Physics-Constrained Diffusion for Lightweight Composite Material Design

Anonymous Author(s)

Affiliation Address email

Abstract

Composite materials play a pivotal role in diverse engineering applications, particularly in the development of lightweight yet high-performance structures. However, their generative design has received far less attention than that of other material classes. A major challenge is that existing generative models often push component weights toward extreme minima, sometimes even yielding negative values, which are physically impossible for material compositions. In this work, we propose a diffusion-based generative framework tailored for lightweight composites. Specifically, we introduce a physics-constrained diffusion model (PCDiff) that integrates domain-specific constraints into the denoising process, ensuring generated candidates are both high-fidelity and physically plausible. In particular, we enforce two key constraints, i.e., non-negativity and sum-to-one conditions on composite compositions, through regularization within the diffusion process. Experimental evaluations demonstrate that our approach consistently outperforms existing generative models in terms of validity, density, and coverage with respect to target physical properties. This study underscores the potential of physics-guided generative modeling for accelerating the discovery of lightweight composites.

Introduction

2

3

5

6

7

10

11

12

13

14

15

16

28

30

31

34

Artificial Intelligence (AI) is rapidly transforming scientific fields by enhancing hypothesis generation, accelerating experimental design, and uncovering insights beyond traditional trial-and-error processes 19 [10]. Materials design has been a driving force behind technological and industrial progress, from the 20 steel that fueled the industrial revolution to the semiconductors that power the information age [1]. 21 The task of inverse material design seeks to identify the optimal material compositions or structures in 22 order to achieve the specified target properties. Recently, deep generative models such as variational 23 autoencoders (VAEs), generative adversarial networks (GANs) and diffusion models have drastically 24 accelerated the landscape of inverse design [2, 9]. Deep generative models are capable of capturing 25 the latent relationships between materials' data and properties, and new material candidates can be 26 sampled from the learned latent space. 27

In this work, we focus on a relatively underexplored material class — lightweight composite materials—which play a vital role in many engineering applications but have received far less attention in generative modeling compared to materials such as crystals or alloys. Composites are typically composed of matrices, fibers and varying fillers. Accurate modeling of composites' behavior remains challenging due to the complex interplay between material composition, manufacturing processes and the respective characteristics [4]. This presents two key challenges for inverse lightweight composite design. First, most existing generative modeling approaches have been developed for other material classes, and their architectures or representations cannot be directly adapted to capture the 35 heterogeneous microstructures and multi-phase nature of composites. Second, composites impose

unique physics constraints that are often overlooked in conventional generative frameworks. As a result, existing models may push mixture weights toward unrealistic extremes, even producing negative values that violate physical rules.

To this end, we propose a novel physics-constrained diffusion model, termed PCDiff, specifically 40 designed for the inverse design of lightweight composite materials. PCDiff explicitly incorporates 41 domain-specific constraints into the diffusion process. In particular, non-negativity and sum-to-42 one physical constraints on component proportions are enforced through regularization, ensuring 43 that generated candidates remain physically valid while maintaining diversity. By embedding 44 these inductive biases directly into the generative dynamics, PCDiff not only produces high-fidelity 45 composite compositions but also aligns the generated solutions with fundamental physical laws, thereby narrowing the gap between theoretical design and real-world applicability. Empirically, 47 we evaluate the proposed PCDiff model on an in-house composite material dataset that captures 48 diverse composition–property relationships. Across multiple evaluation metrics, PCDiff demonstrates 49 superior performance, consistently generating candidates with higher validity and broader coverage of the design space. These results highlight the effectiveness of integrating physics-based constraints into the diffusion process and underscore the promise of PCDiff for lightweight composites design.

53 **2 Related Work**

AI-driven Composite Materials. Composites are widely used in aerospace, automotive and construction industries, and precise engineering is required to achieve desired mechanical and 55 thermal properties. Traditionally, the development of composites relies on experimental approaches 56 and computational methods such as finite element analysis (FEA). In recent years, AI techniques 57 have been increasingly integrated into the forward design process. For instance, supervised learning 58 algorithms have been applied to predict material properties, classify material types, and perform failure 59 analysis, and surrogate models have been constructed to approximate computationally expensive FEA simulations [4]. Beyond forward design, AI has also shown promise in inverse design. For example, 62 the integration of deep convolutional GANs (DCGANs) with convolutional neural networks (CNNs) has been explored to accelerate the discovery of two-phase composite materials [8], demonstrating 63 the potential of generative models to navigate large design spaces efficiently. Despite these advances, 64 AI-driven research for composite materials remains relatively limited compared to other material 65 classes, leaving significant opportunities for developing specialized generative models that account 66 for their unique physical characteristics. 67

Generative Models for Inverse Material Design. Generative models are promising for inverse 68 material design tasks by directly generating materials from the learned latent space given desired 69 properties. Early efforts explored GAN-based and VAE-based architectures, which are shown feasible 70 for inverse material design [8]. More recently, diffusion-based generative models have shown superior 71 performance and flexibility for such tasks. CDVAE designs a noise conditional score network as the 72 decoder of VAE and incorporates the physical inductive bias of crystal's stability into consideration 73 [11]. MatterGen is a diffusion-based model that generates stable and diverse inorganic materials with crystalline structures by gradually refining atom types, coordinates and the periodic lattice 75 [12]. Beyond these domain-specific models, diffusion models themselves have undergone significant 76 methodological developments that broaden their applicability. For instance, denoising diffusion 77 probabilistic models (DDPMs) [3] further improve the noise schedules to generate more realistic 78 outputs. For composites data, the TabDDPM [5] is more proper as TabDDPM is designed to model 79 tabular data with vectors of heterogeneous features. 80

81 3 Proposed PCDiff Model

82 3.1 Problem Statement

The inverse design of lightweight composite materials can be naturally formulated as a constrained optimization problem, where the goal is to identify candidate compositions that achieve target material properties (e.g., Yield Strength, Thermal Conductivity). Let $\mathbf{x}_0 = [\mathbf{x}_0^1, \mathbf{x}_0^2, ..., \mathbf{x}_0^D] \in \mathbb{R}^D$ denote the vector of proportions of the D components in a composite. The inverse design task seeks to minimize the discrepancy between predicted properties $f(\mathbf{x}_{\text{gen}})$ and the desired target properties $f(\mathbf{x}_{\text{gen}})$ wherein the new material \mathbf{x}_{gen} is generated via a diffusion generative process parameterized by $f(\mathbf{x}_{\text{gen}})$.

$$\min_{\mathbf{x}_{gen} \in \mathbb{R}^{D}} \quad \mathcal{L}(f(\mathbf{x}_{gen}), y^{*}),$$
s.t. $\mathbf{x}_{gen} \sim p_{\theta}(\mathbf{x}_{0}|y^{*}), \quad p_{\theta}(\mathbf{x}_{0:T} \mid y^{*}) = p(\mathbf{x}_{T}) \prod_{t=1}^{T} p_{\theta}(\mathbf{x}_{t-1} \mid \mathbf{x}_{t}, y^{*}),$

$$0 \leq \mathbf{x}_{gen}^{i}, \mathbf{x}_{t}^{i} \leq 1, \quad (i = 1, ..., D), (t = 1, ..., T),$$

$$\sum_{i=1}^{D} \mathbf{x}_{gen}^{i} = 1, \sum_{i=1}^{D} \mathbf{x}_{t}^{i} = 1, \quad (t = 1, ..., T),$$
(1)

where we set the loss function \mathcal{L} to be the mean squared error (MSE) loss and the forward predictor f to be a shallow MLP. The non-negative constraint ensures that no component is assigned a negative proportion during diffusion, while the summation constraint enforces that the mixture proportions form a valid composite composition. Without these constraints, existing generative models often push component weights toward extreme minima, sometimes yielding negative values that violate fundamental physical rules. This formulation highlights the dual challenge of inverse design for composites: simultaneously maintaining physical feasibility and steering the design toward lightweight composites. In this work, we address this challenge through a diffusion-based generative model that learns to sample directly from the feasible composition space while conditioning on target properties, thereby integrating optimization and physical validity within a unified framework.

3.2 Physics-Constrained Diffusion for Composites

90

91

93

94

95

96

97

98

99

Diffusion models are likelihood-based models and handle the data through forward and reverse Markov processes. The forward process gradually adds noise to the initial sample \mathbf{x}_0 , i,e., $q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \mu_t = \sqrt{1-\beta_t}\,\mathbf{x}_{t-1},\, \Sigma_t = \beta_t \mathbf{I}\right), t=1,...,T,$ where the variance parameter β_t can be fixed to a constant or chosen as a schedule of the timesteps. The reverse process is to invert the forward noising process and we want $p(\mathbf{x}_{0:T}) = \mathbf{x}_t$

 $p(\mathbf{x}_T) \prod_{t=1}^T p(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ to gradually remove noise at each step. Since the exact reverse distributions are intractable, and a neural network parameterized by θ is used to learn the approximation. Following the training of DDPM [3], the loss can be simplified to be a noise-prediction MSE:

108 $\mathcal{L}_{\text{base}}(\theta) = \mathbb{E}_{\mathbf{x}_0, \, \epsilon, \, t} \left[\|\epsilon - \epsilon_{\theta}(\mathbf{x}_t, t)\|^2 \right]$, where $\mathbf{x}_t = \sqrt{\overline{\alpha}_t} \, \mathbf{x}_0 + \sqrt{1 - \overline{\alpha}_t} \, \epsilon$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$. Despite the advances in diffusion models, existing models cannot naturally satisfy these physical constraints

as shown in Eq. (1). To this end, we seek to incorporate these physical constraints as additional

regularization terms in the diffusion process, i.e.,

$$\phi_{\text{non-neg}}(\mathbf{x}_{\text{gen}}) = \sum_{i=1}^{D} \max(0, -\mathbf{x}_{\text{gen}}^{i})^{2} + \sum_{t=1}^{T} \sum_{i=1}^{D} \max(0, -\mathbf{x}_{t}^{i})^{2}, \quad \text{(non-negativity regularization)}$$
(2)

$$\phi_{\text{sum}}(\mathbf{x}_{\text{gen}}) = \left(\sum_{i=1}^{D} \mathbf{x}_{\text{gen}}^{i} - 1\right)^{2} + \sum_{t=1}^{T} \left(\sum_{i=1}^{D} \mathbf{x}_{t}^{i} - 1\right)^{2}, \quad \text{(sum-to-one regularization)} \quad (3)$$

 $\mathcal{L}_{\text{reg}}(\mathbf{x}_{\text{gen}}) = \alpha_1 \phi_{\text{non-neg}}(\mathbf{x}_{\text{gen}}) + \alpha_2 \phi_{\text{sum}}(\mathbf{x}_{\text{gen}}), \tag{4}$

and the weighting coefficients α_1 and α_2 can control the relative importance of each regularization term. Then, the overall regularized loss expands to: $\mathcal{L}_{overall} = \mathcal{L}_{base} + \mathcal{L}_{reg}$. These regularizers allow the generation process to be steered toward specific directions, i.e., ensuring all outputs remain non-negative and that their components sum to one. By adjusting the weights, we can control the trade-off between adhering to the data distribution and enforcing physical constraints, leading to more meaningful and valid generated samples in composites material applications.

118 4 Experiments

Dataset. We have constructed a dataset for carbon composite. In our dataset of 3,400 carbon composite entries, each material is described by its composition, including the proportion of carbon

fiber, epoxy, and up to four types of fillers—Carbon Nanotubes (CNTs), Graphene, Copper, and 121 Nickel. These compositional inputs directly influence the resulting material properties. For example, 122 adding CNTs or Graphene generally enhances the mechanical strength due to their high intrinsic 123 stiffness, while metallic fillers like Copper and Nickel can improve thermal conductivity. The interplay 124 between the proportions of each component determines how the composite balances strength and heat 125 transfer, making the mapping from composition to properties a complex but physically meaningful 126 relationship. This dataset therefore captures the fundamental connection between the materials and 127 composite properties such as Yield Strength (MPa) and Thermal Conductivity (W/mK). 128

Evaluation Metrics. We evaluate generative models through the following aspects:

- Validity_{reg} quantifies the proportion of generated samples that satisfy both the non-negativity and sum-to-one constraints. For this evaluation, we generate one sample conditioned on each training instance's property values.
- Validity_{fab} measures the proportion of generated samples that fall within the valid region for practical fabrication. In the context of composite materials, we define this valid region based on domain experts' criteria, such as percolation limits, and use it to assess whether the generated samples are physically feasible. We fix the desired property values to be 3,500 and 1.0 for Yield Strength and Thermal Conductivity and generate 100 samples from each model for evaluation.
- Density and Coverage [5, 7]. The metrics are designed to differentiate the fidelity and diversity of the generated samples.

Baselines and Experimental Setup. We compare the proposed method with a conditional VAE model [6] and the state-of-the-art TabDDPM [5] model. For fair comparisons, we keep the neural networks within each model to be MLP-based, to align with our method. We split the dataset into training and testing set with ratio of 0.8 : 0.2. The number of training epochs is set to 100.

Experimental Results. The overall experimental results are summarized in Table 1. In terms of validity, both CVAE and our proposed PCDiff achieve 100%. However, closer inspection reveals that CVAE suffers from mode collapse, repeatedly generating identical samples, which undermines 147 its practical utility. In contrast, the vanilla diffusion model (TabDDPM) exhibits very low va-148 lidity, highlighting the necessity of incorporating physical constraints into the diffusion process. 149 Regarding Density and Coverage, diffusion-based models consistently outperform the VAE-based 150 model. Notably, PCDiff achieves the best performance, with Density exceeding TabDDPM by 4.3% 151 and Coverage by 9.1%, demonstrating both higher fidelity and greater diversity in the generated 152 153 composites.

Table 1: Experimental evaluations on the composite dataset. (Unit:%)

Model \ Metric	Validity _{reg}	Validity _{fab}	Density	Coverage
CVAE	100	100	37.5	7.8
TabDDPM	15	0	44.7	10.5
PCDiff (ours)	100	100	49.0	19.6

5 Conclusion

154

155

156

157

158

160

161

130

131

132

133

134

135

136

137

138

139

In this work, we address the relatively underexplored problem of generative modeling for lightweight composite materials by introducing PCDiff, a physics-constrained diffusion framework. By embedding domain-specific constraints, such as non-negativity and sum-to-one composition rules, directly into the diffusion process, PCDiff ensures that generated candidates are both physically valid and of high fidelity. Empirical results on composite datasets demonstrate that our approach outperforms state-of-the-art generative baselines in terms of validity, density, and design space coverage. This study lays the groundwork for future research in integrating richer physics and domain knowledge into generative AI for material design.

References

- [1] Mouyang Cheng, Chu-Liang Fu, Ryotaro Okabe, Abhijatmedhi Chotrattanapituk, Artittaya Boonkird,
 Nguyen Tuan Hung, and Mingda Li. Ai-driven materials design: a mini-review. *CoRR*, abs/2502.02905,
 2025.
- [2] Xiao-Qi Han, Xin-De Wang, Meng-Yuan Xu, Zhen Feng, Bo-Wen Yao, Peng-Jie Guo, Ze-Feng Gao, and
 Zhong-Yi Lu. Ai-driven inverse design of materials: Past, present and future. *CoRR*, abs/2411.09429,
 2024.
- [3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Hugo Larochelle,
 Marc' Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, Advances in
 Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems
 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.
- [4] Manickaraj Karuppusamy, Ramakrishnan Thirumalaisamy, Sivasubramanian Palanisamy, Sudha Nagamalai, Ehab El Sayed Massoud, and Nadir Ayrilmis. A review of machine learning applications in polymer composites: advancements, challenges, and future prospects. *Journal of Materials Chemistry A*, 2025.
- [5] Akim Kotelnikov, Dmitry Baranchuk, Ivan Rubachev, and Artem Babenko. Tabddpm: Modelling tabular
 data with diffusion models. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt,
 Sivan Sabato, and Jonathan Scarlett, editors, *International Conference on Machine Learning, ICML 2023*,
 23-29 July 2023, Honolulu, Hawaii, USA, volume 202 of Proceedings of Machine Learning Research,
 pages 17564–17579. PMLR, 2023.
- [6] Jaechang Lim, Seongok Ryu, Jin Woo Kim, and Woo Youn Kim. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of cheminformatics*, 10(1):31, 2018.
- Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity
 and diversity metrics for generative models. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 7176–7185. PMLR, 2020.
- [8] José Pablo Quesada-Molina, Hossein Mofatteh, Abdolhamid Akbarzadeh, and Stefano Mariani. Navigating
 beyond the training set: A deep learning framework for inverse design of architected composite materials.
 Eng. Appl. Artif. Intell., 150:110473, 2025.
- [9] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak,
 Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne Bergen, Carla P.
 Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun Manrai, Debora S.
 Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Velickovic, Max Welling, Linfeng
 Zhang, Connor W. Coley, Yoshua Bengio, and Marinka Zitnik. Scientific discovery in the age of artificial
 intelligence. Nat., 620(7972):47–60, 2023.
- [10] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak,
 Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial
 intelligence. *Nature*, 620(7972):47–60, 2023.
- [11] Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi S. Jaakkola. Crystal diffusion
 variational autoencoder for periodic material generation. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.
- [12] Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Sasha Shysheya,
 Jonathan Crabbé, Lixin Sun, Jake Smith, Ryota Tomioka, and Tian Xie. Mattergen: a generative model for
 inorganic materials design. CoRR, abs/2312.03687, 2023.