196.9658
120	$NaScMgTeO_6$	0.0477	-2.4890	2.5755	0	168.0166
121	$NaTaBiTeO_6$	0.0127	-2.4991	0.1953	0.99	206.3917
122	$NaTaMgTeO_6$	0.0127	-2.6183	3.2346	0	369.8424
123	$NaTlSbTeO_6$	0.0803	-2.0968	0.1243	0.02	214.5647
124	$SrCdWRuO_6$	0.1047	-2.3234	0.6816	0	219.6032
125	$SrFeIrRuO_6$	0.0970	-2.2202	0	3.02	165.2095
126	$SrLaBiSbO_6$	0.0287	-2.4481	1.1895	0.86	241.1619
127	$SrLaBiWO_6$	0.0399	-2.5638	0.6738	0	218.4850
128	$SrLaIrTeO_6$	0.0854	-2.3460	0.7392	0	313.8927
129	$SrLaIrWO_6$	0.0613	-2.6405	0.4640	0	156.0003
130	$SrLaTaWO_6$	0.0523	-3.3072	0.8349	0	236.7150
131	$SrLiBiTeO_6$	0.0365	-2.2757	1.2793	0	169.1703
132	$SrLiWTeO_6$	0.0331	-2.4863	1.3587	0.97	314.7984
133	$SrScIrTeO_6$	0.0436	-2.3979	1.2786	0	197.2712
134	$SrScMgTeO_6$	0.0583	-2.6448	0.7720	0.75	194.4175
135	$SrTbMgTeO_6$	0.0321	-2.6312	0.0345	7	210.4011
136	$TiLaBiTeO_6$	0.0675	-2.4840	1.3341	0	205.9486
137	$TiLiBiTeO_6$	0.0559	-2.3201	2.7195	0	159.4994

Table A3: Fully-relaxed new perovskites from the LCMGM model. Unique and novel compositions are marked in orange background, and are 72 in total. The remainders are polymorphs of previously discovered chemical compositions.

CIF ID	New perovskites	Space group	DFT-relaxed lattice parameters	DFT-relaxed density (g/cm^3)	Model-predicted E_f (eV/atom)
CUBIC					
ABX_3 compounds					
001	$CsMnO_3$	$Pm\bar{3}m$	a = 3.987 Å	6.181	-1.8112
002	$CsZnBr_3$	$Pm\bar{3}m$	a = 5.364 Å	4.714	-1.0317
003	$SrCuI_3$	$Pm\bar{3}m$	a = 4.943 Å	7.313	-0.4971
004	$MnPdS_3$	$Pm\bar{3}m$	a = 4.412 Å	4.979	-0.4425
005	$SrMnS_3$	$Pm\bar{3}m$	a = 5.053 Å	3.073	-1.2563
006	$SrPdI_3$	$Pm\bar{3}m$	a = 5.304 Å	6.395	-0.8230
007	$SrPdS_3$	$Pm\bar{3}m$	a = 4.729 Å	4.558	-1.1177
008	$BeAuI_3$	$Pm\bar{3}m$	a = 5.113 Å	7.287	-0.4204
$A_2BB'X_6$ compounds					
009	Co_2VPbO_6	$Fm\bar{3}m$	a = 7.944 Å	6.254	-1.7546
010	Cs_2LiScF_6	$Fm\bar{3}m$	a = 8.786 Å	4.229	-3.5326
011	Dy_2HfIrO_6	$Fm\bar{3}m$	a = 8.02 Å	10.193	-2.6064
012	Eu_2LaVO_6	$Fm\bar{3}m$	a = 8.343 Å	6.746	-3.2581
013	K_2GePtF_6	$Fm\bar{3}m$	a = 8.751 Å	4.557	-2.2710
014	Lu_2HoTlO_6	$Fm\bar{3}m$	a = 8.469 Å	8.913	-3.4117
015	Nb_2CuWO_6	$Fm\bar{3}m$	a = 7.827 Å	7.331	-1.9957
016	Rb_2LiMnF_6	$Fm\bar{3}m$	a = 8.348 Å	3.960	-3.0653
017	Rb_2AlPdF_6	$Fm\bar{3}m$	a = 8.575 Å	4.406	-2.9696
018	Sr_2VCuO_6	$Fm\bar{3}m$	a = 7.771 Å	5.459	-2.3635
019	Sr_2CWO_6	$Fm\bar{3}m$	a = 7.912 Å	6.264	-2.7890
020	Sr_2VPbO_6	$Fm\bar{3}m$	a = 8.275 Å	6.204	-2.6294
021	U_2CuWO_6	$Fm\bar{3}m$	a = 7.952 Å	10.824	-2.5247
022	Yb_2LaHoO_6	$Fm\bar{3}m$	a = 8.766 Å	7.354	-3.3688

023	Yb_2TlVO_6	$Fm\bar{3}m$	a = 8.054 Å	8.865	-2.7123
MONOCLINIC					
ABX_3 compounds					
024	$AlAsO_3$	$P2_1/m$	a = 5.212 Å, b = 5.274 Å, c = 5.215 Å, β = 92.21°	3.475	-2.1329
025	$NbAlO_3$	$P2_1/m$	a = 3.857 Å, b = 6.247 Å, c = 5.305 Å, β = 104.05°	4.497	-2.6265
026	$TaAlO_3$	$P2_1/m$	a = 3.929 Å, b = 6.456 Å, c = 6.701 Å, β = 91.45°	5.002	-2.9427
027	$TaAuO_3$	$P2_1/m$	a = 3.548 Å, b = 5.890 Å, c = 8.132 Å, β = 91.77°	8.327	-2.5635
028	$LaNbO_3$	$P2_1/m$	a = 3.937 Å, b = 5.983 Å, c = 6.977 Å, β = 106.32°	5.892	-3.1734
029	$TaPtO_3$	$P2_1/m$	a = 3.493 Å, b = 6.486 Å, c = 6.260 Å, β = 105.97°	10.328	-2.3475
030	$RbCrBr_3$	$P2_1/m$	a = 5.328 Å, b = 7.905 Å, c = 8.035 Å, β = 108.78°	3.910	-1.0881
031	$VSbO_3$	$P2_1/m$	a = 3.953 Å, b = 4.971 Å, c = 7.285 Å, β = 103.31°	5.262	-1.9405
032	$AgRhO_3$	Pm	a = 3.832 Å, b = 3.746 Å, c = 5.196 Å, β = 103.94°	5.937	-0.8130
033	$TiSN_3$	Pm	a = 4.806 Å, b = 3.049 Å, c = 4.809 Å, β = 113.07°	3.124	-1.2520
034	VPN_3	Pm	a = 3.660 Å, b = 2.959 Å, c = 5.027 Å, β = 111.01°	4.050	-0.9579
035	VSN_3	Pm	a = 4.922 Å, b = 2.797 Å, c = 4.980 Å, β = 115.64°	3.359	-1.1032
$A_2BB'X_6$ compounds					
036	Nd_2SiWO_6	Cm	a = 9.182 Å, b = 5.277 Å, c = 6.096 Å, β = 96.93°	6.754	-2.7115
037	Sb_2TiCuO_6	$P2_1/c$	a = 5.452 Å, b = 6.046 Å, c = 9.152 Å, β = 125.65°	6.109	-2.4080
038	Y_2TiCrO_6	$P2_1/c$	a = 5.289 Å, b = 5.751 Å, c = 9.147 Å, β = 124.78°	5.431	-3.3044
039	Y_2SrTiO_6	$P2_1/c$	a = 5.519 Å, b = 5.801 Å, c = 9.968 Å, β = 121.54°	4.997	-3.6104

040	Yb_2NdTiO_6	$P2_1/c$	a = 5.473 Å, b = 5.708 Å, c = 9.828 Å, β = 120.99°	8.003	-3.5225
041	Zr_2TiCuO_6	$P2_1/c$	a = 5.038 Å, b = 5.29 Å, c = 10.115 Å, β = 118.39°	5.459	-2.7825
ORTHORHOMBIC					
$\boldsymbol{ABX_3}$ **compounds**					
042	$KPdCl_3$	$Amm2$	a = 7.000 Å, b = 7.022 Å, c = 4.952 Å	3.437	-1.4599
043	$KSrCl_3$	$Amm2$	a = 7.844 Å, b = 7.975 Å, c = 5.542 Å	2.233	-2.4612
044	$KSrF_3$	$Amm2$	a = 6.635 Å, b = 6.926 Å, c = 4.746 Å	2.798	-3.2612
045	$RbGeH_3$	$Amm2$	a = 6.631 Å, b = 6.337 Å, c = 4.005 Å	3.179	-0.2960
046	$RbMgCl_3$	$Amm2$	a = 7.144 Å, b = 7.143 Å, c = 5.036 Å	2.793	-2.1809
047	$RbZnH_3$	$Amm2$	a = 6.628 Å, b = 6.780 Å, c = 3.336 Å	3.409	-0.3451
048	$LiTiH_3$	$Amm2$	a = 4.759 Å, b = 5.094 Å, c = 3.167 Å	2.502	-0.7479
049	$RbTiF_3$	$Amm2$	a = 7.291 Å, b = 7.496 Å, c = 3.471 Å	3.332	-3.0079
050	$GdTlO_3$	$Amm2$	a = 6.503 Å, b = 9.747 Å, c = 3.526 Å	6.087	-2.4146
051	$GdYO_3$	$Amm2$	a = 6.455 Å, b = 9.861 Å, c = 3.508 Å	4.375	-3.7891
$\boldsymbol{A_2BB'X_6}$ **compounds**					
052	Al_2GdTaO_6	$Pmm2$	a = 6.138 Å, b = 7.153 Å, c = 7.997 Å	4.617	-3.3010
053	Al_2PdCO_6	$Imm2$	a = 5.598 Å, b = 6.064 Å, c = 7.457 Å	3.521	-2.4442
054	Al_2ErPdO_6	$Pmm2$	a = 5.872 Å, b = 6.504 Å, c = 6.596 Å	5.585	-2.9223
055	Al_2TaPdO_6	$Imm2$	a = 5.712 Å, b = 6.312 Å, c = 7.261 Å	5.548	-2.5847
056	Ge_2TaSiO_6	$Pmm2$	a = 5.660 Å, b = 5.734 Å, c = 7.333 Å	6.283	-2.4764

057	La_2TaMoO_6	$Pmm2$	a = 5.543 Å, b = 5.976 Å, c = 8.119 Å	8.034	-3.0983
058	Mg_2GdTaO_6	$Imm2$	a = 6.194 Å, b = 6.857 Å, c = 7.513 Å	5.025	-3.1157
059	Sr_2GdTaO_6	$Pmm2$	a = 5.942 Å, b = 6.077 Å, c = 8.427 Å	6.652	-3.4645
060	Sr_2PdCO_6	$Imm2$	a = 5.696 Å, b = 6.327 Å, c = 6.545 Å	5.486	-2.3119
061	Tl_2GdTaO_6	$Imm2$	a = 6.014 Å, b = 6.139 Å, c = 8.407 Å	9.021	-2.7741
062	Tl_2TaMgO_6	$Pmm2$	a = 5.751 Å, b = 6.081 Å, c = 7.866 Å	8.573	-2.6797
063	Tl_2YTaO_6	$Pmm2$	a = 5.979 Å, b = 6.092 Å, c = 8.363 Å	8.446	-3.0054
064	V_2YTaO_6	$Imm2$	a = 5.918 Å, b = 6.246 Å, c = 8.185 Å	5.135	-3.2101
TETRAGONAL					
$\boldsymbol{ABX_3}$ **compounds**					
065	$ErAgS_3$	$P4/mmm$	a = 5.067 Å, c = 5.039 Å	4.767	-1.3723
066	$EuErS_3$	$P4/mmm$	a = 5.243 Å, c = 5.106 Å	4.914	-2.2360
067	$ErTaS_3$	$P4/mmm$	a = 4.889 Å, c = 5.099 Å	6.055	-2.2956
068	$ErTiS_3$	$P4/mmm$	a = 4.973 Å, c = 5.144 Å	4.064	-2.1838
069	$HoAgO_3$	$P4/mmm$	a = 4.453 Å, c = 4.043 Å	6.645	-2.1645
070	$HoAgS_3$	$P4/mmm$	a = 5.034 Å, c = 5.068 Å	4.770	-1.3282
071	$HoTiS_3$	$P4/mmm$	a = 4.944 Å, c = 5.188 Å	4.047	-2.2303
072	$KTaS_3$	$P4/mmm$	a = 4.910 Å, c = 5.897 Å	3.693	-1.5779
073	$TaPS_3$	$P4/mmm$	a = 4.897 Å, c = 4.823 Å	4.423	-1.4148
074	$TiSiO_3$	$P4/mmm$	a = 4.319 Å, c = 3.295 Å	3.348	-2.8424
075	$SmTaO_3$	$P4/mmm$	a = 4.266 Å, c = 4.158 Å	8.322	-3.3318
076	$TbAgS_3$	$P4/mmm$	a = 5.132 Å, c = 5.084 Å	4.500	-1.1152
077	$TbNbO_3$	$P4/mmm$	a = 4.265 Å, c = 4.113 Å	6.655	-2.9112
078	$TbTaO_3$	$P4/mmm$	a = 4.254 Å, c = 4.109 Å	8.660	-3.1196
079	$TbTaS_3$	$P4/mmm$	a = 4.888 Å, c = 5.118 Å	5.923	-1.9500
080	$RbVCl_3$	$P4/mmm$	a = 4.878 Å, c = 4.818 Å	3.517	-1.9067

$A_2BB'X_6$ compounds					
081	Ho_2EuAgO_6	$P4mm$	a = 4.032 Å, c = 9.372 Å	7.472	-2.8802
082	Ho_2EuScO_6	$P4mm$	a = 3.759 Å, c = 10.690 Å	6.847	-3.8027
083	Ho_2RbTaO_6	$P4mm$	a = 3.920 Å, c = 10.546 Å	7.093	-3.3018
084	Ho_2YTlO_6	$P4mm$	a = 3.875 Å, c = 10.470 Å	7.596	-3.3219
085	K_2RbAgI_6	$P4mm$	a = 5.715 Å, c = 12.109 Å	4.338	-0.8990
086	K_2TaVO_6	$P4mm$	a = 3.747 Å, c = 10.437 Å	4.602	-2.7054
087	La_2RbTaO_6	$P4mm$	a = 4.101 Å, c = 9.463 Å	6.678	-3.4602
088	Sr_2GdSbO_6	$P4mm$	a = 4.196 Å, c = 8.922 Å	5.817	-2.8812
089	Sr_2LuUO_6	$P4mm$	a = 4.330 Å, c = 8.605 Å	7.042	-3.6831
090	V_2EuTaO_6	$P4mm$	a = 3.943 Å, c = 10.168 Å	5.575	-2.8295
TRIGONAL					
ABX_3 compounds					
091	$RbHgCl_3$	$P\bar{3}m1$	a = 7.648 Å, c = 8.574 Å	3.001	-1.5440
092	$RbPtCl_3$	$P\bar{3}m1$	a = 7.495 Å, c = 7.168 Å	3.685	-1.4997
093	$LuBiSe_3$	$P\bar{3}m1$	a = 7.124 Å, c = 6.662 Å	7.041	-0.8647
094	$RbBiSe_3$	$P\bar{3}m1$	a = 7.578 Å, c = 7.924 Å	4.477	-0.4820
095	$VBiSe_3$	$P\bar{3}m1$	a = 7.046 Å, c = 6.337 Å	6.056	-0.5337
096	$LuCrSe_3$	$P6_3/mmc$	a = 7.121 Å, c = 5.536 Å	6.336	-1.3592
097	$RbCrSe_3$	$P\bar{3}m1$	a = 7.436 Å, c = 7.413 Å	3.502	-0.6842
098	$LaLuSe_3$	$P\bar{3}m1$	a = 6.946 Å, c = 7.555 Å	5.794	-1.8013
099	$LaVSe_3$	$P\bar{3}m1$	a = 6.676 Å, c = 7.111 Å	5.163	-1.2236
100	$PtPbO_3$	$P\bar{3}m1$	a = 4.563 Å, c = 8.906 Å	9.311	-1.2505
101	$PtPbSe_3$	$P\bar{3}m1$	a = 5.619 Å, c = 9.014 Å	8.613	-0.0579
102	$RhPbSe_3$	$P\bar{3}m1$	a = 6.573 Å, c = 6.662 Å	7.288	-0.1999
103	$SiPbSe_3$	$P\bar{3}m1$	a = 7.202 Å, c = 6.553 Å	5.326	-0.3167
104	$VPbSe_3$	$P\bar{3}m1$	a = 7.376 Å, c = 5.708 Å	6.114	-0.7150
105	$ZnPbSe_3$	$P\bar{3}m1$	a = 6.636 Å, c = 7.200 Å	6.162	-0.3210
106	$ScAuSe_3$	$P\bar{3}m1$	a = 5.911 Å, c = 7.153 Å	7.348	-0.5236
107	$LuScSe_3$	$P\bar{3}m1$	a = 6.817 Å, c = 6.913 Å	5.453	-1.7275

108	$LuUSe_3$	$P\bar{3}m1$	a = 6.681 Å, c = 7.214 Å	7.740	-1.1209
109	$URhSe_3$	$P\bar{3}m1$	a = 6.195 Å, c = 6.469 Å	8.925	-0.6334
$A_2BB'X_6$ compounds					
110	Ag_2AsClF_6	$P\bar{3}m1$	a = 5.026 Å, c = 7.020 Å	4.758	-1.6225
111	Ag_2RbHoF_6	$P\bar{3}m1$	a = 6.854 Å, c = 7.250 Å	3.266	-2.7717
112	Ag_2TcClF_6	$P\bar{3}m1$	a = 5.253 Å, c = 7.035 Å	4.576	-2.2331
113	Cd_2LuClF_6	$P\bar{3}m1$	a = 7.270 Å, c = 5.555 Å	3.583	-2.5610
114	Cd_2RbAuF_6	$P\bar{3}m1$	a = 7.137 Å, c = 6.217 Å	3.761	-2.0743
115	K_2HoAuF_6	$P\bar{3}m1$	a = 6.121 Å, c = 6.203 Å	4.571	-3.1062
116	K_2LiHoF_6	$P\bar{3}m1$	a = 6.003 Å, c = 5.534 Å	3.501	-3.5136
117	K_2RbHoF_6	$P\bar{3}m1$	a = 6.511 Å, c = 6.629 Å	3.020	-3.3808
118	$Rb_2LiAuBr_6$	$P\bar{3}m1$	a = 7.421 Å, c = 6.511 Å	4.568	-0.9689
119	Rb_2LiDyF_6	$P\bar{3}m1$	a = 6.150 Å, c = 5.592 Å	4.119	-3.4323
120	$Rb_2LiNiBr_6$	$P\bar{3}m1$	a = 7.300 Å, c = 6.366 Å	4.047	-1.1682
121	Rb_2AuLuF_6	$P\bar{3}m1$	a = 6.281 Å, c = 6.223 Å	5.129	-2.9630
122	Rb_2DyNiF_6	$P\bar{3}m1$	a = 6.206 Å, c = 5.755 Å	4.378	-2.9235
123	Rb_2TcClF_6	$P\bar{3}m1$	a = 5.705 Å, c = 7.136 Å	3.454	-2.8511
124	Rb_2CaTiF_6	$P\bar{3}m1$	a = 6.295 Å, c = 5.893 Å	3.062	-3.2943