Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Generative Inverse Design of Shell Metamaterials through Continuous-Latent Conditional Flow Matching

Sibo Zhang
The University of Texas at Austin
Yunlan Zhang
yunlan.zhang@austin.utexas.edu
The University of Texas at Austin
Krishna Kumar
The University of Texas at Austin

Article

Keywords:

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11234832/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Generative Inverse Design of Shell Metamaterials through Continuous-Latent Conditional Flow Matching

Sibo Zhang[a], Yunlan Zhang*[a], and Krishna Kumar*[a]

Abstract

Mechanical metamaterials derive programmable properties from a repeated unit cell's geometry. A stress–strain curve summarizes a three-dimensional shell geometry in a few numbers, so distinct geometries can share a curve and the inverse map from curve to geometry is one-to-many. We develop a continuous-latent conditional flow-matching framework that designs periodic shell unit cells for prescribed compressive responses. Each shell geometry is a continuous periodic function, and a function encoder learns a basis that compresses it into 32 coefficients. Conditional flow matching learns the response-conditioned coefficient distribution and generates many candidate geometries. On 200 held-out targets, our 5.9-million-parameter model reaches a mean best-of-20 NRMSE of 0.074 at 73% validity, against 0.118 and 48% for a 40-million-parameter tokenized-diffusion baseline. Sampling requires 26.8 ms per candidate versus 2.03 s for the baseline. Finite-element analysis of 981 shortlisted candidates yields a mean best-design NRMSE of 0.092, compared with a simulation-to-simulation reproducibility of 0.016–0.063.

1 Introduction

Mechanical metamaterials derive their effective properties from the geometry of a repeated unit cell, so the unit-cell geometry is the design space [1–3]. Periodic shell metamaterials, such as triply periodic minimal surface (TPMS) metamaterials, are formed by smooth, continuous, and interconnected thin-walled surfaces that are periodic in three spatial dimensions [4,5]. These architectures can provide high stiffness, efficient load transfer, and substantial energy absorption at low relative density [6–8]. Buckling and other structural transitions give them nonlinear responses that include peak loads, stress softening, stress plateaus, and multistability [9–11]. These behaviors support applications in lightweight structures, energy absorption, adaptive systems, and biomedical devices [2,8,12–15]. Such applications specify their mechanical requirement as a target stress–strain curve. An energy absorber needs an extended plateau at a prescribed stress, and a prosthesis needs compliance matched to the tissue it replaces [14].

In this work, we address the inverse design of periodic thin-shell unit cells from prescribed compressive stress–strain responses. Such responses characterize the unit cell's initial stiffness, peak strength, and post-peak softening or plateau behavior [8,16]. Under compression, thin shells buckle, develop self-contact, and switch between instability modes, so a small change in geometry can reshape the stress–strain curve. A stress–strain curve is also a coarse summary of the geometry, one function of strain standing for a full three-dimensional field. Many distinct shells therefore produce nearly the same curve, and the inverse mapping from curve to geometry is one-to-many [17,18]. Generating valid periodic shell unit cells directly from target nonlinear mechanical responses therefore remains a key challenge in generative inverse design.

Inverse design of mechanical metamaterials has evolved from simulation-based optimization toward data-driven approaches that learn structure–property relationships and design distributions [3,19]. Topology optimization provides a systematic framework for distributing material within a design domain [20], but incorporating geometric nonlinearity and structural instability generally requires repeated nonlinear analyses and specialized stability treatments [21]. Data-driven studies have learned inverse or latent representations of structure–property relationships for truss, spinodoid, and cellular metamaterials [22–24], while surrogate-assisted evolutionary methods have enabled the design of structures with prescribed nonlinear loading responses [25–27]. More recently, generative adversarial networks, graph-based generative models, and diffusion models have extended inverse design toward learned design distributions and conditional generation [17,28–33]. These methods employ a variety of geometric representations, including parameter vectors, voxel grids, graphs, and algebraic expressions [17,18,23,33,34]. For instance, DiffuMeta represents periodic shell geometries as tokenized algebraic expressions and uses a diffusion transformer to generate candidate structures conditioned on nonlinear compressive responses [18], demonstrating that diffusion models can address non-unique inverse-design problems. However, its generation relies on discrete token sequences and requires conversion between continuous representations and discrete expressions. Such discrete geometric descriptions sever the continuity of the design space, leaving interpolation between designs undefined. Similarly, target responses are often represented by discrete values at predefined strain levels, tying the conditioning input to a fixed response discretization [17,18]. These limitations motivate generative inverse-design approaches that can continuously represent complex periodic geometries while supporting more flexible response representations.

To avoid restricting complex periodic geometries to predefined parameters or discrete token representations, we represent the shell geometry as a continuous function. The midsurface of each shell is defined as the zero level set of a periodic implicit field $\Psi(x, y, z)$ on the unit cell, thereby transforming a complex three-dimensional shape into a continuous function [35,36]. Under appropriate regularity, these geometry fields form a family inside a Hilbert space of square-integrable periodic functions. A function encoder learns a shared set of neural-network basis func-

[a] Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, Austin, TX, USA.

* Corresponding authors: yunlan.zhang@austin.utexas.edu and krishnak@utexas.edu.

tions whose span covers this family[37], and projection onto this basis represents each complex periodic shell by a compact coefficient vector. The basis networks take integer-frequency Fourier features of the coordinates, so every basis function, and every field built from them, is exactly periodic across unit-cell boundaries[38,39]. Inverse design can be viewed as a mapping from a target response to a geometry function, namely from a prescribed stress–strain function $\sigma^*(\varepsilon)$ to a geometry field Ψ. Operator learning addresses such maps[40–42], and basis-to-basis methods reduce them to a deterministic map between coefficient vectors[43]. However, the inverse problem considered here is one-to-many. The same target response may correspond to multiple geometry coefficient vectors, and therefore cannot be represented by a single deterministic mapping. We instead learn the distribution of geometry coefficients conditioned on the target response, so one target yields many candidate geometries.

Conditional flow matching (CFM) provides a continuous generative approach to this problem. Flow matching learns a continuous transport from a simple prior distribution to the data distribution by regressing a time-dependent velocity field[44]. Conditioned on the target response, CFM transports random samples from a Gaussian prior to the corresponding distribution of geometry coefficients. Different initial samples are transported to different coefficients, so a single target response yields multiple geometrically distinct candidate shells. Related continuous-time formulations include rectified flow, stochastic interpolants, and optimal-transport paths[45–47]. Diffusion models define a forward noising process and learn its reverse through denoising or score estimation[48,49]. Flow matching instead learns the velocity field that connects the prior and target distributions and samples by solving the corresponding ordinary differential equation[44], which in practice takes tens of solver steps rather than thousands of denoising iterations. Flow matching has recently been applied to crystalline materials, three-dimensional shapes, airfoils, and structural design[50–55], while conditional continuous normalizing flows have also been explored for porous-microstructure inverse design[56]. To our knowledge, CFM has not previously been used to generate periodic shell metamaterial unit-cell geometries conditioned on nonlinear stress–strain responses within a learned continuous geometry space.

Generative inverse design produces many candidate structures per target and ranks them with a learned forward surrogate. Prediction errors can alter this ranking, particularly for generated geometries containing local features that are difficult for the surrogate to resolve accurately[57]. Consequently, the candidate with the lowest surrogate-predicted error may not achieve the lowest error under high-fidelity evaluation. This discrepancy is particularly relevant to thin-shell metamaterials, whose buckling and post-buckling responses are sensitive to geometric imperfections and small perturbations of the instability path[10,11,58,59]. Mesh discretization and large-deformation contact also perturb the simulated response. We therefore measure how closely the finite-element workflow reproduces its own results and read every generated design against that reproducibility. These considerations motivate a hierarchical workflow in which a surrogate rapidly screens a large candidate set and finite-element analysis independently evaluates a fixed shortlist[60].

We therefore develop a continuous-latent CFM framework for the inverse design of periodic shell metamaterials. For a single target response, it generates multiple response-compatible geometries. A hierarchical evaluation procedure combines surrogate-based candidate screening with independent finite-element analysis (FEA) validation. We validate 981 shortlisted candidates from the primary model through FEA and benchmark five generative models on 200 target responses held out from both our training split and the baseline's released training set, under a shared candidate budget, validity criterion, and forward surrogate. Through controlled comparisons, we examine the roles of geometry representation, response representation, and CFM velocity-network capacity. The function-space formulation additionally enables geometry interpolation and off-grid response conditioning.

2 Results

2.1 Inverse-design framework

The proposed framework maps a prescribed nominal stress–strain response of the unit cell under uniaxial compression along the z direction to a distribution of periodic shell unit-cell geometries. Spatial repetition of a unit cell defines the corresponding metamaterial. All unit cells have the same relative density, $\rho = 0.1$, so the shell wall occupies 10% of the unit-cell volume and the wall thickness follows from the midsurface area. All designs are evaluated up to a nominal compressive strain of 0.30. As summarized in Fig. 1(a), each shell geometry is represented by $K = 32$ coefficients over a learned periodic basis. The target response is represented either by the raw 11-point stress vector or by $L = 4$ coefficients over a learned continuous response basis. CFM learns the response-conditioned distribution of geometry coefficients, so one target response yields multiple geometrically distinct candidates. During inverse design [Fig. 1(b)], generated coefficients are decoded into periodic shell fields and filtered by the periodic-wrap validity criterion, which keeps a candidate only if its zero level set forms a single connected surface after periodic wrapping. Valid candidates are then evaluated by a frozen 3D-CNN surrogate and ranked by their predicted NRMSE relative to the prescribed response. The top-ranked valid design is defined as Candidate 1, and up to the five highest-ranked valid candidates form the predefined shortlist for FEA validation. The primary model uses the learned $K = 32$ geometry representation, raw 11-point response conditioning, and an eight-block, width-512 velocity network, denoted Raw-11D CFM (8×512).

2.2 FEA validation of surrogate-ranked candidates

FEA validation shows that the candidate structures generated by CFM reproduce the prescribed responses. For 200 evaluation targets held out from both our training split and the released DiffuMeta training set, FEA was performed on up to five top-ranked valid candidates from the primary Raw-11D CFM (8×512) model. Of the 985 submitted models, 981 completed. To assess the reliability of the adopted FEA workflow, we re-simulated 16 structures from the original dataset. The NRMSE between the repro-

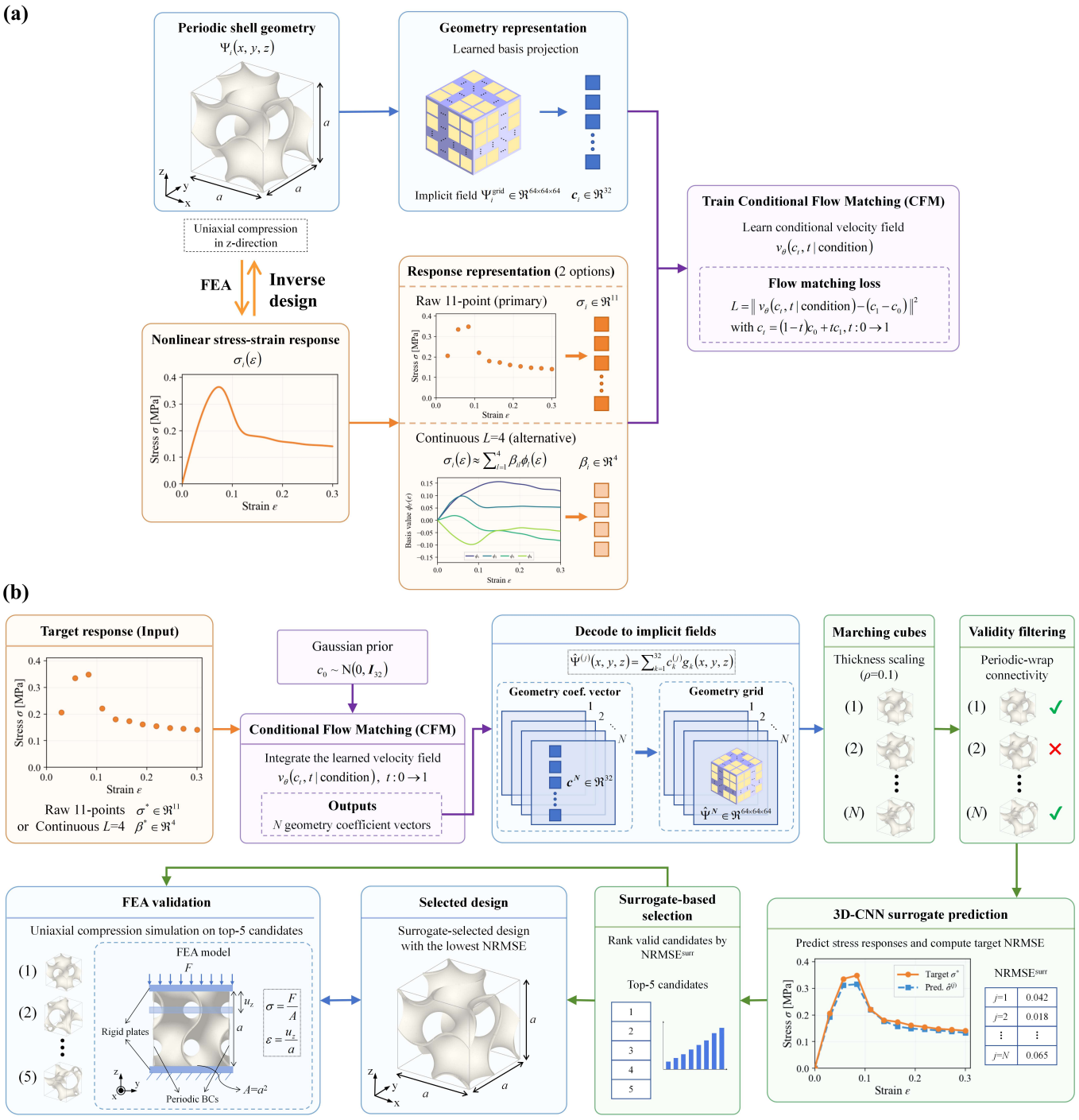


Fig. 1 Conditional flow-matching framework for inverse design of periodic shell metamaterials. **(a)** Continuous representations of shell geometry and target response, together with the response-conditioned CFM training workflow. **(b)** Candidate generation, validity filtering, surrogate ranking, and FEA evaluation of the predefined shortlist.

duced and stored responses ranged from 0.016 to 0.063, providing an empirical measure of numerical reproducibility. We define FEA-best as the completed candidate with the lowest FEA NRMSE within the predefined shortlist. Across the 200 targets, FEA-best achieved a mean NRMSE of 0.0917, a median of 0.0793, and a p_{90} of 0.1464, against a simulation-to-simulation reproducibility of 0.016 to 0.063. This high-fidelity agreement extended across distinct types of nonlinear response. The three representative targets in Fig. 2(a) exhibit post-peak softening, plateau response, and post-peak softening followed by hardening, with FEA-best NRMSE of 0.045, 0.040, and 0.056, respectively. For the same target, shortlisted candidates can also exhibit distinct geometries [Fig. 2(b)], which is direct FEA evidence that the inverse mapping is one-to-many.

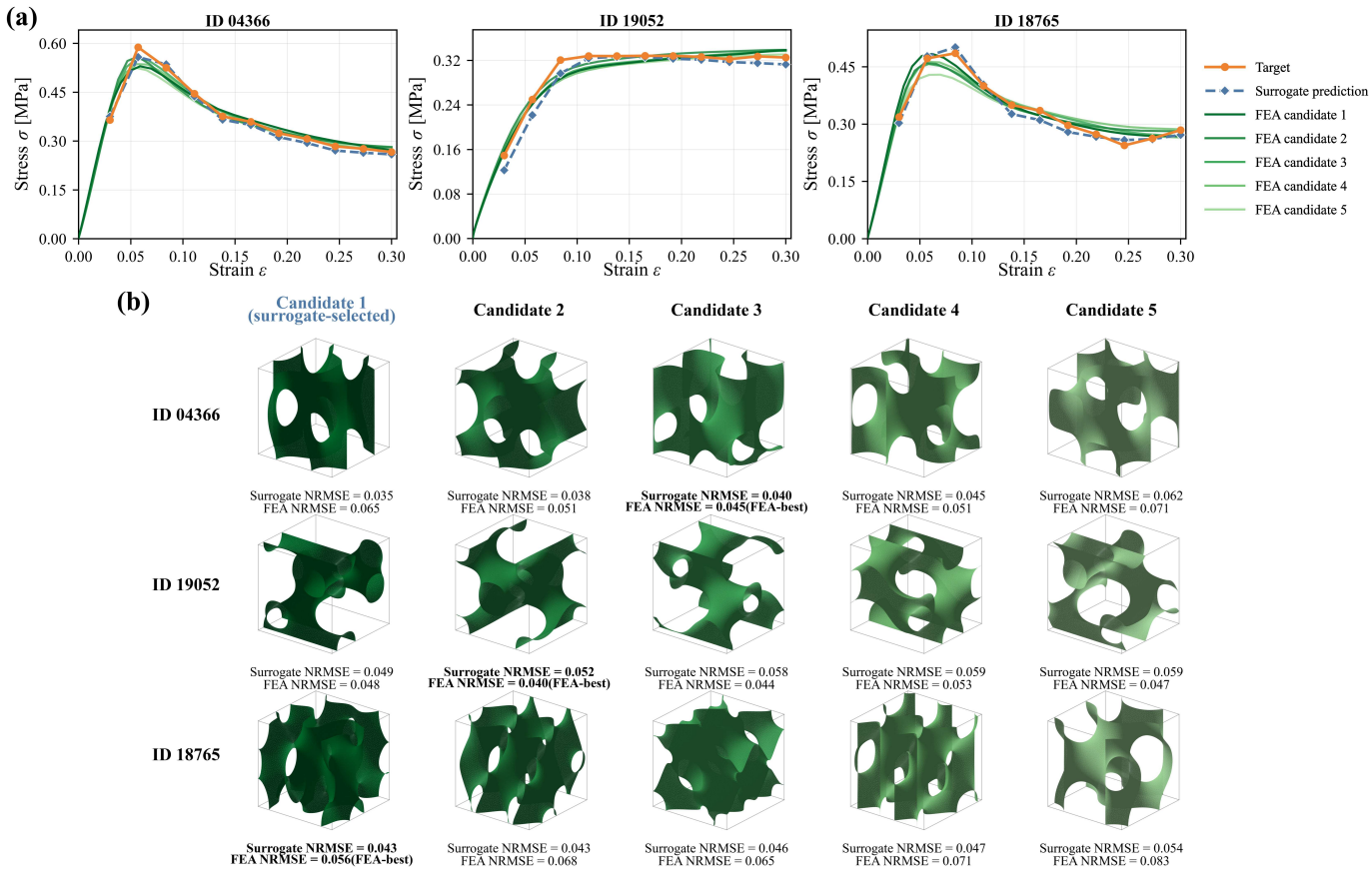


Fig. 2 FEA validation of surrogate-ranked candidates. **(a)** Prescribed target responses, the surrogate prediction for Candidate 1, and the FEA responses of the shortlisted candidates for three representative targets. **(b)** Corresponding candidate geometries with their surrogate and FEA NRMSE values.

FEA results also show that high-fidelity evaluation of the shortlist can improve final candidate selection. The surrogate-selected Candidate 1 was also FEA-best for 46 of the 200 targets. Its mean FEA NRMSE was 0.1192, whereas the mean NRMSE of FEA-best was 0.0917, corresponding to a 23.1% reduction. The median decreased from 0.1085 to 0.0793, and the p_{90} from 0.1841 to 0.1464. The representative cases in Fig. 2(b) show the same behavior. For IDs 04366 and 19052, Candidates 3 and 2 achieved lower FEA errors than Candidate 1, respectively, whereas Candidate 1 was also FEA-best for ID 18765. These results show that the surrogate produces useful candidate shortlists, while FEA improves final selection within those shortlists.

2.3 Inverse-design performance

The unified benchmark shows that the continuous-latent CFM framework outperformed the published DiffuMeta baseline, with the primary Raw-11D CFM (8×512) achieving the lowest response-matching error and the highest candidate validity. Table 1 compares four CFM configurations with DiffuMeta [18]. Natural–Fourier CFM and Raw-11D CFM (4×256) compare the Natural–Fourier 33D and learned $K = 32$ geometry representations. Stress-conditioned CFM ($L = 4$) compares the continuous $L = 4$ and raw 11-point response representations. Raw-11D CFM (8×512) increases the velocity-network capacity while keeping both representations fixed and serves as the primary model. DiffuMeta is included as an external baseline. All methods were evaluated on the same 200 doubly held-out targets, with 20 candidates generated per target and scored using the same frozen 3D-CNN surrogate.

The primary model achieved mean, median, and p_{90} best-of-20 NRMSE values of 0.0739, 0.0604, and 0.1328, respectively, together with a candidate validity of 73.0%. In comparison, DiffuMeta achieved a mean NRMSE of 0.1182 and a validity of 47.9%. The primary model generated at least one valid candidate for all 200 targets, whereas DiffuMeta yielded $n_{\text{eff}} = 186$. Target-level paired comparisons supported this advantage. The primary model achieved lower error on 68.0–76.9% of the common evaluable targets, and the 95% confidence intervals for the mean paired error differences relative to the other four methods were all above zero (Supplementary Information S3.4).

The performance improvement was also reflected in the overall error distribution, candidate validity, and sampling efficiency. As shown in Fig. 3(a,b), the NRMSE distributions of the CFM models were generally shifted toward lower errors, with the primary model achieving both the lowest p_{90} and the highest candidate validity, indicating fewer high-error targets. The representative cases in Fig. 3(c,d) are consistent with the aggregate trend. The primary model achieved the lowest error for IDs 16675 and

Table 1 Inverse-design benchmark results on 200 doubly held-out target responses with 20 candidates generated per target. Accuracy statistics summarize the per-target best-of-20 NRMSE, and n_{eff} is the number of targets for which at least one valid candidate was generated. Parameter counts refer to the generative network. Sampling times were measured on an NVIDIA Tesla T4 using a batch size of 20. CFM timings are reported as mean ± standard deviation over four repeated sampling runs using 50-step Heun integration; the DiffuMeta timing was obtained from a complete 4,000-candidate run using its 2,000-step reverse diffusion process.

Method	Geometry representation	Response representation	Generator	Generator params	Validity	n_{eff}	Mean	Median	p_{90}	Sampling time per candidate
Raw-11D CFM (8×512)	**Learned geometry basis, $K=32$**	**Raw 11-point**	**CFM, 8×512, 50 steps**	**5.87M**	**73.0%**	**200**	**0.0739**	**0.0604**	**0.1328**	26.76 ± 0.56 **ms**
Raw-11D CFM (4×256)	Learned geometry basis, $K=32$	Raw 11-point	CFM, 4×256, 50 steps	0.991M	61.8%	199	0.0891	0.0694	0.1482	15.85 ± 0.24 ms
Stress-conditioned CFM ($L=4$)	Learned geometry basis, $K=32$	Continuous $L=4$	CFM, 4×256, 50 steps	0.990M	61.4%	200	0.0904	0.0746	0.1813	15.31 ± 0.30 ms
Natural–Fourier CFM	Natural–Fourier coefficients, 33D	Raw 11-point	CFM, 4×256, 50 steps	0.992M	51.6%	199	0.0972	0.0828	0.1723	15.97 ± 0.37 ms
DiffuMeta [18]	Algebraic tokens	Raw 11-point	DiT, 2,000 steps	40.48M	47.9%	186	0.1182	0.0864	0.2118	2.03 s

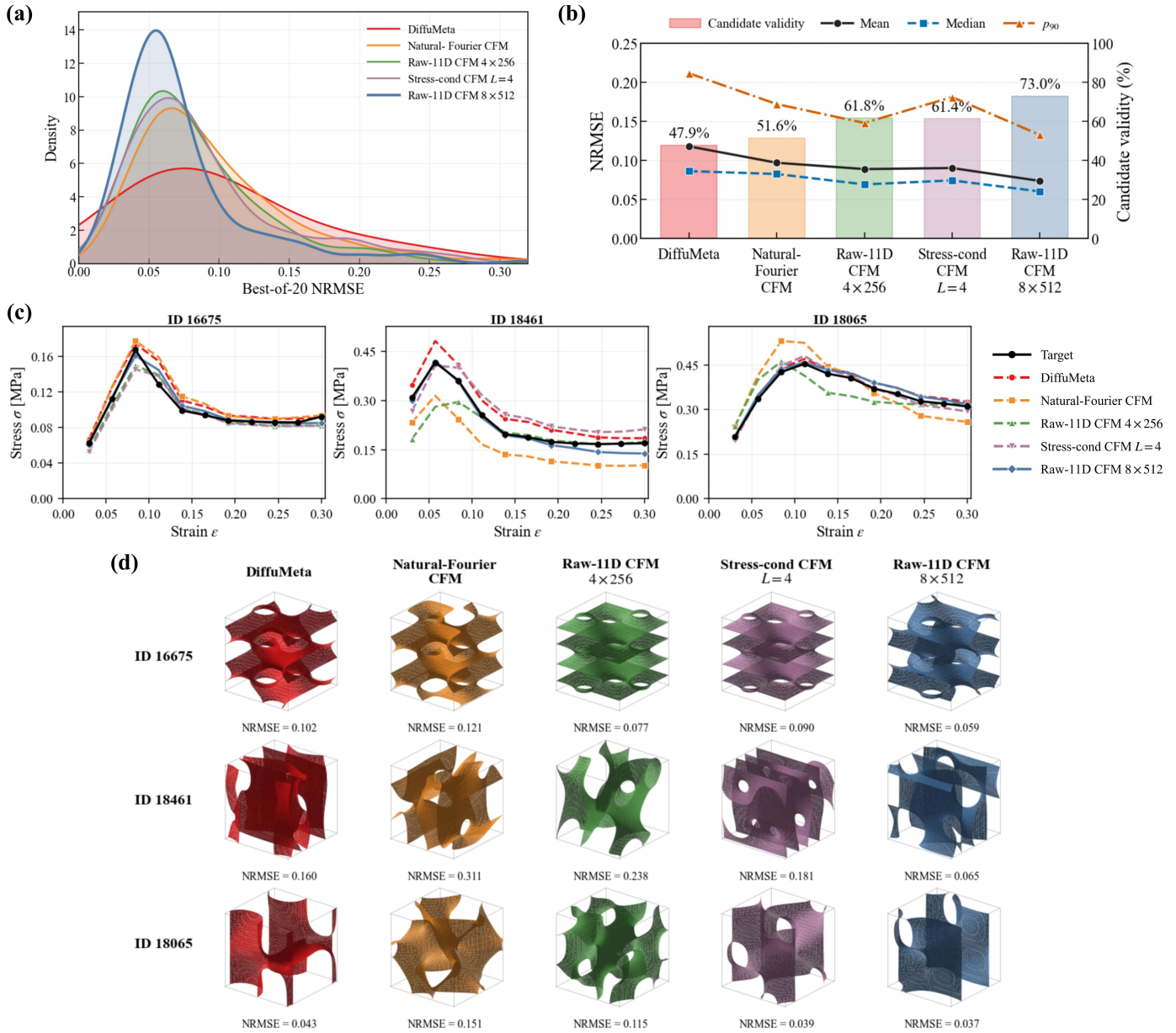


Fig. 3 Comparison of generative inverse-design methods. **(a)** Distributions of per-target best-of-20 NRMSE over the evaluable targets for each method. **(b)** Candidate validity and aggregate best-of-20 NRMSE statistics. **(c)** Target responses and surrogate predictions for Candidate 1 from each method for three representative targets. **(d)** Corresponding generated geometries and surrogate NRMSE values.

18461, whereas multiple methods matched the target response well for ID 18065. On an NVIDIA Tesla T4, the primary model required 26.76 ± 0.56 ms per candidate on average, compared with 2.03 s for DiffuMeta, a 76-fold difference under the evaluated

implementations and hardware. The overall ordering of the five methods was unchanged when candidates were rescored using an independent coefficient-space MLP surrogate or when validity was evaluated using the DiffuMeta-style criterion (Supplementary Information S3.2 and S3.3).

Comparisons within the CFM framework show that the learned geometry representation and increased velocity-network capacity both improve inverse-design performance, with network capacity producing the largest improvement. At the same 4×256 network capacity and with the raw 11-point response representation, replacing the Natural–Fourier 33D geometry representation with the learned $K = 32$ representation reduced the mean NRMSE from 0.0972 to 0.0891 and increased validity from 51.6% to 61.8%. With the geometry and response representations fixed, increasing the velocity network from 4×256 to 8×512 further reduced the mean NRMSE to 0.0739 and increased validity to 73.0%. In comparison, the raw 11-point and continuous $L = 4$ response representations produced similar mean NRMSE values at the smaller network capacity (0.0891 and 0.0904, respectively). Thus, performance within the CFM framework reflects the combined effects of geometry representation, response representation, and model capacity (Section 2.4).

2.4 Effects of representation and CFM model configuration

To explain the performance differences observed in the preceding benchmark, we conducted controlled comparisons to examine the effects of geometry coefficient dimension, geometry representation, response representation, and velocity-network capacity on inverse-design performance (Fig. 4). As shown in Fig. 4(a), increasing K from 8 to 32 reduced the reconstruction error by several orders of magnitude and progressively recovered the principal surface features of the original geometries. At $K = 32$, the mean, median, and p_{90} reconstruction errors were all approximately 10^{-3}. Increasing the dimension to $K = 64$ provided little additional reconstruction benefit. For matched Raw-11D CFM (4×256) generators, however, candidate validity decreased from 61.8% to 34.4%, and the mean best-of-20 NRMSE increased from 0.0891 to 0.1381. We therefore selected $K = 32$ as the geometry coefficient dimension. Complete numerical results are provided in Supplementary Information S3.5.

At fixed CFM capacity, the learned geometry representation improved both response-matching accuracy and candidate validity. Figure 4(b) compares the 33-dimensional Natural–Fourier representation with the learned $K = 32$ geometry representation, both using the raw 11-point response representation and a 4×256 CFM. After standardization, the learned coefficients followed the standard-normal reference more closely, whereas the Natural–Fourier coefficients remained more concentrated near zero with broader tails. The validation flow-matching loss decreased from 0.713 to 0.674, the mean best-of-20 NRMSE decreased by approximately 8.3% from 0.0972 to 0.0891, and candidate validity increased by 10.2 percentage points from 51.6% to 61.8%.

Compressing the original response into a continuous response representation largely preserved inverse-design performance. The four learned response basis functions provided a compact continuous representation of the stress–strain response family, and the three representative reconstructions in Fig. 4(c) retained major nonlinear features, including pronounced peaks, post-peak softening, and plateau behavior, with reconstruction NRMSE values of 0.008–0.017. Complete reconstruction results are provided in Supplementary Information S1.2. At the same learned $K = 32$ geometry representation and 4×256 CFM capacity, the raw 11-point and continuous $L = 4$ response representations produced similar results, with mean best-of-20 NRMSE values of 0.0891 and 0.0904 and candidate validities of 61.8% and 61.4%, respectively [Fig. 4(d)]. Thus, the $L = 4$ representation largely preserved average inverse-design performance while providing a continuous conditioning representation that can be evaluated at off-grid strain locations, as examined in Section 2.5.

Increasing the velocity-network capacity produced the largest improvement in average inverse-design accuracy. With the learned $K = 32$ geometry representation and raw 11-point response representation fixed, increasing the velocity network from 4×256 to 8×512 reduced the mean best-of-20 NRMSE by approximately 17.1%, from 0.0891 to 0.0739, while candidate validity increased by 11.2 percentage points, from 61.8% to 73.0% [Fig. 4(e)].

2.5 Continuous geometry interpolation and off-grid response conditioning

The learned coefficient spaces support interpolation between designs and response conditioning at strain locations outside the training grid.

Linear interpolation in the learned geometry coefficient space produced a continuous geometric transition between two valid structures, accompanied by smoothly varying mechanical responses. We interpolated between the encoded coefficients of Design A (ID 3483) and Design B (ID 10153) as

$$\mathbf{c}(\alpha) = (1-\alpha)\mathbf{c}_A + \alpha\mathbf{c}_B, \qquad \alpha \in [0,1]. \tag{1}$$

All eight structures sampled along the interpolation path satisfied the periodic-wrap validity criterion, while their pore arrangement and surface morphology changed gradually with α [Fig. 5(a)]. The corresponding surrogate-predicted stress–strain responses also varied continuously, although nonlinearly, along the interpolation path [Fig. 5(b,c)]. The peak stress increased markedly toward Design B, whereas the terminal stress at $\varepsilon = 0.30$ changed more moderately. To maintain $\rho = 0.1$, the shell thickness was recalculated from the midsurface area of each decoded structure and gradually decreased from approximately 0.425 to 0.263 mm. Thus, the learned coefficient space connected the two valid geometries through a continuous structural path.

The continuous response representation enabled conditioning at alternative strain-sampling locations without retraining, while maintaining inverse-design performance comparable to that obtained on the fixed training grid. We evaluated the trained Stress-conditioned CFM ($L = 4$, 4×256) on 20 target responses, including 17 held-out validation responses and three additionally designed responses. Each target was sampled at both off-grid strain locations and the fixed 11-point dataset grid and projected onto

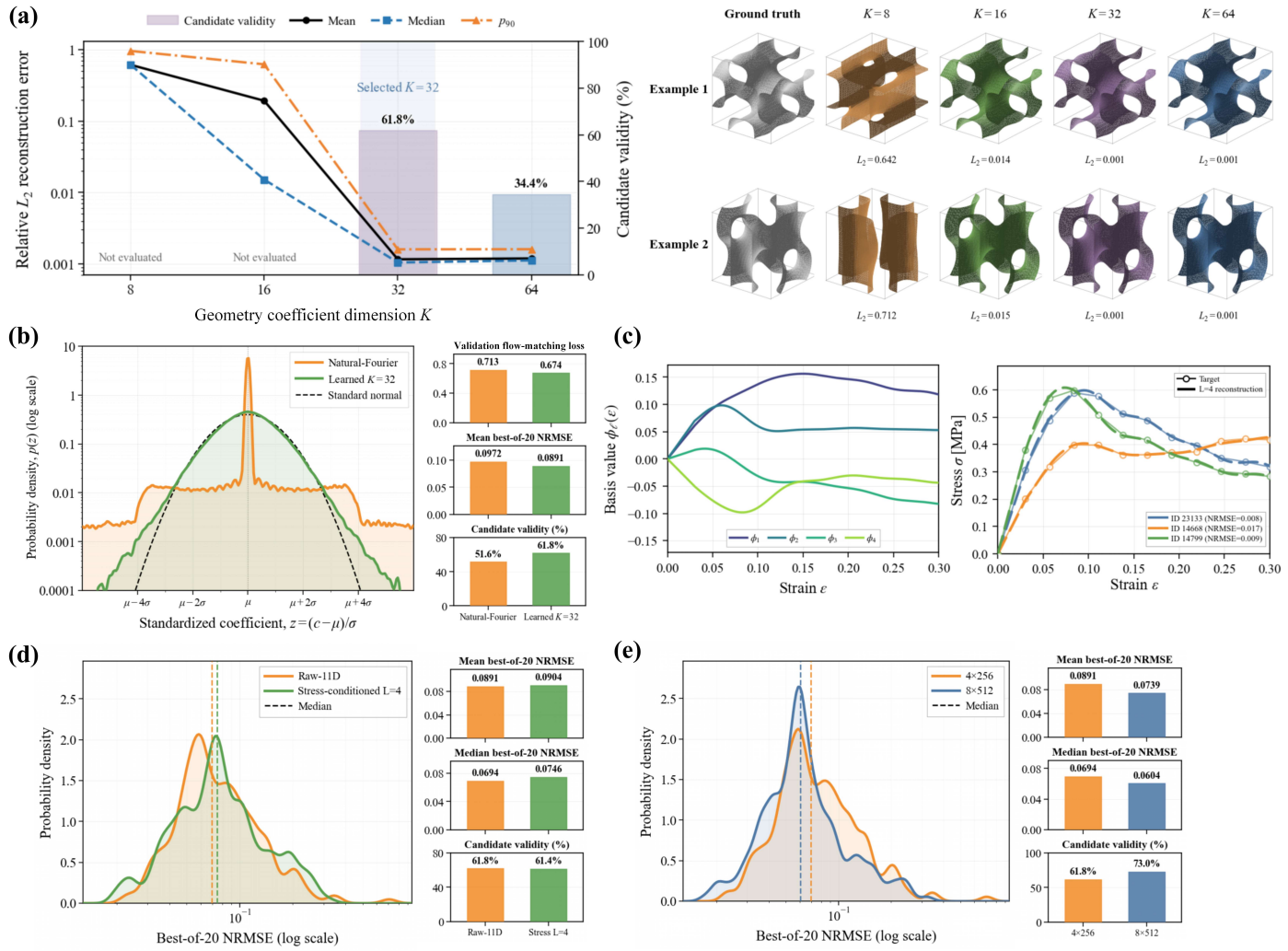


Fig. 4 Effects of representation and CFM model configuration. **(a)** Geometry reconstruction across coefficient dimensions $K = 8$, 16, 32, and 64. **(b)** Standardized Natural–Fourier and learned $K = 32$ geometry-coefficient distributions and their matched CFM results. **(c)** Four learned response basis functions and representative reconstructions using the continuous $L = 4$ response representation. **(d)** Comparison of raw 11-point and continuous $L = 4$ conditioning at the same 4×256 CFM capacity. **(e)** Comparison of 4×256 and 8×512 CFM architectures using the learned $K = 32$ geometry representation and raw 11-point conditioning.

the same learned $L = 4$ response basis, with 20 candidates generated under each condition. The two conditioning schemes produced similar results [Fig. 5(d)]. The mean best-of-20 NRMSE values were 0.181 and 0.182 for off-grid and fixed-grid conditioning, respectively, while candidate validity was 51.0% and 53.0%. Each scheme achieved the lower NRMSE for 10 of the 20 targets, indicating no systematic performance difference associated with changing the strain-sampling locations. Figure 5(e) also presents representative results for additionally designed and held-out validation targets spanning different response types, including softening and plateau–hardening behavior. Across these cases, off-grid conditioning generated valid candidates whose NRMSE values were comparable to or lower than those obtained with fixed-grid conditioning.

3 Discussion

From a methodological perspective, the proposed framework extends basis-to-basis operator learning from deterministic coefficient mappings to one-to-many generative mappings. The target response and shell geometry are represented separately in their respective function spaces. Deterministic regression cannot capture the relationship between their coefficients, because distinct geometries can produce similar responses. We therefore model this relationship as a response-conditioned distribution, with CFM acting as the generative operator between the two coefficient spaces. In principle, the same factorization could also accommodate other conditional generative models. The benchmark and ablation results support the effectiveness of the proposed continuous-latent CFM framework and show that the learned geometry representation contributes directly to its performance. Under the shared evaluation protocol, the complete framework achieved lower best-of-20 error and higher candidate validity than DiffuMeta. The two pipelines differ, however, in geometry representation, generator architecture, and decoding procedure, so this comparison should be interpreted at the frame-

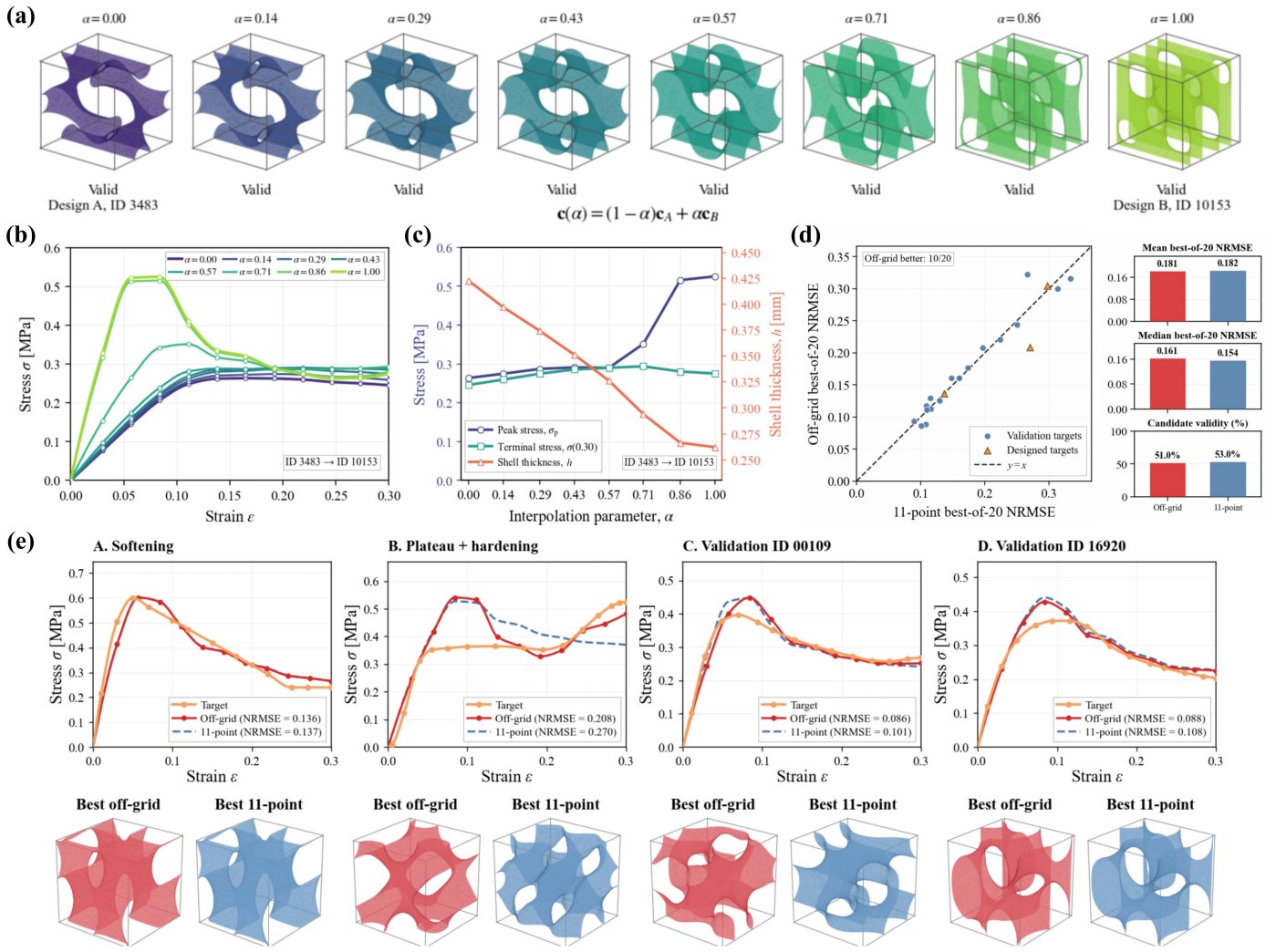


Fig. 5 Continuous geometry interpolation and off-grid response conditioning. **(a)** Linear interpolation between Design A (ID 3483) and Design B (ID 10153) in the learned geometry coefficient space. **(b)** Surrogate-predicted stress–strain responses along the sampled interpolation path. **(c)** Peak stress, terminal stress at $\varepsilon = 0.30$, and shell thickness along the interpolation path. **(d)** Paired aggregate comparison of off-grid and fixed-grid conditioning over 20 target responses. **(e)** Selected examples of target responses, surrogate predictions, and Candidate 1 geometries obtained using off-grid and fixed-grid conditioning.

work level rather than as a generator-only comparison between CFM and diffusion.

The learned geometry representation provides benefits beyond reducing the coefficient dimension. Natural–Fourier coefficients are sparse by construction, because each symbolic geometry activates only a small subset of predefined trigonometric terms. By contrast, the learned basis functions are dense and periodic across the unit cell, allowing geometric information to be distributed more broadly across the coefficient dimensions. The resulting latent distribution is smoother and more Gaussian-like than the sharply concentrated, heavy-tailed Natural–Fourier coefficient distribution. CFM transports samples from a Gaussian prior, so this reduced distributional mismatch may simplify the conditional flow. The lower validation flow-matching loss and improved inverse-design performance are consistent with this interpretation. This benefit appears to arise from the learned coordinate system rather than from simple linear post-processing. Increasing the latent dimension to $K = 64$ did not improve reconstruction but reduced candidate validity. Whitening and field-based PCA did not reduce the flow-matching loss, while the Matryoshka representation had a lower loss without improving generation performance. These results show that reconstruction accuracy and flow-matching loss alone do not identify a useful generative representation. Validity of the decoded structures must be measured as well.

The stress function encoder is useful because the 11 sampled stress values are discrete observations of one smooth response curve and are therefore correlated. The response family can therefore be viewed as lying on an approximately low-dimensional manifold parameterized by a few correlated mechanical descriptors, such as initial stiffness, peak stress, and plateau stress. The learned $L = 4$ basis provides coordinates for this manifold. More importantly, the basis functions are continuous in strain, so responses sampled at different strain locations project

into the same coefficient space rather than being tied to a fixed 11-point grid. This property may be useful in multi-objective inverse design, where stress curves and other target properties could be combined within a common conditioning representation. The appropriate embedding dimension nevertheless depends on the complexity of the response family, because an overly compact basis may smooth localized features that are not captured by the dominant manifold coordinates. Compact response conditioning did not eliminate the need for sufficient velocity-network capacity. Within the tested configurations, velocity-network capacity had a greater influence on inverse-design performance than reducing the condition dimension. The inverse mapping remains one-to-many under either response representation, so a lower-dimensional condition does not necessarily reduce the complexity of the response-conditioned geometry distribution.

The difference between surrogate and FEA evaluation is expected because nonlinear shell responses are sensitive to local geometry. Small changes in curvature or level-set topology can affect buckling, contact, and the post-buckling response, even when the corresponding low-dimensional coefficients are similar. The 3D-CNN achieved lower validation error than the coefficient-space MLP, which may reflect its direct access to local spatial information in the reconstructed field. Nevertheless, the surrogate remains an approximation of the FEA response and is therefore used to construct a compact candidate shortlist, while FEA evaluates the shortlisted designs. Candidate 1 is the predefined surrogate-ranked output, whereas the FEA-best candidate indicates that a better design may exist within the same shortlist.

Several limitations define the scope of the present study. The geometries are restricted to the symbolic family represented in the DiffuMeta dataset, and the held-out targets are drawn from the same dataset distribution as the training samples. The physical setting is also limited to a fixed material model, relative density, loading direction, and strain range. In addition, the periodic-wrap criterion verifies periodic connectivity but does not guarantee manufacturability, insensitivity to geometric imperfections, or experimental feasibility, and we did not fabricate or test any generated design. The benchmark is primarily surrogate-based, with large-scale FEA applied only to the shortlist generated by the primary model. A controlled comparison between diffusion and CFM using the same representation and model capacity also remains to be performed. Future work should therefore examine out-of-distribution targets, multi-property conditioning, manufacturing-aware constraints, and more efficient selection of candidates for FEA validation.

Overall, the proposed framework provides a strategy for generative inverse design in function space. The target response and shell geometry are encoded separately in their respective function spaces, while CFM learns their non-unique relationship as a response-conditioned distribution in coefficient space. The surrogate model and high-fidelity simulation then serve complementary roles in candidate screening and high-fidelity numerical validation. By separating representation, CFM-based generation, and validation into modular components, the framework can be adapted to other inverse-design problems that map a target physical response function to a family of feasible design fields.

4 Methods

4.1 Dataset and inverse-design task

This study uses the publicly available DiffuMeta dataset[18], which contains 23,534 periodic shell unit cells. The ith design is represented by a periodic implicit scalar field $\Psi_i(x,y,z)$ defined over the normalized unit cell $\Omega = [0,1]^3$, with the zero level set $\Psi_i = 0$ defining the shell midsurface. The original symbolic expressions were evaluated on a uniform $64 \times 64 \times 64$ grid over Ω to obtain the discrete geometry fields used in this work. The mechanical response of each design is represented by its nominal stress–strain curve under uniaxial compression along the z direction. Each response vector $\boldsymbol{\sigma}_i \in \mathbb{R}^{11}$ contains nominal stress values at 11 equally spaced strain levels from 0.03 to 0.30. All unit cells have a prescribed relative density of $\rho = 0.1$ and are evaluated up to a maximum nominal compressive strain of 0.30.

The dataset was divided into training and validation subsets. From the validation subset, 200 structures that were also excluded from the released DiffuMeta training set were selected as the common doubly held-out evaluation targets. The complete partitioning and target-selection procedure is provided in Supplementary Information S1.

Let $\mathscr{G}$ denote the forward mechanical mapping from a periodic geometry field to its discrete stress response. For a prescribed target $\boldsymbol{\sigma}^* \in \mathbb{R}^{11}$, the inverse-design task is to generate candidate fields $\{\widehat{\Psi}^{(j)}\}_{j=1}^{N}$ satisfying

$$\mathscr{G}\left(\widehat{\Psi}^{(j)}\right) \approx \boldsymbol{\sigma}^*, \qquad j = 1,\ldots,N, \tag{2}$$

while requiring each decoded zero level set to satisfy the prescribed geometric-validity criterion. Geometrically distinct structures can exhibit similar mechanical responses, so the inverse mapping is non-unique and is formulated as a conditional generative problem.

4.2 Continuous representations of geometry and stress response

Both shell geometry and stress–strain response are treated as functions defined over continuous coordinate domains. Following the function-encoder formulation[37,43], functions within each space are represented using shared learned basis functions. For a function f_i sampled at coordinates $\{x_m\}_{m=1}^{M}$, the reconstruction and ridge-regularized coefficient projection are

$$\begin{aligned} \widehat{f}_i(x) &= \sum_{k=1}^{K} a_{ik} g_k(x), \\ \mathbf{a}_i &= \underset{\mathbf{a}\in\mathbb{R}^K}{\arg\min} \left[\|\Phi\mathbf{a} - \mathbf{f}_i\|_2^2 + \lambda \|\mathbf{a}\|_2^2 \right]. \end{aligned} \tag{3}$$

Here, $\Phi_{mk} = g_k(x_m)$ and $\mathbf{f}_i = [f_i(x_1),\ldots,f_i(x_M)]^{\mathrm{T}}$. The learned basis functions span a finite-dimensional subspace of the Hilbert space of square-integrable functions on each domain, and $\mathbf{a}_i$ collects the coordinates of f_i in this subspace[37]. The projection is differentiable with respect to the basis-network parameters, so the shared basis functions are learned by minimizing reconstruction error. After basis training, the projected coefficient vectors

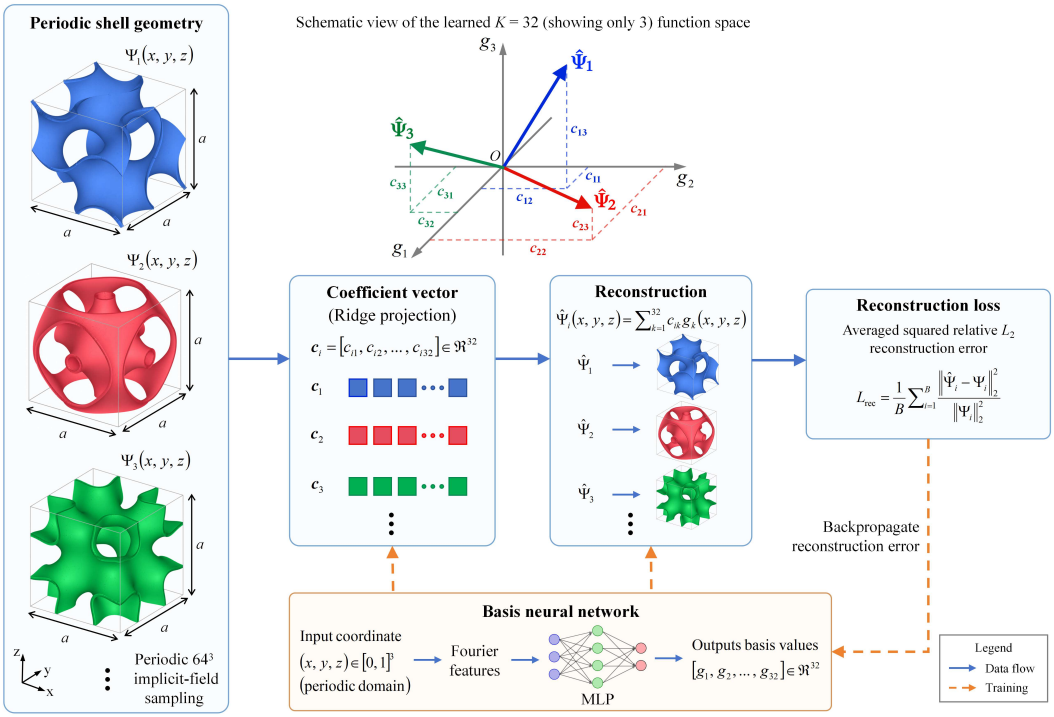


Fig. 6 Geometry function encoder and learned function-space representation. Periodic geometry fields are represented by coefficient vectors over a shared learned basis, with distinct coefficient vectors corresponding to distinct shell geometries.

are standardized dimension-wise using training-set statistics. Basis architectures, reconstruction objectives, optimization settings, and checkpoint selection are provided in Supplementary Information S1.2.

4.2.1 Continuous geometry representation

The geometry basis network maps spatial coordinates $\mathbf{x} = (x,y,z) \in [0,1]^3$ through fixed integer-frequency Fourier features. The ith geometry field is represented as

$$\widehat{\Psi}_i(\mathbf{x}) = \sum_{k=1}^{K} c_{ik} g_k(\mathbf{x}), \qquad \mathbf{c}_i \in \mathbb{R}^K. \tag{4}$$

Geometry coefficients are obtained by ridge projection onto the shared learned basis. During basis training, the reconstructed fields are compared with the original fields, and the reconstruction error is backpropagated through the differentiable projection to update the basis network. After training, the frozen basis is evaluated on the 64^3 grid for coefficient projection. The primary inverse-design model uses $K = 32$. Generated standardized coefficients are transformed back to their original scale before reconstruction of the implicit field. The geometry function encoder representation and training procedure are shown in Fig. 6.

4.2.2 Continuous stress-response representation

The primary model uses the raw response vector $\boldsymbol{\sigma}_i \in \mathbb{R}^{11}$ directly as its conditioning input. As a compact continuous alternative, the stress function encoder represents the response using four learned basis functions:

$$\widehat{\boldsymbol{\sigma}}_i(\varepsilon) = \sum_{\ell=1}^{4} \boldsymbol{\beta}_{i\ell} \varphi_\ell(\varepsilon), \qquad \boldsymbol{\beta}_i \in \mathbb{R}^4. \tag{5}$$

The coefficients are obtained through the same ridge-regularized projection and standardized using training-set statistics. The basis functions $\varphi_\ell(\varepsilon)$ can be evaluated at arbitrary strain coordinates within the studied range. Target responses sampled at the fixed 11-point grid or at off-grid locations therefore project into the same four-dimensional condition space.

4.3 Conditional flow matching (CFM)

CFM models the response-conditioned distribution of geometry coefficients through a learned time-dependent velocity field. Let $\mathbf{c}_1 \in \mathbb{R}^K$ denote a standardized geometry coefficient vector and $\mathbf{q}$ its paired response condition. The condition is obtained either from the raw response vector $\boldsymbol{\sigma} \in \mathbb{R}^{11}$ or from the continuous response coefficients $\boldsymbol{\beta} \in \mathbb{R}^4$.

Following a linear conditional flow-matching formulation [44,47], an independent Gaussian prior sample and a training geometry coefficient are connected through

$$\mathbf{c}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_K), \qquad t \sim \mathcal{U}(0,1), \qquad \mathbf{c}_t = (1-t)\mathbf{c}_0 + t\mathbf{c}_1. \tag{6}$$

The corresponding target velocity is $\mathbf{c}_1 - \mathbf{c}_0$. The conditional velocity network v_θ is trained using

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{(\mathbf{c}_1,\mathbf{q})\sim p_{\mathrm{data}},\mathbf{c}_0\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K),t\sim\mathcal{U}(0,1)} \left[\left\| v_\theta(\mathbf{c}_t, t \mid \mathbf{q}) - (\mathbf{c}_1 - \mathbf{c}_0) \right\|_2^2 \right]. \tag{7}$$

For generation, a new sample is drawn from the Gaussian prior and transported from $t = 0$ to $t = 1$ by integrating

$$\frac{\mathrm{d}\mathbf{c}_t}{\mathrm{d}t} = \widetilde{v}_\theta\left(\mathbf{c}_t, t \mid \mathbf{q}^*\right), \tag{8}$$

where $\mathbf{q}^*$ is the standardized target condition and $\widetilde{v}_\theta$ is the classifier-free-guided velocity field [61]. For the learned-geometry models, the terminal coefficient vector is transformed back to its original scale and decoded through Eq. (4). The Natural–Fourier CFM baseline applies the same flow-matching formulation to the standardized 33-dimensional coefficient vector of the original symbolic Fourier representation and reconstructs the field using the corresponding fixed Fourier terms. Velocity-network architectures, classifier-free guidance, optimization settings, and numerical integration details are provided in Supplementary Information S1.3.

4.4 Candidate generation, validity assessment, and surrogate ranking

For each target response, each benchmark method generated 20 candidates. For the CFM models, 20 independent samples were drawn from the standard Gaussian prior and transported through the learned conditional ODE. The terminal coefficient vectors were transformed back to their original scales and decoded using the corresponding learned-basis or Natural–Fourier representation. DiffuMeta candidates were generated using the released pretrained diffusion-transformer checkpoint and decoded from the sampled algebraic expressions [18]. All decoded candidates were evaluated as implicit fields on the same 64^3 grid.

Candidate validity was assessed using the periodic-wrap criterion. Each reconstructed implicit field was periodically wrapped before extraction of its zero-level surface, and the candidate was classified as valid when the extracted surface formed a single connected component. An alternative evaluation using the DiffuMeta-style procedure is reported in Supplementary Information S3.3.

Each valid candidate was evaluated using the same frozen 3D-CNN surrogate, which maps the reconstructed 64^3 implicit field to an 11-point stress response. The surrogate was trained independently of the generative models and applied without modification to all candidate sets. Candidates were ranked by their surrogate-predicted NRMSE relative to the prescribed target response. The surrogate architecture and training procedure are provided in Supplementary Information S1.4, and the cross-surrogate assessment is reported in Supplementary Information S3.2.

For each target with at least one valid candidate, the valid candidates were ranked in ascending order of surrogate-predicted NRMSE. The top-ranked candidate was defined as Candidate 1 and retained as the predefined surrogate-selected inverse-design output. For the primary Raw-11D CFM (8×512) model, a predefined shortlist containing up to the five highest-ranked valid candidates was submitted to FEA. If fewer than five valid candidates were available, all valid candidates were retained. FEA-best was defined as the completed candidate with the lowest FEA NRMSE within this fixed shortlist. It represents the lowest high-fidelity error identified within the predefined shortlist and does not replace Candidate 1 as the surrogate-selected inverse-design output.

4.5 FEA validation

The predefined candidate shortlists generated by the primary Raw-11D CFM (8×512) model were evaluated using the FEA protocol adopted in the DiffuMeta reference simulations [18]. For each candidate, the zero-level midsurface was extracted from the reconstructed 64^3 implicit field and scaled to a cubic unit cell with side length $a = 10$ mm. The surface was remeshed and discretized using three-node reduced-integration shell elements (S3R), and a uniform shell thickness was assigned to maintain the prescribed relative density of $\rho = 0.1$. The shell material followed the same isotropic elastoplastic description used in the reference simulations. Each unit cell was compressed between two rigid plates under lateral periodic boundary conditions, with general contact included. The analyses were performed in Abaqus/Standard using an implicit dynamic procedure to a maximum nominal compressive strain of 0.30. Nominal stress and strain were calculated from the reaction force and displacement of the upper plate, and the resulting responses were interpolated onto the common 11-point strain grid. Surface processing, meshing, thickness assignment, periodic constraints, contact settings, response extraction, and numerical reproducibility are detailed in Supplementary Information S2.

4.6 Evaluation metrics and statistical analysis

Response-matching accuracy was quantified using the normalized root-mean-square error (NRMSE). For a predicted response $\widehat{\boldsymbol{\sigma}} \in \mathbb{R}^J$ and a target response $\boldsymbol{\sigma}^* \in \mathbb{R}^J$,

$$\mathrm{NRMSE}\left(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\sigma}^*\right) = \frac{\left\|\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}^*\right\|_2}{\left\|\boldsymbol{\sigma}^*\right\|_2}. \tag{9}$$

The primary evaluations use $J = 11$, and the same metric is applied to surrogate-predicted and FEA responses.

For target i and method m, let $\mathscr{V}_i^{(m)}$ denote the indices of the valid candidates among the 20 generated samples. When $\mathscr{V}_i^{(m)}$ is nonempty, the best-of-20 error is defined as

$$e_i^{(m)} = \min_{j \in \mathscr{V}_i^{(m)}} \mathrm{NRMSE}\left(\widehat{\boldsymbol{\sigma}}_{ij}^{(m)}, \boldsymbol{\sigma}_i^*\right). \tag{10}$$

The per-target errors are summarized over evaluable targets using their mean, median, and 90th percentile, denoted p_{90}. If all 20 candidates generated for a target are invalid, the corresponding best-of-20 error is treated as missing. The effective target count, n_{eff}, is defined as the number of targets for which at least one valid candidate is generated. Candidate validity is calculated as the fraction of all generated candidates that satisfy the periodic-wrap criterion.

Target-level paired comparisons between the primary Raw-11D CFM (8×512) model and the alternative methods are performed over common evaluable targets. The paired-difference definition, bootstrap confidence-interval procedure, and win-fraction calculation are provided in Supplementary Information S3.4.

Geometry-representation accuracy is quantified using the rel-

ative L_2 reconstruction error on held-out geometry fields, and validation flow-matching loss is obtained by evaluating Eq. (7) on held-out samples. Fixed-grid and off-grid conditioning are compared on the same underlying target curves using equal candidate budgets and target-level pairing. Multi-seed robustness, cross-surrogate evaluation, alternative-validity assessment, paired benchmark comparisons, and geometry-dimension sensitivity are reported in Supplementary Information S3. Additional geometry-representation diagnostics are provided in Supplementary Information S4.

Data availability

The DiffuMeta dataset analysed in this work is publicly available from the ETH Zurich Research Collection under DOI 10.3929/ethz-b-000746797 [62]. The gridded fields, generated candidate sets, and per-target predictions used for the comparisons reported here are available from the corresponding authors on reasonable request.

Code availability

Code to reproduce the geometry and stress function encoders, the conditional flow matching generator, the forward surrogates, and all reported experiments, together with the trained weights for the primary model, is available at `https://github.com/kks32/2026-cfm-inverse-metamaterials` .

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. 2339678 and 2321040. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Additional support was provided by the Gulf Research Program of the National Academies of Sciences, Engineering, and Medicine through the Gulf Futures Challenge Award.

Author contributions

K.K. conceived the study and developed the method and software. S.Z. carried out the finite-element simulations and mesh generation. K.K. and S.Z. wrote the manuscript. Y.Z. contributed to the analysis and interpretation of results. K.K. and Y.Z. supervised the work. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

[1] Bertoldi, K., Vitelli, V., Christensen, J. & van Hecke, M. Flexible mechanical metamaterials. *Nature Reviews Materials* **2**, 17066 (2017).

[2] Jiao, P., Mueller, J., Raney, J. R., Zheng, X. & Alavi, A. H. Mechanical metamaterials and beyond. *Nature Communications* **14**, 6004 (2023).

[3] Lee, D., Chen, W. W., Wang, L., Chan, Y.-C. & Chen, W. Data-driven design for metamaterials and multiscale systems: A review. *Advanced Materials* **36**, 2305254 (2024).

[4] Feng, J., Fu, J., Shang, C., Lin, Z. & Li, B. Porous scaffold design by solid T-splines and triply periodic minimal surfaces. *Computer Methods in Applied Mechanics and Engineering* **336**, 333–352 (2018).

[5] Feng, J., Fu, J., Yao, X. & He, Y. Triply periodic minimal surface (TPMS) porous structures: from multi-scale design, precise additive manufacturing to multidisciplinary applications. *International Journal of Extreme Manufacturing* **4**, 022001 (2022).

[6] Berger, J. B., Wadley, H. N. G. & McMeeking, R. M. Mechanical metamaterials at the theoretical limit of isotropic elastic stiffness. *Nature* **543**, 533–537 (2017).

[7] Tancogne-Dejean, T., Diamantopoulou, M., Gorji, M. B., Bonatti, C. & Mohr, D. 3D plate-lattices: An emerging class of low-density metamaterial exhibiting optimal isotropic stiffness. *Advanced Materials* **30**, 1803334 (2018).

[8] Bonatti, C. & Mohr, D. Mechanical performance of additively-manufactured anisotropic and isotropic smooth shell-lattice materials: Simulations and experiments. *Journal of the Mechanics and Physics of Solids* **122**, 1–26 (2019).

[9] Kochmann, D. M. & Bertoldi, K. Exploiting microstructural instabilities in solids and structures: From metamaterials to structural transitions. *Applied Mechanics Reviews* **69**, 050801 (2017).

[10] Derveni, F., Gross, A. J., Peterman, K. D. & Gerasimidis, S. Postbuckling behavior and imperfection sensitivity of elastic–plastic periodic plate-lattice materials. *Extreme Mechanics Letters* **50**, 101510 (2022).

[11] Groh, R. M. J. & Pirrera, A. Probing the stability landscape of cylindrical shells for buckling knockdown factors. *Philosophical Transactions of the Royal Society A* **381**, 20220032 (2023).

[12] Jiang, Y. & Wang, Q. Highly-stretchable 3D-architected mechanical metamaterials. *Scientific Reports* **6**, 34147 (2016).

[13] Fang, X. *et al.* Programmable gear-based mechanical metamaterials. *Nature Materials* **21**, 869–876 (2022).

[14] Zhang, S., Xue, J., Yu, X. & Zhang, Y. Development of architected materials for external breast prostheses. *Biomedical Engineering Advances* **10**, 100168 (2025).

[15] Ahmad, S. Z. *et al.* Recent developments in novel TPMS lattice materials: Design optimization, performance control, and applications in biomimetic scaffolds. *Materials* **18**, 5209 (2025).

[16] Zhang, L. *et al.* Energy absorption characteristics of metallic triply periodic minimal surface sheet structures under compressive loading. *Additive Manufacturing* **23**, 505–515 (2018).

[17] Bastek, J.-H. & Kochmann, D. M. Inverse design of nonlinear mechanical metamaterials via video denoising diffusion models. *Nature Machine Intelligence* **5**, 1466–1475 (2023).

[18] Zheng, L., Kumar, S. & Kochmann, D. M. Algebraic language models for inverse design of metamaterials via diffu-

sion transformers. *Nature Machine Intelligence* **8**, 628–640 (2026). ArXiv:2507.15753.

[19] Regenwetter, L., Heyrani Nobari, A. & Ahmed, F. Deep generative models in engineering design: A review. *Journal of Mechanical Design* **144**, 071704 (2022).

[20] Sigmund, O. & Maute, K. Topology optimization approaches: A comparative review. *Structural and Multidisciplinary Optimization* **48**, 1031–1055 (2013).

[21] Dunning, P. D. Stability constraints for geometrically nonlinear topology optimization. *Structural and Multidisciplinary Optimization* **66**, 253 (2023).

[22] Bastek, J.-H., Kumar, S., Telgen, B., Glaesener, R. N. & Kochmann, D. M. Inverting the structure–property map of truss metamaterials by deep learning. *Proceedings of the National Academy of Sciences* **119**, e2111505119 (2022).

[23] Kumar, S., Tan, S., Zheng, L. & Kochmann, D. M. Inverse-designed spinodoid metamaterials. *npj Computational Materials* **6**, 73 (2020).

[24] Wang, L. *et al.* Deep generative modeling for mechanistic-based learning and design of metamaterial systems. *Computer Methods in Applied Mechanics and Engineering* **372**, 113377 (2020).

[25] Wang, Y., Zeng, Q., Wang, J., Li, Y. & Fang, D. Inverse design of shell-based mechanical metamaterial with customized loading curves based on machine learning and genetic algorithm. *Computer Methods in Applied Mechanics and Engineering* **401**, 115571 (2022).

[26] Deng, B. *et al.* Inverse design of mechanical metamaterials with target nonlinear response via a neural accelerated evolution strategy. *Advanced Materials* **34**, 2206238 (2022).

[27] Ha, C. S. *et al.* Rapid inverse design of metamaterials based on prescribed mechanical behavior through machine learning. *Nature Communications* **14**, 5765 (2023).

[28] Mao, Y., He, Q. & Zhao, X. Designing complex architectured materials with generative adversarial networks. *Science Advances* **6**, eaaz4169 (2020).

[29] Zheng, L., Karapiperis, K., Kumar, S. & Kochmann, D. M. Unifying the design space and optimizing linear and nonlinear truss metamaterials by generative modeling. *Nature Communications* **14**, 7563 (2023).

[30] Vlassis, N. N. & Sun, W. Denoising diffusion algorithm for inverse design of microstructures with fine-tuned nonlinear material properties. *Computer Methods in Applied Mechanics and Engineering* **413**, 116126 (2023).

[31] Bastek, J.-H., Sun, W. & Kochmann, D. M. Physics-informed diffusion models. In *International Conference on Learning Representations (ICLR)* (2025). ArXiv:2403.14404.

[32] Abu-Mualla, M. & Huang, J. Inverse design of curved mechanical metamaterials with geometric AI: a generative diffusion operates in compact latent space of cellular structures. *npj Metamaterials* **1**, 5 (2025).

[33] Maurizi, M. *et al.* Designing metamaterials with programmable nonlinear responses and geometric constraints in graph space. *Nature Machine Intelligence* **7**, 1023–1036 (2025).

[34] Liu, Q., Koric, S., Abueidda, D., Meidani, H. & Geubelle, P. Toward signed distance function based metamaterial design: Neural operator transformer for forward prediction and diffusion model for inverse design. *Computer Methods in Applied Mechanics and Engineering* **446**, 118316 (2025). ArXiv:2504.01195.

[35] Park, J. J., Florence, P., Straub, J., Newcombe, R. & Lovegrove, S. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 165–174 (2019).

[36] Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S. & Geiger, A. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4460–4470 (2019). ArXiv:1812.03828.

[37] Ingebrand, T., Thorpe, A. J. & Topcu, U. Function encoders: A principled approach to transfer learning in Hilbert spaces. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, vol. 267 of *Proceedings of Machine Learning Research*, 26464–26489 (2025). ArXiv:2501.18373.

[38] Dong, S. & Ni, N. A method for representing periodic functions and enforcing exactly periodic boundary conditions with deep neural networks. *Journal of Computational Physics* **435**, 110242 (2021).

[39] Tancik, M. *et al.* Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)* (2020). ArXiv:2006.10739.

[40] Lu, L., Jin, P., Pang, G., Zhang, Z. & Karniadakis, G. E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence* **3**, 218–229 (2021).

[41] Li, Z. *et al.* Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)* (2021). ArXiv:2010.08895.

[42] Kovachki, N. *et al.* Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research* **24**, 1–97 (2023).

[43] Ingebrand, T., Thorpe, A. J., Goswami, S., Kumar, K. & Topcu, U. Basis-to-basis operator learning using function encoders. *Computer Methods in Applied Mechanics and Engineering* **435**, 117646 (2025). ArXiv:2410.00171.

[44] Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M. & Le, M. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)* (2023). ArXiv:2210.02747.

[45] Liu, X., Gong, C. & Liu, Q. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)* (2023). ArXiv:2209.03003.

[46] Albergo, M. S., Boffi, N. M. & Vanden-Eijnden, E. Stochastic interpolants: A unifying framework for flows and dif-

fusions. *Journal of Machine Learning Research* **26**, 1–80 (2025). ArXiv:2303.08797.

[47] Tong, A. *et al.* Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research (TMLR)* (2024). ArXiv:2302.00482.

[48] Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)* (2020). ArXiv:2006.11239.

[49] Song, Y. *et al.* Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)* (2021). ArXiv:2011.13456.

[50] Miller, B. K., Chen, R. T. Q., Sriram, A. & Wood, B. M. FlowMM: Generating materials with riemannian flow matching. In *International Conference on Machine Learning (ICML)* (2024). ArXiv:2406.04713.

[51] Sriram, A., Miller, B. K., Chen, R. T. Q. & Wood, B. M. FlowLLM: Flow matching for material generation with large language models as base distributions. In *Advances in Neural Information Processing Systems (NeurIPS)* (2024). ArXiv:2410.23405.

[52] Luo, X. *et al.* CrystalFlow: a flow-based generative model for crystalline materials. *Nature Communications* **16**, 9267 (2025).

[53] Chang, J.-H. R. *et al.* 3D shape tokenization via latent flow matching. *arXiv preprint arXiv:2412.15618* (2024).

[54] Zhang, J. *et al.* FuncGenFoil: Airfoil generation and editing model in function space. *arXiv preprint arXiv:2502.10712* (2025).

[55] van Delden, J. *et al.* Minimizing structural vibrations via guided flow matching design optimization. *arXiv preprint arXiv:2506.15263* (2025).

[56] Mirzaee, H. & Kamrava, S. Inverse design of microstructures using conditional continuous normalizing flows. *Acta Materialia* **285**, 120704 (2025).

[57] Yao, M. S. *et al.* Generative adversarial model-based optimization via source critic regularization. In *Advances in Neural Information Processing Systems (NeurIPS)* (2024). ArXiv:2402.06532.

[58] Koiter, W. T. *On the Stability of Elastic Equilibrium*. Ph.D. thesis, Polytechnic Institute of Delft, Delft, The Netherlands (1945). English translation: NASA Technical Translation TT F-10833, 1967.

[59] Hutchinson, J. W. & Thompson, J. M. T. Imperfections and energy barriers in shell buckling. *International Journal of Solids and Structures* **148–149**, 157–168 (2018).

[60] Danesh, H. & Wessels, H. Data-efficient bayesian-guided design selection from large candidate sets: Application to hyperelastic stochastic metamaterials (2026). ArXiv:2603.15917.

[61] Ho, J. & Salimans, T. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598* (2022).

[62] Zheng, L., Kumar, S. & Kochmann, D. M. DiffuMeta: Algebraic language models for inverse design of metamaterials via diffusion transformers (2025). Dataset and model checkpoints.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.pdf