European Journal of Applied Sciences – Vol. 13, No. 02
Publication Date: April 25, 2025
DOI:10.14738/aivp.1302.18468.
Yuan, X., & Li, Q. (2025). Applications of Diffusion Models in Material Properties Prediction. European Journal of Applied Sciences, Vol - 13(02). 375-389.

Applications of Diffusion Models in Material Properties Prediction

Xin Yuan
School of Science, Chongqing University of Posts and Telecommunications, Chongqing, 400065, People's Republic of China

Qingxu Li*
School of Science, Chongqing University of Posts and Telecommunications, Chongqing, 400065, People's Republic of China

ABSTRACT

Diffusion models, as probabilistic generative models, have shown significant potential applications in materials science in recent years [1,2]. By gradually including and removing noise from data, diffusion models can generate material structures with targeted physical and/or chemical properties. We systematically summarizes the research progress of diffusion models in materials science, focusing on the application of denoising diffusion probabilistic models (DDPM), score-based generative models (SGM), and diffusion models based on stochastic differential equations (SDE) in material generation. Furthermore, we summarize successful applications of crystal diffusion variational autoencoders (CDVAE) and their variants in material generation, and discusses the use of diffusion language models for generating inorganic compounds. Despite the robust generative capabilities and stability of diffusion models in generative tasks, challenges remain in terms of computational cost, diversity of generated samples, and scalability. We conclude by analyzing these key challenges and exploring future directions for the development of diffusion models in the field of materials science.

Keywords: diffusion models, generative models, material sciences, crystal diffusion variational autoencoders (CDVAE)

INTRODUCTION

Materials science, as an interdisciplinary field, encompasses physics, chemistry, computer sciences, and other disciplines, focusing on the study of the structure, properties, performance, and interrelations of materials. With rapid technological advancements, materials science plays a crucial role in various fields such as energy, environment, electronics, and aerospace. To meet the growing demand for new materials in modern society, scientists continuously explore various methods to accelerate the discovery and optimization of materials. Although traditional experimental methods can provide precise data, they are often costly, time-consuming, and inefficient. Therefore, leveraging computational methods to assist in material research and development has become a significant topic in the field of materials science.

Diffusion models are a class of generative models based on probability theory and statistical physics, and have achieved significant progress in the field of machine learning in recent years[1-8]. These models simulate the process of transforming data from a disordered to an

ordered state to gradually generate target data. Initially, noise is added to data to make it disordered, and then through learning the reverse process, the disordered data is incrementally restored to an ordered state, achieving data generation. Due to their theoretical advantages and practical efficiency, diffusion models have become a research hotspot in the field of generative models.

Diffusion models have demonstrated strong capabilities across various domains, including image generation, natural language processing, time series prediction, and multimodal generation. They are particularly notable in image generation tasks, capable of producing high-quality image samples and being used in tasks such as image super-resolution, inpainting, and synthesis in computer vision. In natural language processing, diffusion models are utilized for text generation, completion, and translation, effectively handling discrete text data by mapping it into continuous representation spaces.

Additionally, diffusion models excel in time series prediction, generating highly accurate predictive results by learning the distribution of time series data, applicable in fields like finance and weather forecasting. In multimodal generation tasks, such as generating images or videos from text, diffusion models also exhibit extensive application potential. With ongoing research, diffusion models have not only improved the quality of generated data but also expanded their application scope, suggesting a promising future for their application in more fields.

The application of diffusion models in materials science is still in its early stages, but existing studies indicate their significant potential for material design and optimization. Through diffusion models, researchers can simulate the microstructures of materials, predict their physical and chemical properties, and provide theoretical foundations for the discovery of new materials. For instance, diffusion models have been successfully applied to the generation and prediction of crystalline materials, offering new perspectives and methods for research in materials science. However, despite some progress, the application of diffusion models in materials science still faces numerous challenges, such as high computational complexity and limited ability to handle large-scale data.

We aims to systematically summarize and analyze the research progress and current applications of diffusion models in materials science, exploring their specific applications in material design, performance prediction, and microstructural simulation. Additionally, the paper will analyze the main challenges facing the application of diffusion models in materials science and envisage future research directions. The structure of this paper is arranged as follows: first, it introduces the basic principles of diffusion models and their significance in materials science; next, it discusses in detail the applications of diffusion models within the field; subsequently, it analyzes the challenges currently faced in research involving diffusion models and outlines potential future research directions.

THE BASIC PRINCIPLES OF DIFFUSION MODELS

Generative Diffusion Models are a class of generative models based on stochastic processes, primarily used for generating complex data from simple distributions. The core idea involves incrementally injecting noise into the data and then learning the inverse process of restoring

the data from the noise. Diffusion models have demonstrated strong capabilities in image generation, natural language processing, and other complex data generation tasks. The workflow of generative diffusion models can be divided into two stages: forward diffusion and reverse denoising. During the forward diffusion stage, the model progressively adds noise to degrade the original data, eventually transforming it into Gaussian noise. This process is typically implemented through a series of Markov chains. In the reverse diffusion stage, the model starts from Gaussian noise and generates new data samples by gradually denoising it, as illustrated in Figure 1.

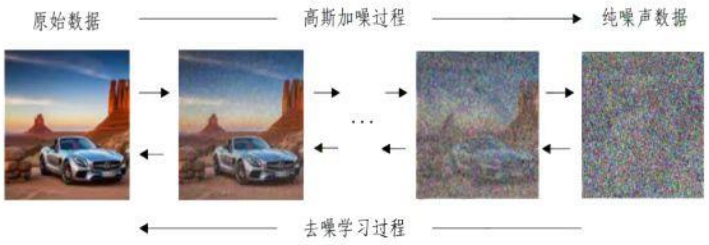

Figure 1: Workflow of the Diffusion Model, from [2].

Denoising Diffusion Probabilistic Models (DDPMs) [6,7] are a type of generative model that generate new data by progressively injecting noise into data and learning the reverse denoising process. This method consists of two key processes: the forward diffusion process and the reverse denoising process. In the forward diffusion process, the model gradually adds noise to the data, transforming its distribution into a simple prior distribution (such as a standard Gaussian distribution). Given an initial data distribution$x_0 \sim q(x_0)$, a series of random variables $x_0, x_1, \cdots, x_n$ are generated via a Markov chain [8], with the transition kernel at each step defined as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

Where β_t is a predetermined hyperparameter representing the intensity of noise added at each step. Utilizing the chain rule and the Markov property, the joint distribution can be factorized as:

$$q(x_1, \cdots x_T|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{2}$$

To simplify the computation, the variables $\alpha_t = 1-\beta_t$ and $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$ are introduced, and for any time step t, they can be expressed as:

$$q(x_t|x_0) = \mathbf{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I\right) \tag{3}$$

sampling method to directly compute x_t from the initial data x_0 is:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon \tag{4}$$

In this context $\varepsilon \sim \mathbf{N}(0, I)$. In the reverse denoising process, the model begins from a simple prior distribution $p(x_T) = \mathbf{N}(x_T; 0, I)$ and generates data samples by progressively denoising through a learnable Markov chain. The transition kernel of the reverse Markov chain is given by:

$$p_\theta(x_{t-1} \mid x_t) = \mathbf{N}\left(x_{t-1}; \mu_\theta(x_t, t), \sum\nolimits_\theta (x_t, t)\right) \tag{5}$$

Where μ_θ and Σ_θ are the mean and variance parameterized by a deep neural network. The training objective of the model is to align the joint distribution of the forward process $q(x_0, \cdots, x_T)$ with that of the reverse process $p_\theta(x_0, \cdots, x_T)$ by minimizing the Kullback-Leibler (KL) divergence between them to optimize the model parametersθ. The KL divergence can be decomposed as follows:

$$KL\left(q(x_0, \cdots, x_T) \,\|\, p_\theta(x_0, \cdots, x_T)\right) = \mathbf{E}_q[-\log p_\theta + \lg q] \tag{6}$$

The expectation term of the loss function is further simplified to the Variational Lower Bound (VLB), which is given by:

$$L_{VLB}(x_0) = \mathbf{E}_q\left[-\log p_\theta - \sum_{t=1}^{T} \log \frac{p_\theta(x_{t-1} \mid x_t)}{q(x_t \mid x_{t-1})}\right] \tag{7}$$

Ho et al. (2020) [6] reweighted the terms in the Variational Lower Bound (VLB) to improve generation quality. The corresponding form of the loss function is:

$$E_{t \sim U[1,T], x_0 \sim q(x_0), \varepsilon \sim \mathbf{N}(0,I)}[\boldsymbol{\lambda}(t) \,\|\, \varepsilon - \varepsilon_\theta(x_t, t)\,\|^2] \tag{8}$$

Here, $\boldsymbol{\lambda}(t)$ is a weighting function, and ε_θ is the noise vector predicted by the model. By applying the above process to specific tasks, such as image generation, DDPMs can generate high-quality image samples. During the training phase, the model learns the denoising process step by step, and ultimately, in the generation phase, it can restore clear images from noise.

Score-based Generative Models (SGMs) [9] are a type of generative models that generate samples by estimating the score function of the data distribution. The score function refers to the gradient of the log-density of the data distribution, indicating the direction in which the data density function increases the fastest. The key idea of SGMs is to perturb the data with a series of progressively stronger Gaussian noise and train a deep neural network model conditioned on the noise level, known as a Noise-Conditional Score Network (NCSN [9]), to jointly estimate the score functions of all noisy data distributions. In practice, the estimation of the score function and sample generation is achieved through the following steps. First, given

a data distribution $q(x)$ and a series of noise levels $0 < \sigma_1 < \sigma_2 < \cdots < \sigma_T$, perturb the data points from x_0 to x with Gaussian noise $q(x_t|x_0) = \mathbf{N}(x_t; x_0, \sigma_t^2 I)$. Then, train a neural network $s_\theta(x,t)$ to estimate the score function $\nabla_x \log q(x_i)$ of the noisy data. If denoising score matching is used, the training objective is as follows:

$$\mathbf{E}_{t\sim U[1,T], x_0\sim q(x_0), x_t\sim q(x_t|x_0)}[\boldsymbol{\lambda}(t)\sigma_t^2 \| \nabla_{x_t} \log q(x_t) - s_\theta(x_t,t)\|^2] \tag{10}$$

$\boldsymbol{\lambda}(t)$ is a positive weighting function used to control the weights at different noise levels. The noise level, denoted as σ_t, increases gradually with time t. By substituting this $q(x_t \mid x_0) = N(x_t; x_0, \sigma_t^2 I)$ into $\varepsilon_\theta(x,t) = -\sigma_t s_\theta(x,t)$, the loss function expression can be further simplified as follows:

$$E_{t\sim U[1,T], x_0\sim q(x_0), \varepsilon\sim \mathbf{N}(0,I)}[\boldsymbol{\lambda}(t)\| \varepsilon + s_\theta(x_t,t)\|^2] + \text{const} \tag{11}$$

Comparing the training objectives of denoising score matching and Denoising Diffusion Probabilistic Models (DDPMs) (Equation 9), it can be seen that when we set $\varepsilon_\theta(x,t) = -\sigma_t s_\theta(x,t)$, the training objectives of these two models are equivalent. Finally, we can use Annealed Langevin Dynamics (ALD) [9] to generate data by applying Langevin Monte Carlo starting from an initial noise that follows a Gaussian distribution. The sampling process reduces the noise level by iteratively applying the score function. The process is illustrated in Figure 2.

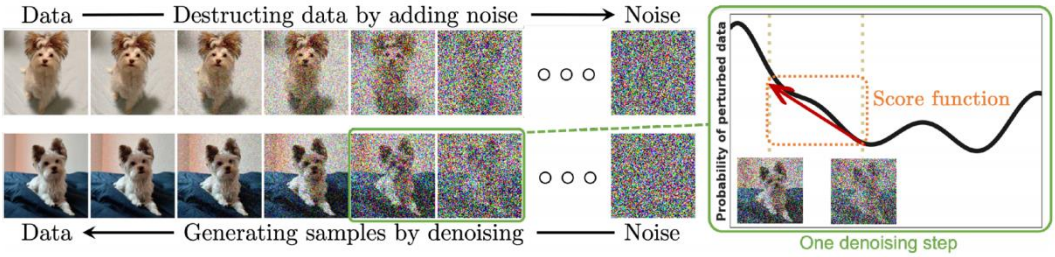

Figure 2: Illustrates the process flow of SGMs, sourced from [1]

The fundamental idea of generative diffusion models based on SDEs is to model both the data perturbation process and the generation process using SDEs. In the forward diffusion process, the data perturbation can be represented by the following SDE:

$$dx = f(x,t)dt + g(t)dw \tag{12}$$

Here, $f(x,t)$ is the drift function, describing the deterministic changes of data over time, $g(t)$ is the diffusion coefficient, describing the random changes of data over time, and dw is the standard Brownian motion. Stochastic differential equations (SDEs) further generalize the denoising diffusion probabilistic models (DDPMs) and score-based generative models (SGMs) to the case of infinite time steps or noise levels. At this point, both perturbation and denoising

processes can be described by stochastic differential equations. For DDPMs, the corresponding SDE is:

$$dx = -\frac{1}{2}\beta(t)xdt + \sqrt{\beta(t)}dw \tag{13}$$

For SGMs, the corresponding SDE is:

$$dx = \sqrt{\frac{d\left[\sigma(t)^2\right]}{dt}}dw \tag{14}$$

Anderson (1982) [11] proved that for any form of diffusion process (Equation 12), its reverse process can be realized by solving the following reverse-time SDE:

$$dx = -[f(x,t) - g^2(t)\nabla_x \log q_t(x)]dt + g(t)dw \tag{15}$$

where $\bar{w}$ is the standard Brownian process flowing in reverse time, and dt represents an infinitesimally small negative time step. Song et al. (2020) [10] demonstrated that there exists an ordinary differential equation (ODE), namely the probability flow ODE, whose trajectories have the same marginal distribution as the reverse-time SDE. The probability flow ODE is:

$$dx = [f(x,t) - \frac{1}{2}g^2(t)\nabla_x \log q_t(x)]dt \tag{16}$$

Once the score function $s_\theta(x_t, t)$ at each time step is known, samples can be generated by solving the reverse-time SDE (or probability flow ODE). We estimate the score function by parameterizing a time-dependent score model $s_\theta(x_t, t)$.

APPLICATIONS OF DIFFUSION MODELS IN MATERIALS SCIENCE

The Crystal Diffusion Variational Autoencoder (CDVAE) [4] is a novel generative model combining Variational Autoencoders (VAE) and diffusion models. This model generates materials that comply with physical stability biases by learning the data distribution of stable materials. In addition to its core functionality, it contributes by creating standardized datasets and evaluation metrics. The generation of new materials using CDVAE can be divided into two steps: first, sampling the material composition, lattice, and number of atoms from the latent space; then, gradually denoising through Langevin dynamics to generate the final structure. The CDVAE network architecture is primarily composed of three components: a Periodic Graph Neural Network Encoder, an Attribute Predictor, and a Periodic Graph Neural Network Decoder.

The encoder uses a SE (3)-equivariant periodic graph neural network (PGNN) [12][13][14][15] to map the input atomic coordinates and types from a high-dimensional space to a low-dimensional latent representation space. The PGNN network is capable of handling periodic

boundary conditions of atoms and maintaining structural symmetry. This allows the encoder to learn the spatial characteristics of crystal structures, enabling the generation of new materials with periodicity and symmetry.

The role of the attribute predictor (MLP_{AGG}) is to predict key aggregate properties of materials from their latent representation z: the number and proportion of different types of atoms in the material, the periodic structure of the material, and the total number of atoms in the material.

In the CDVAE model, the decoder plays a crucial role in generating actual material structures from the latent representation. It receives the low-dimensional latent representation produced by the encoder and noisy atomic types $\tilde{A}$, atomic coordinates $\tilde{X}$, and lattice L. It then uses the Noise Conditional Score Network (NCSN) to learn the score function, which, through Langevin dynamics, iteratively reduces noise, causing the generated structure to gradually approach the real crystal structure.

The CDVAE model directly processes atomic coordinates and learns a continuous representation of the structure, eliminating the need for intermediate representations such as graphs or descriptors. It fully leverages the advantages of combining VAE with diffusion models, providing a powerful tool for the generation of new materials.

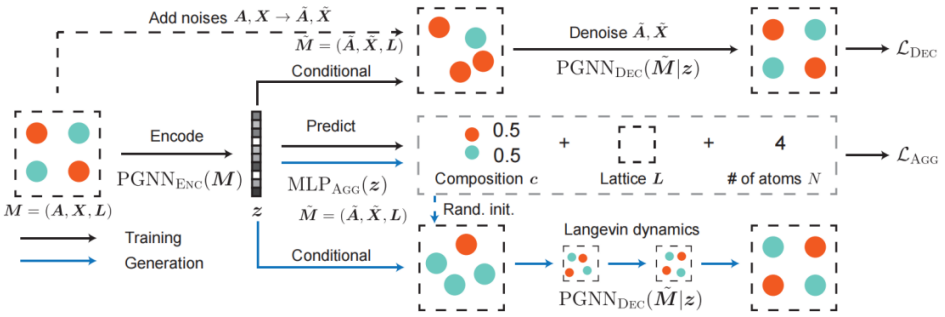

Figure 3: provides an overview of the CDVAE method, sourced from [4].

The study by Chen et al. (2024) [5] is a specific application of the CDVAE model in generating BCP (Boron-Carbon-Phosphorus) materials. The researchers used the CDVAE model to generate a vast array of new BCP structures, which were then screened and optimized. By training the CDVAE model, they managed to generate 10,000 new BCP structures. The generated structures first underwent preliminary screening through validity and diversity checks to ensure they met predefined validity standards and exhibited sufficient diversity. These checks included verification of structural geometric reasonableness and chemical composition. For the screened candidate structures, the researchers employed a pretrained materials graph neural network (M3GNet) [20] model to perform structural relaxation. The M3GNet model can quickly predict the formation energies, forces, and stresses of the structures while progressively updating the properties of bonds, atoms, and states. After M3GNet relaxation, structures that met BCP bonding specifications were subjected to further DFT calculations to evaluate their thermodynamic and kinetic stability. The experimental results

demonstrated that the BCP structures generated by CDVAE showed significant advantages in terms of binding energy and chemical diversity. The generated structures exhibited a binding energy distribution superior to that of the original data and showed greater diversity in t-SNE analysis. These findings indicate that the application of the CDVAE model in BCP material generation highlights its powerful ability to generate high-quality new material structures, providing an effective approach for the discovery and exploration of new materials.

The Conditional Crystal Diffusion Variational Autoencoder (Con-CDVAE) [17] is a generative model based on diffusion models and the variational autoencoder (VAE) architecture. It aims to enable the design of crystal structures using conditional generation techniques. This model extends the classical CDVAE model by adding the ability to control specific attributes of the generated crystal structures, such as band gap and formation energy. The key components of the model include the property embedding module, the predictor block, and the prior block. These components work together to generate crystals with specific properties based on user-defined conditions. The generation strategy of the model mainly includes three types: default strategy, full-conditioned strategy, and few-conditioned strategy. The default strategy uses only a single attribute for generation, the full-conditioned strategy utilizes all available attributes for crystal generation, and the few-conditioned strategy generates crystals by randomly sampling missing data when some attributes are absent.

The training process of Con-CDVAE is divided into two steps: Step-one and Step-two, aimed at building the model's encoder and decoder. First, the Encoder is responsible for receiving the input crystal structure features, including atomic types (A), atomic coordinates (X), and lattice vectors (L), and converting this information into a low-dimensional latent variable representation. Simultaneously, the PropEmb (property embedding) module embeds input target attributes, such as band gap and formation energy, into the attribute space to provide prior information for conditional generation. The Predictor block is used to adjust the generated latent variables, ensuring they are close to the target attributes in attribute space, so that the generated results align with the specified conditions. Subsequently, the Decoder takes the latent variables and embedded attribute results together to generate crystal structures that meet these conditions.

In the Step-one phase, the focus is on training the encoder, decoder, and predictor block by optimizing two loss functions: one is the CDVAE loss for model reconstruction, and the other is the predictor loss to ensure the generated results match the target attributes. The Step-two phase further optimizes the model by introducing the Prior module. PriorEmb (prior embedding) embeds target attributes into the latent space for generating condition-compliant latent variables, while PriorDecode (prior decoder) uses these embedded attributes to generate latent variables, optimizing the model's conditional generation ability by minimizing the prior loss.

During the model's generation phase, the trained Prior and PropEmb modules are used for crystal generation. PriorEmb and PriorDecode generate new latent variables, sampling new samples from the latent space based on given conditions. The Decoder then receives the latent variables generated by PriorDecode, decodes them, and generates crystal structures that meet the conditions, thus achieving the goal of conditional crystal generation.

To evaluate the performance of Con-CDVAE in conditional generation, the researchers designed a series of experiments. The generation conditions used in the experiments included band gap values (0 eV, 3 eV, 6 eV), formation energies (0 eV/atom, -2 eV/atom, -4 eV/atom), and different crystal systems (such as cubic and hexagonal systems). The experimental setup employed three generation strategies: default strategy, full-conditioned strategy, and few-conditioned strategy, aimed at testing the model's generation capability with single attribute conditions as well as combined attribute conditions.

The experimental parameters involved different datasets (MP20, MP40, OQMD), each filtered to ensure the inclusion of representative crystal structures and attribute distributions. During the experiments, the model received specified generation condition inputs and used the diffusion process to progressively generate crystal structures that met these conditions. Tools such as pymatgen [19] and CGCNN [18] were used to analyze and evaluate the generated results.

As illustrated in Figure 4, the general workflow of Con-CDVAE is depicted, showing the process through which the model generates condition-compliant crystal structures.

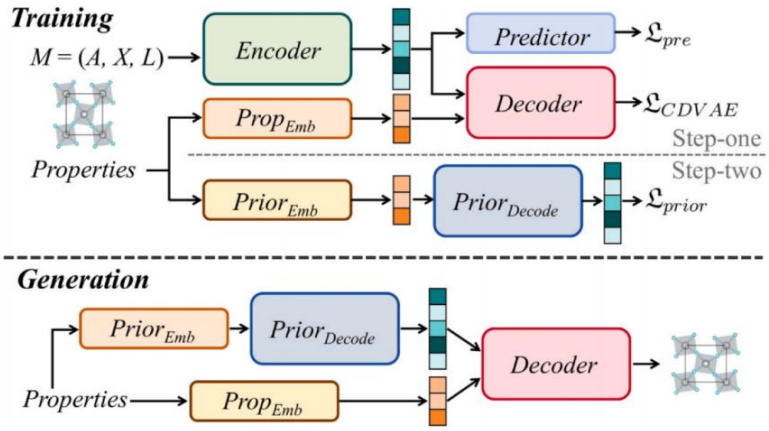

Figure 4: General Workflow of Con-CDVAE [17]

RESULTS OF COMBINED CONDITIONAL GENERATION EXPERIMENTS

In the combined conditional generation experiments, the Con-CDVAE model demonstrated its capability to generate crystal structures based on multiple attribute conditions. The experimental results indicated that the model could produce crystal materials meeting requirements under various attribute conditions, such as band gap and formation energy. Particularly under the full-conditioned strategy, the generated crystal structures exhibited more accurate distribution in the designated attributes, outperforming other strategies. However, the task of combined conditional generation posed greater challenges compared to single-attribute generation, leading to decreased precision in matching the band gap and formation energy. This was especially evident when attempting to simultaneously generate crystals with high band gaps and low formation energies. Overall, Con-CDVAE showed promising potential in controlling crystal attributes and generating structures based on multiple conditions, offering an effective tool for future material design and discovery.

The Vector Field Oriented Diffusion Model for Crystal Material Generation [21] enables the global generation of crystal geometric structures, providing an effective method for new material design. The core concept of this model views the generation process of crystal materials as a diffusion process, executing stepwise diffusion in the geometric space of the crystal, including atomic positions and lattice units, and learning the reverse process to generate targeted crystal materials. Compared to traditional generative models, this approach not only handles the positions of atoms but also generates plausible lattice unit structures. The generation of lattice units is crucial, as the geometric structure of the lattice determines the material's periodicity and density, which directly affect the physical properties of the crystal, such as the band gap, formation energy, and elastic modulus. By introducing a diffusion process under periodic conditions in three-dimensional space, the model ensures that the generated lattice units and atomic positions maintain physical consistency, while also allowing the crystal structure to satisfy specific chemical and physical conditions.

To maintain the symmetry and periodicity of crystal structures during the generation process, the model employs Geometrically Equivariant Graph Neural Networks (GNNs). This type of network effectively captures the geometric features of crystal materials during the diffusion process and preserves the equivariance of the generated results. Specifically, the model utilizes a toroidal structure in the atomic position space, while the diffusion process for generating lattice units is based on a vector field. This approach allows the model to generate crystal structures that adhere to geometric constraints, while also ensuring the physical consistency of the crystal materials throughout the generation process.

Experimental results indicate that the vector field-oriented diffusion model performs exceptionally well in existing crystal generation benchmark tests. In particular, the model demonstrates significant advantages in generating materials with high density and reasonable lattice structures. Compared to traditional generative models based on convolutional neural networks (CNNs) or generative adversarial networks (GANs), the diffusion model shows substantial improvement in the validity of the generated results, as well as enhanced consistency in physical properties such as crystal structure density, formation energy, and band gap. Furthermore, the model effectively avoids the common mode collapse issues encountered in generative adversarial networks during the generation process, thereby enabling more stable production of crystal materials that meet the expected properties.

Dong et al. [25] proposed a method for generating inorganic compounds using deep diffusion language models, aimed at addressing the challenges of component generation and structure prediction in novel material design. Due to the vast complexity of chemical space, the discovery of materials faces numerous challenges, such as the need to simultaneously satisfy requirements like charge neutrality, balanced electronegativity, synthesizability, and mechanical stability. To tackle these challenges, the researchers designed a comprehensive workflow for generating materials: first, chemical components are generated using a diffusion language model; next, a template-based crystal structure prediction algorithm is applied to generate the corresponding structures; finally, a latent model based on graph neural networks is used to optimize the structure through relaxation.

In the component generation stage, the researchers trained two types of diffusion language models: Diffusion-LM and Diffusion-BERT, and compared their performance. The generated chemical formulas underwent rigorous validation for chemical viability, including checks for charge neutrality, electronegativity balance, and oxidation states. Through these diffusion language models, the researchers successfully generated several new materials with chemical stability.

Diffusion-LM is a non-autoregressive language model, meaning it can generate sequences without strictly following a left-to-right order. It generates word vectors by iteratively denoising a series of Gaussian vectors, producing continuous intermediate latent variables throughout the generation process, which are continuously updated to meet the requirements of complex generation tasks. On the other hand, Diffusion-BERT is a diffusion model based on BERT, which generates chemical components through a diffusion process. The masked language model mechanism of BERT is utilized to generate the symbol sequences within chemical formulas, with the model performing multiple rounds of denoising to produce the final chemical components. In the structure generation process, the researchers employed the CSPML [26] algorithm for template matching, which generates crystal structures consistent with the new chemical components through element substitution. The generated structures are then relaxed using the M3GNet [20] model based on graph neural networks, and their thermodynamic stability is further validated using density functional theory (DFT).

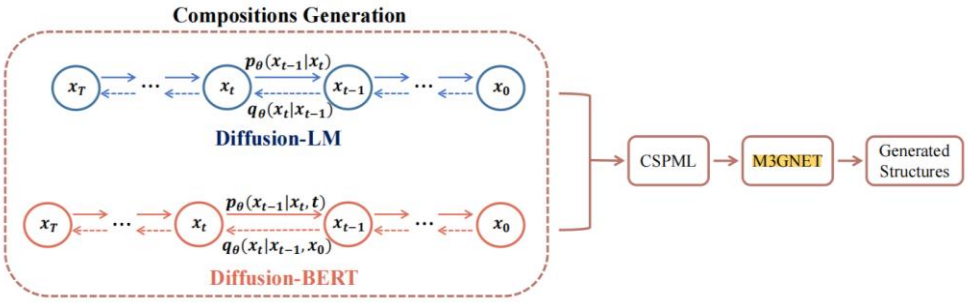

Figure 4: Framework for Generating Inorganic Compounds Based on Diffusion Language Models, from [25].

Through this innovative generative model, the researchers discovered six new materials, including Ti_2HfO_5, TaNbP, $YMoN_2$, and $TaReO_4$, all of which exhibit negative formation energies, indicating their thermodynamic stability. Notably, the energies of four of these materials are positioned above the convex hull (e-above-hull values) and are less than 0.3 eV, further demonstrating the potential of this method in the field of material generation. The findings indicate that the generative approach based on diffusion language models can effectively explore new chemical spaces, providing a novel technological pathway for the discovery of new materials.

PROGRESS AND CHALLENGES OF DIFFUSION MODELS

Diffusion models, as a novel approach within generative modeling, have seen extensive application in recent years in areas such as image generation, natural language processing, and the design of molecular and crystal structures in materials science. These models generate

high-quality data through a stepwise diffusion and denoising process, which mitigates some of the instability issues associated with the training of traditional Generative Adversarial Networks (GANs) and addresses the limitations of Variational Autoencoders (VAEs) in producing lower-quality results. However, despite the strong generative capabilities and stability exhibited by diffusion models, several limitations and challenges remain in their practical applications.

Firstly, the high computational cost is a significant bottleneck for diffusion models. These models typically require hundreds to thousands of diffusion and denoising steps to progressively recover target samples from noise, resulting in longer inference times and higher resource consumption. This issue becomes particularly pronounced when dealing with large-scale datasets or generating complex structures, such as crystal materials or molecular configurations. Although several improvements have been proposed — such as reducing the number of diffusion steps or optimizing the denoising process to enhance efficiency — further research is needed to reduce computational overhead while ensuring the quality of the generated results.

Secondly, the diversity and quality of generated samples continue to present challenges. While diffusion models demonstrate good stability in their outputs, the diversity of the generated samples may be insufficient in practical applications, especially for high-dimensional complex data, such as the generation of inorganic compounds. For example, in the research on inorganic compound generation conducted by Dong et al., although a variety of new materials were successfully generated, additional screening mechanisms were required during the component generation stage to validate the chemical viability of the generated results. This indicates that enhancing the diversity and precision of generated results in chemical generation processes remains an area that requires further optimization.

Moreover, the scalability of diffusion models represents another challenge when addressing complex tasks. While significant progress has been made in image generation and small-scale structure generation, their performance is still limited when confronted with large-scale, high-complexity tasks. For instance, in applications such as crystal structure prediction or large molecule generation, a key focus of current research is to enhance model flexibility while maintaining the physical validity of the generated results.

To address these issues, researchers have proposed various improvement strategies, including reducing the number of diffusion steps or adopting jump diffusion techniques to lower computational costs. Additionally, combining diffusion models with other generative models (such as GANs or VAEs) can enhance sample diversity and quality. Furthermore, integrating physical constraints and chemical rules into the generation process provides additional opportunities for the application of diffusion models in materials science and molecular generation. However, it remains essential to ensure the accuracy and reliability of generated results while pursuing efficiency.

Despite the significant potential of diffusion models in generative tasks, challenges related to computational costs, sample diversity, and scalability are core issues that require further research and resolution in the future. Optimizing the generation process and enhancing the

model's generalizability and adaptability will be key directions for the future development of diffusion models.

References

[1] Xiaoyang Yang, Jian Li, and Yifan Wang. Diffusion Models: A Comprehensive Survey of Methods and Applications. Advances in Neural Information Processing Systems, 37:1123-1134, 2024.

[2] Zhihao Yan, Changbing Zhou, and Xiaocui Li. Generative Diffusion Model Research Overview. Computer Science, Online First: September 26, 2023.

[3] Astrid Klipfel, Yaël Fregier, Adlane Sayede, and Zied Bouraoui. Vector Field Oriented Diffusion Model for Crystal Material Generation. The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24), 1234-1245, 2024.

[4] Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal Diffusion Variational Autoencoder for Periodic Material Generation. Published as a conference paper at ICLR 2022.

[5] Chen Chen, Jinzhou Zheng, Chaoqin Chu, Qinkun Xiao, Chaozheng He, and Xi Fu. An Effective Method for Generating Crystal Structures Based on the Variational Autoencoder and the Diffusion Model. Chinese Chemical Letters, 2024.

[6] Jonathan Ho, Ajay Jain, and Pieter Abbeel.Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 33:6840–6851, 2020.

[7] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli.Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. International Conference on Machine Learning, 32:2256–2265, 2015.

[8] Daphne Koller and Nir Friedman. Probabilistic Graphical Models: Principles and Techniques. MIT Press, 2009.

[9] Yang Song and Stefano Ermon.Generative Modeling by Estimating Gradients of the Data Distribution. Advances in Neural Information Processing Systems, 32:1-10, 2019.

[10] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole.Score-Based Generative Modeling Through Stochastic Differential Equations. International Conference on Learning Representations, 2020.

[11] Brian D. O. Anderson.Reverse-Time Diffusion Equation Models.Stochastic Processes and Their Applications, 12(3):313–326, 1982.

[12] Johannes Klicpera, Janek Groß, and Stephan Günnemann.Directional Message Passing for Molecular Graphs.8th International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020.

[13] Johannes Klicpera, Florian Becker, and Stephan Günnemann. GemNet: Universal Directional Graph Neural Networks for Molecules. arXiv preprint arXiv:2106.08903, 2021.

[14] Muhammed Shuaibi, Adeesh Kolluru, Abhishek Das, Aditya Grover, Anuroop Sriram, Zachary Ulissi, and C. Lawrence Zitnick. Rotation Invariant Graph Neural Networks Using Spin Convolutions.arXiv preprint arXiv:2106.09575, 2021.

[15] Jonathan Godwin, Michael Schaarschmidt, Alexander Gaunt, Alvaro Sanchez-Gonzalez, Yulia Rubanova, Petar Veličković, James Kirkpatrick, and Peter Battaglia. Very Deep Graph Neural Networks via Noise Regularization. arXiv preprint arXiv:2106.07971, 2021.

[16] C. Chen, S.P. Ong. Nat. Comput. Sci. 2, 718-728, 2022.

[17] Cai-Yuan Ye, Hong-Ming Weng, and Quan-Sheng Wu. Con-CDVAE: A Method for the Conditional Generation of Crystal Structures. Computational Materials Today, 1:100003, 2024.

[18] Tian Xie and Jeffrey C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. Physical Review Letters, 120:145301, 2018.

[19] C. Saharia, W. Chan, S. Saxena, L. Li, Y. Liu.Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. Advances in Neural Information Processing Systems, 2022.

[20] R. Mokady, A. Hertz, K. Aberman.Null-Text Inversion for Editing Real Images Using Guided Diffusion Models. Proceedings of CVPR, 2023.

[21] N. Carlini, J. Hayes, M. Nasr. Extracting Training Data from Diffusion Models.USENIX Security Symposium, 2023.

[22] N. Liu, S. Li, Y. Du. Compositional Visual Generation with Composable Diffusion Models.European Conference on Computer Vision (ECCV), 2022.

[23] P. Esser, J. Chiu, P. Atighehchian. Structure and Content-Guided Video Synthesis with Diffusion Models.Proceedings of ICCV, 2023.

[24] N. Ruiz, Y. Li, V. Jampani, Y. Pritch. Dreambooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. Proceedings of CVPR, 2023.

[25] P. Dhariwal, A. Nichol.Diffusion Models Beat GANs on Image Synthesis. Advances in Neural Information Processing Systems, 2021.

[26] A. Vahdat, F. Williams, Z. Gojcic. Latent Point Diffusion Models for 3D Shape Generation. Advances in Neural Information Processing Systems, 2022.

[27] J. Ho, T. Salimans, A. Gritsenko. Video Diffusion Models. Advances in Neural Information Processing Systems, 2022.

[28] J. Austin, D. D. Johnson, J. Ho, D. Tarlow, R. Salakhutdinov. Structured Denoising Diffusion Models in Discrete State-Spaces. Advances in Neural Information Processing Systems, 2021.

[29] M. Xu, L. Yu, Y. Song, C. Shi, S. Ermon, J. Tang. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. arXiv preprint arXiv:2203.02923, 2022.

[30] S. Luo, W. Hu. Diffusion Probabilistic Models for 3D Point Cloud Generation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[31] T. Karras, M. Aittala, T. Aila, S. Laine. Elucidating the Design Space of Diffusion-Based Generative Models. Advances in Neural Information Processing Systems, 2022.

[32] L. Zhou, Y. Du, J. Wu. 3D Shape Generation and Completion Through Point-Voxel Diffusion. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021.

[33] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi.Cascaded Diffusion Models for High Fidelity Image Generation. Journal of Machine Learning Research, 2022.

[34] B. Poole, A. Jain, J. T. Barron, B. Mildenhall. Dreamfusion: Text-to-3D Using 2D Diffusion.arXiv preprint arXiv:2209.14988, 2022.

[35] X. He, Y. Zhu, Y. Mo. Origin of fast ion diffusion in super-ionic conductors. Nature Communications, 2017.

[36] J. Hoja, A. M. Reilly, A. Tkatchenko. First-principles modeling of molecular crystals: structures and stabilities. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2017.

[37] L. Zhang, D. Lin, H. Wang, et al. Active learning of uniformly accurate interatomic potentials for materials simulation. Physical Review Materials, 2019.

[38] J. Mancuso, A. M. Mroz, K. N. Le, C. H. Hendon. Electronic structure modeling of metal–organic frameworks. Chemical Reviews, 2020.

[39] C. Bartel, S. L. Millican, A. Deml, J. Rumptz, et al. Physical descriptor for the Gibbs energy of inorganic crystalline solids. Nature Communications, 2018.

[40] J. Watson, D. Juergens, N. R. Bennett, B. Trippe, et al. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. BioRxiv, 2022.

[41] J. Noh, J. Kim, H. S. Stein, B. Sanchez-Lengeling, et al. Inverse design of solid-state materials via a continuous representation. Matter, 2019.

[42] C. Gu, N. Hosono, S. Sakaki. Design and control of gas diffusion in a nanoporous soft crystal. Science, 2019.

[44] D. Weber, A. Senyshyn, K. Weldert. Structural insights within the lithium superionic conductor. Chemistry of Materials, 2016.