HASH-GUIDED DIFFUSION: EFFICIENT GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Physics-based simulation of sewing patterns yields globally correct but overly smooth garment surfaces, and recovering high-frequency detail such as wrinkles and seams with optimization- or diffusion-based refinement typically requires hundreds to thousands of iterations. We present Hash-guided Diffusion (HD), which keeps the simulated surface fixed and generates only a residual displacement field over a frozen UV parameterization, a mapping of the 3D surface onto 2D coordinate charts obtained with Nuvo, a neural UV parameterization. The residual is represented by a multiresolution hash encoding for constant-time lookup of local high-frequency detail, and is generated by a conditional diffusion process with either a discrete (DDPM) or a deterministic probability-flow ODE reverse solver. On a representative stiff garment from GarmentCodeData v2, HD-ODE with 10 reverse steps attains lower panel-level and image-level error than 500-step discrete diffusion (Panel L2 4.72 vs. 5.18 cm). Performance depends on material stiffness, with the best results on stiff fabrics. These results indicate that a geometry-aligned residual domain can substantially shorten diffusion-based surface refinement.

1 INTRODUCTION

High-quality digital garment reconstruction and synthesis lies at the core of modern virtual try-on, animation, and digital fashion pipelines. Unlike rigid objects or parametric body models (Loper et al., 2015; Zuffi et al., 2017), garments exhibit complex non-rigid behavior governed by sewing patterns, material properties, and physical interactions. As a result, high-fidelity garment modeling typically requires solving three interconnected sub-problems: (i) *pattern-level structure* recovery (panels and seams), which defines the topology and layout of garment panels in 2D; (ii) *physics-based draping* to obtain a coarse but globally consistent 3D surface; and (iii) *surface-level refinement* to recover high-frequency wrinkles, seams, and appearance details, with efficient inference enabling refinement to be deployed at scale or interactively.

Recent multimodal and foundation-model approaches have made strong progress on the first two components, enabling robust pattern prediction and simulation-conditioned 3D reconstruction. Specifically, AIpparel (Nakayama et al., 2025) brings a multimodal foundation model that jointly processes sewing patterns, materials, and simulated 3D geometry on datasets like GarmentCodeData (Korosteleva et al., 2024). However, the resulting surfaces are often overly smooth and can still fail to provide realistic high-frequency details, even though sewing-pattern-conditioned simulation is accurate. Post-processing with optimization- or diffusion-based refinement can require hundreds to thousands of iterations, limiting scalability and interactive use.

Our approach. In this paper, we focus on surface refinement and inference efficiency of sewing-pattern-based garments. We refer to our method as Hash-Guided Diffusion (HD). Instead of creating geometry from scratch, HD performs conditional refinement over a strong geometric prior supplied by sewing patterns and simulation. We use fixed geometry at the panel and stitch level, with HD modeling only continuous high-frequency residuals aligned with the surface. By restricting diffusion to a geometry-aligned residual space constructed via Nuvo (Srinivasan et al., 2023), HD achieves rich refinement with only 10 diffusion steps, thereby drastically speeding up the inference process while maintaining visual fidelity.

Key insight. The original 200–1000 steps of surface refinement optimization can be compressed into 10 steps through geometry-aligned diffusion optimization. By performing diffusion in a surface-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

attached residual field over a frozen UV coordinate system produced by Nuvo, HD significantly reduces the effective generative complexity. Nuvo provides a stable, low-distortion multi-chart UV atlas for irregular, non-rigid geometry; HD then models high-frequency residual displacements in this UV domain while keeping the coarse simulated surface fixed. This enables accurate reverse-time refinement as a per-instance optimization with a diffusion solver.

Pipeline. HD follows a two-stage procedure: (1) *Nuvo parameterization:* given a coarse garment surface, we optimize a multi-chart neural UV map and *freeze* it once chart boundaries stabilize. (2) *Residual diffusion refinement:* we learn a diffusion/ODE process over a surface-aligned residual field defined in the frozen UV domain, using a hash-accelerated representation for fast lookup and reduced spectral bias.

Contributions.

- **Surface-attached residual diffusion.** We formulate garment surface refinement as a diffusion (and probability-flow ODE) process over a residual displacement field defined in a frozen, Nuvo-produced UV atlas, explicitly separating global geometry from high-frequency detail.
- **Hash-accelerated residual space.** We introduce a lightweight, multiresolution hash indexing of the UV residual field, enabling constant-time residual queries and supporting short-step reverse diffusion.
- **Discrete and ODE reverse solvers.** We instantiate HD with both DDPM-style discrete denoising and a deterministic neural ODE interpretation, and study the quality–efficiency trade-off under different step budgets.
- **Evaluation protocol for refinement.** We evaluate refinement quality using image-level fidelity (L1/LPIPS) while reporting sewing-pattern reconstruction metrics strictly as geometry-preservation sanity checks, ensuring refinement does not degrade panel-level structure.

2 Related Work

2.1 Garment Reconstruction and Sewing Patterns

A large body of work studies garment reconstruction from images, point clouds, or simulated geometry (Korosteleva & Lee, 2022; Liu et al., 2023; Wang et al., 2024). Recent methods predict sewing patterns—including panel layouts, edge connectivity, and stitches—as an intermediate representation (Liu et al., 2023; He et al., 2024; Can et al., 2026), yielding strong topology control and improved generalization. Large-scale multimodal models such as AIpparel (Nakayama et al., 2025) further integrate pattern semantics, material parameters, and simulation signals to recover coherent garments. However, their outputs may still lack fine-scale surface detail. HD is complementary: we assume a correct coarse garment surface obtained from patterns and simulation and focus exclusively on fast, high-frequency refinement while preserving the underlying panel and stitch structure.

Closer to our setting, several works add fine detail on top of coarse garments. DeepWrinkles (Lähner et al., 2018) and detail enhancement (Zhang et al., 2021) predict wrinkles as normal maps in UV space, and recent diffusion models generate garment deformations as UV displacement maps (Vidaurre et al., 2024; Dumoulin et al., 2026) or in sewing-pattern coordinates (Li et al., 2025). These methods train a model over a dataset of garments; HD instead optimizes per instance, represents the residual with a hash grid in a learned Nuvo atlas, and targets reverse trajectories of 10 steps.

2.2 Surface Parameterization for Irregular Geometry

Classical UV unwrapping can produce highly fragmented atlases on noisy or non-manifold meshes, making downstream editing and learning unstable. Nuvo (Srinivasan et al., 2023) introduces neural UV mapping for "unruly" geometry, producing large, contiguous, and low-distortion charts by learning a continuous mapping from surface points to texture space. Its neural-field representation (Mildenhall et al., 2020) makes the parameterization robust to holes and artifacts commonly found in neural reconstructions (Peng et al., 2021). Rather than assigning UV coordinates only to

mesh vertices, Nuvo learns a continuous mapping from 3D surface points to 2D texture space using multilayer perceptrons. Consequently, it can generate large, contiguous, and low-distortion UV charts for highly irregular geometry. HD leverages Nuvo as an explicit *surface-attached coordinate system* for refinement: diffusion operates over a residual field in UV space, while the refined output remains a 3D surface.

2.3 Diffusion, Neural ODEs, and Continuous Refinement

Diffusion models originate from nonequilibrium thermodynamics (Sohl-Dickstein et al., 2015) and have become a dominant paradigm for generative modeling (Ho et al., 2020; Chen et al., 2024; Bartosh et al., 2024; Holderrieth & Erives, 2025). They have also been adapted to geometry generation and surface refinement. In the forward process, clean samples are progressively corrupted by Gaussian noise until their distribution approaches a tractable noise distribution. In the reverse process, a denoising network learns to remove this corruption and recover samples from the data distribution. A recurring limitation is inference cost, as high-quality sampling may require long denoising chains (Asmussen & Glynn, 2007). Several lines of work accelerate diffusion sampling through distillation (Salimans & Ho, 2022) and ODE-based samplers, including deterministic formulations based on probability-flow ODEs. Our approach is complementary: rather than modifying only the reverse solver, we reduce the effective complexity of the state space by performing diffusion in a geometry-aligned residual domain induced by Nuvo. This formulation enables short-step refinement while keeping the global garment surface fixed.

2.4 Hash Grids and High-Frequency Signals

Multiresolution hash encodings provide compact and efficient feature lookup for high-frequency neural fields (Müller et al., 2022). They are therefore well suited to local surface details, such as wrinkles and seams, that require high spatial bandwidth but occupy only a surface-attached residual domain. HD uses a hash grid as an efficient residual query representation in Nuvo UV space rather than as a full-scene representation. This design keeps the global garment geometry fixed while allocating representational capacity to the high-frequency residual field.

De Rebotti et al. (De Rebotti et al., 2025) introduced a GPU-native, variance-adaptive, multi-resolution voxel grid in which a flat spatial hash table provides average-case constant-time access to mixed-resolution voxel blocks. Their method targets online TSDF reconstruction rather than generative modeling. Inspired by its principle of allocating spatial resolution according to local TSDF variation, we define a distinct variance-adaptive hash representation over Nuvo-aligned garment residuals. Unlike MrHash, our representation operates on a multi-chart 2D surface domain, uses local residual variation to allocate resolution, and is coupled to conditional residual diffusion.

3 Problem Formulation

Let $\mathcal{S} \subset \mathbb{R}^3$ denote the coarse garment surface produced by sewing patterns and simulation. Nuvo provides a frozen, multi-chart parameterization that maps each surface point to a UV coordinate and chart id:

$$\pi : \mathcal{S} \to \Omega \times \{1, \dots, K\}, \qquad \pi(\mathbf{x}) = (\mathbf{u}, k), \tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is the UV domain and K is the number of charts (we use $K{=}8$).

Residual refinement field. We define a residual displacement field in UV space,

$$\mathcal{R} : \Omega \times \{1, \dots, K\} \to \mathbb{R}^3, \tag{2}$$

and write the refined surface as

$$\mathbf{x}_{\text{ref}}(\mathbf{u}, k) = \mathbf{x}_{\text{base}}(\mathbf{u}, k) + s \cdot \mathcal{R}(\mathbf{u}, k), \tag{3}$$

where $\mathbf{x}_{\text{base}}(\mathbf{u}, k)$ is the Nuvo-reparameterized coarse surface and s is a residual scale (bounded in our implementation to at most 5% of base mesh size to prevent geometric drift). The output representation is thus:

$$\mathcal{R} = \left\{ \mathcal{R} : \Omega \to \mathbb{R}^3 \;\middle|\; \mathbf{x}(u, v) = \mathbf{x}_0(u, v) + \mathcal{R}(u, v) \right\}, \tag{4}$$

where $\mathcal{R}(u, v)$ is generated by neural diffusion in 5–10 steps. The residual field $\mathcal{R}$ is aligned with the underlying surface geometry through the Nuvo parameterization.

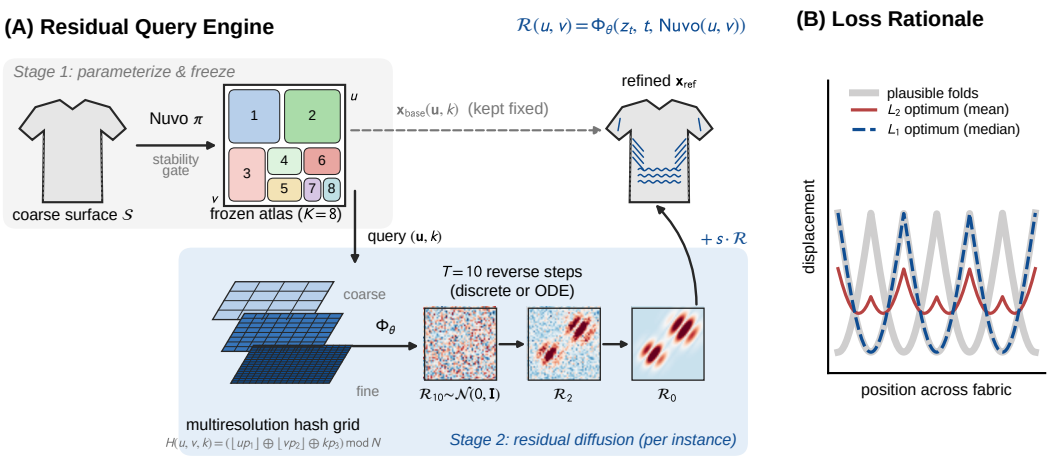

Figure 1: **HD method overview.** (**A**) Residual Query Engine: HD performs refinement by querying a time-dependent residual field in the frozen Nuvo UV space. (**B**) Loss Rationale: L_1+LPIPS preserves high-frequency details and avoids the oversmoothing induced by L_2. Panel A is a schematic: the residual maps are a synthetic wrinkle field noised with our forward process (Equation (5)) and β schedule, not model outputs. Panel B is an exact toy computation: for two plausible fold profiles (probabilities $0.6/0.4$), the pointwise L_2 minimizer is their mean (washed-out folds), whereas the L_1 minimizer is the median, which keeps the sharp crease.

Diffusion objective. Given a target residual field $\mathcal{R}_0$ (supervised by renderings and/or reference geometry), we define a forward variance-preserving diffusion process:

$$q(\mathcal{R}_t \,|\, \mathcal{R}_0) = \mathcal{N}\big(\sqrt{\bar{\alpha}_t}\mathcal{R}_0,\ (1-\bar{\alpha}_t)\mathbf{I}\big)\,, \tag{5}$$

where $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$ and $\{\alpha_t\}_{t=1}^{T}$ is a noise schedule with T steps. HD learns a reverse-time model, either (i) a discrete denoiser $\epsilon_\theta(\mathcal{R}_t, t, \mathbf{u}, k, \mathbf{c})$ that predicts the injected noise, or (ii) a continuous-time drift field $f_\theta(\mathcal{R}_t, t, \mathbf{u}, k, \mathbf{c})$ used in a probability-flow ODE solver. Here $\mathbf{c}$ denotes conditioning signals (e.g., material parameters and pattern-level descriptors).

This formulation makes the refinement task explicit: *we refine only* $\mathcal{R}$, while keeping the base geometry and the Nuvo coordinate system fixed.

4 Method

4.1 Stage 1: Nuvo Parameterization and Stability Gate

Nuvo learns a continuous mapping from surface points to a compact multi-chart UV atlas. We optimize Nuvo per garment instance until the atlas is stable, then freeze it for all subsequent refinement. **Stability is essential:** if chart boundaries drift (e.g., UV pixels switch semantic panels), the residual field becomes ill-posed and diffusion learns inconsistent coordinates. This is a risk condition: if the boundaries of the chart keep changing (for example, pixels that originally belonged to the front piece are now being moved to the sleeve), it absolutely cannot proceed to Stage 2. This is because Diffusion is trained on a fixed coordinate system; if the base changes, Diffusion will learn incorrect coordinates.

In practice, we monitor chart assignment consistency and only proceed to Stage 2 once boundary changes fall below a threshold. We additionally enable vertex duplication (`use_vertex_duplication: true`) at chart boundaries to avoid interpolation artifacts, as documented in the Nuvo repository. This prevents pixel interpolation errors at the boundaries of the 8 charts.

4.2 Stage 2: Residual Field Representation

We represent $\mathcal{R}(\mathbf{u}, k)$ with a lightweight, multiresolution hash grid in the UV domain. Each query $(\mathbf{u}, k)$ is mapped to a small set of hash table entries using the hash function:

$$H(u, v, k) = (\lfloor up_1 \rfloor \oplus \lfloor vp_2 \rfloor \oplus kp_3) \bmod N, \tag{6}$$

where p_1, p_2, p_3 are large primes and N is the hash table size. This can be viewed as a database query (Section A.3). Features are interpolated to obtain a local feature vector, which is then linearly projected to a 3D displacement. This design supports *constant-time* lookup and mitigates spectral bias when fitting high-frequency wrinkles and seam details.

4.3 Neural Diffusion on the Residual Field

Discrete diffusion (DDPM-style). We adopt the forward process in Equation (5). For discrete denoising, we train a time-conditional network to predict noise ϵ under an MSE objective:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t,\epsilon}\left[\|\epsilon - \epsilon_\theta(\mathcal{R}_t, t, \mathbf{u}, k, \mathbf{c})\|_2^2\right]. \tag{7}$$

The forward process uses a discrete variance-preserving diffusion defined over T steps. For step $t \in \{1, \ldots, T\}$, the noised residual is:

$$\mathcal{R}_t = \sqrt{\bar{\alpha}_t}\mathcal{R}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{8}$$

where

$$\bar{\alpha}_t = \prod_{i=1}^{t}(1 - \beta_i). \tag{9}$$

In our experiments, $T = 10$ and β_t follows a linear schedule: $\beta_t \in [0.1, 0.8]$. This aggressive noise schedule is required to ensure sufficient mixing when using a very small number of diffusion steps on a bounded residual field.

The reverse transition from $\mathcal{R}_t$ to $\mathcal{R}_{t-1}$ follows the standard DDPM formulation:

$$\mathcal{R}_{t-1} = \mu_\theta(\mathcal{R}_t, t) + \sigma_t z, \quad z \sim \mathcal{N}(0, I), \tag{10}$$

where the mean μ_θ is computed from the predicted noise:

$$\mu_\theta(\mathcal{R}_t, t) = \frac{1}{\sqrt{1 - \beta_t}}\left(\mathcal{R}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}}\epsilon_\theta(\mathcal{R}_t, t)\right). \tag{11}$$

Reverse diffusion is performed discretely, step by step, for exactly T iterations.

ODE diffusion. For the continuous variant, we reinterpret the reverse dynamics as a deterministic probability-flow ODE. We adopt a standard linear forward diffusion process. The reverse process is parameterized by an implicit neural network that models time-dependent denoising dynamics rather than direct regression:

$$\frac{d\mathcal{R}(t)}{dt} = f_\theta(\mathcal{R}(t), t), \tag{12}$$

where f_θ is a neural network that models the instantaneous denoising direction. The forward noising process is identical:

$$\mathcal{R}_t = \sqrt{\bar{\alpha}_t}\mathcal{R}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon. \tag{13}$$

The same noise schedule and number of steps are used. Instead of applying discrete denoising steps, the reverse process is interpreted as a continuous-time dynamical system. This formulation corresponds to the probability flow ODE associated with the discrete diffusion process.

Given an initial noisy residual $\mathcal{R}_T$, the refined residual $\mathcal{R}_0$ is obtained by numerically integrating the ODE:

$$\mathcal{R}_0 = \int_T^0 f_\theta(\mathcal{R}(t), t)\, dt, \tag{14}$$

using a fixed-step or adaptive ODE solver. In the codebase, the number of effective denoising steps is matched to the discrete setting (e.g., 10 steps). No stochastic noise is injected during reverse

integration. This yields deterministic refinement, in contrast to the stochastic reverse process used in discrete diffusion.

The same noise prediction objective is used for training:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{R_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(\mathcal{R}_t, t)\|_2^2\right]. \tag{15}$$

Thus, the difference between the two main experiments lies only in the reverse-time formulation (discrete step-by-step denoising vs. continuous ODE integration), not in the forward process or the training loss. A conceptual comparison with flow matching is given in Section C.

4.4 Photometric and Perceptual Supervision

HD is designed to refine surface appearance on top of fixed geometry. We therefore supervise refinement using differentiable rendering and combine diffusion loss with image-level reconstruction and perceptual terms:

$$\mathcal{L} = \mathcal{L}_{\text{diff}} + \lambda_1 \mathcal{L}_1 + \lambda_p \mathcal{L}_{\text{LPIPS}}. \tag{16}$$

While L_2 loss encourages mean-mode convergence, it often leads to oversmoothed surfaces in garment modeling (see Figure 1B). We employ L_1 for edge-preservation and LPIPS (Zhang et al., 2018) to capture the perceptual manifold of high-frequency folds, ensuring the refined wrinkles align with human visual expectations of fabric physics. If only L_1 is used, the outcome is not sufficient for human perception; the CNN-based classifier (LPIPS) better aligns with human visual judgment.

We use loss weights $\lambda_1 = 0.5$ and $\lambda_p = 0.5$, with diffusion weight fixed at 1.0. To prevent geometric drift, we bound the residual magnitude in Equation (3) to at most 5% of the base mesh scale.

4.5 Training and Inference

All models are optimized per garment instance. We use a two-stage schedule: first optimize Nuvo (Stage 1, iterations 0–5000), then jointly optimize the diffusion model and residual field with the frozen UV atlas (Stage 2, iterations 5000–10000). Checkpoints are saved every 500 iterations. At inference time, HD starts from Gaussian noise in the residual space and performs $T \in [5, 10]$ reverse steps to obtain $\mathcal{R}_0$, which is then added to the base surface via Equation (3). Algorithm 1 in Section A.1 summarizes the procedure.

5 Experiments

5.1 Dataset and Set Up

We evaluate HD on GarmentCodeData v2 (Korosteleva et al., 2024), containing 3,867 garment designs simulated under varying bending stiffness parameters $b \in \{1, 100, 500\}$. The dataset provides sewing patterns, material parameters, and simulated garments. We use the official recommended split: Train / Val / Test = 80% / 10% / 10% (approximately 92k / 11.5k / 11.5k samples). We report results separately for soft ($b = 1$), medium ($b = 100$), and stiff ($b = 500$) materials because deformation statistics differ substantially across regimes. Unless noted otherwise, all garments are simulated on the neutral body template (default_body).

All experiments are conducted as per-sample optimization on a single NVIDIA H200 GPU. We fix all hyperparameters across experiments and do not perform hyperparameter search, as our goal is to analyze the refinement performance and effectiveness of geometry-aligned residual diffusion under a stable and reproducible setting.

5.2 Baselines and Ablations

We compare against and/or ablate the following components:

- **Naive diffusion (no Nuvo):** diffusion directly on 3D residuals without a surface-attached UV coordinate system.

- **Nuvo + diffusion (no hash):** diffusion in Nuvo UV space using an MLP residual field (no hash acceleration).
- **HD-Discrete:** our full method with DDPM-style discrete reverse steps.
- **HD-ODE:** our full method with probability-flow ODE reverse dynamics.
- **Step budget:** $T \in \{10, 20, 50, 500\}$ to quantify the efficiency–quality trade-off.

We also compare against AIpparel (Nakayama et al., 2025), which reports Panel L2 of 6.10 cm on the same dataset.

5.3 METRICS

HD is designed for surface appearance refinement on top of a fixed and correct garment geometry obtained from sewing patterns and simulation. The method does not predict sewing-pattern structure, panel topology, or stitch connectivity. Instead, it refines a surface-aligned residual field defined over a stable Nuvo parameterization. Accordingly, our evaluation separates surface refinement quality from geometry preservation:

- **Geometry sanity checks:** panel-level metrics adopted by sewing-pattern reconstruction works (Panel L2, rotation/translation drift, panel/edge accuracy, stitch accuracy). Panel L2 (cm) measures the average Euclidean distance between predicted and ground-truth garment panels in 2D pattern space, capturing structural and layout correctness while remaining invariant to 3D folding ambiguity. These verify that refinement does not degrade the underlying structure.
- **Refinement metrics:** image-level L1 and LPIPS (Zhang et al., 2018) computed between refined and reference renderings. These are our primary metrics for surface appearance refinement quality and are measured using image-level losses and inference efficiency, which directly reflect high-frequency appearance quality.

Vertex-wise L1 measures mean absolute error over 3D mesh vertices. LPIPS measures perceptual distance between refined and reference renderings using deep features. We note that vertex-wise L1 and perceptual metrics are known to be unreliable under large non-isometric deformations and self-contact, particularly in the stiff regime where small rigid misalignments or localized sharp folds can induce relatively large L1 deviations even when the global panel structure is accurate and Panel L2 is low (see Section B.2).

5.4 IMPLEMENTATION DETAILS

We follow a two-stage optimization schedule with checkpoints. Stage 1 optimizes Nuvo and is run until UV charts stabilize (iterations 0–5000); Stage 2 optimizes HD with a frozen atlas (iterations 5000–10000). Unless noted otherwise, we use T=10 diffusion steps, a linear β schedule with $\beta \in [0.1, 0.8]$, and fixed loss weights (λ_p=0.5, λ_1=0.5). Diffusion weight is 1.0. Residual scaling is bounded to max 5% of base mesh size. Default sample for experiments: index 5 (rand_02VLBVOOHQ) from batch 31 (garments_5000_31) with stiff material ($b = 500$).

5.5 QUANTITATIVE RESULTS

Table 1 reports evaluation results on a representative sample (sample 5: rand_02VLBVOOHQ) from the stiff material regime. Table 2 shows results across material stiffness regimes; Figure 2 visualizes both tables.

Key observations. Across all evaluated settings, HD consistently produces stable panel-level reconstructions, with performance governed by both the diffusion formulation and material stiffness. Crucially, these results align with the design objective stated in the abstract: achieving high-quality surface refinement through a small number of diffusion steps by operating in a geometry-aligned residual space.

Table 1: **Main results on GarmentCodeData v2** (sample 5, stiff material b=500). Geometry metrics are sanity checks; refinement metrics are primary. HD achieves competitive Panel L2 while using only 10 diffusion steps. Increasing steps beyond 20 yields diminishing returns. best, runner-up.

Method	Panel L2↓	L1↓
AIpparel (Nakayama et al., 2025)	6.10	–
HD-Discrete (Ablation, T=10)	5.13	25.02
HD-Discrete, T=20	5.40	24.64
HD-Discrete, T=50	5.48	25.02
HD-Discrete, T=500 (naive)	5.18	24.93
HD-ODE (T=10, ours)	**4.72**	**23.22**

Table 2: **Material stiffness ablation** (T=10 diffusion steps). Panel L2 improves substantially with bending stiffness, reflecting the more regular, near-isometric deformation behaviour of stiff materials. HD on stiff garments outperforms the AIpparel baseline.

Material	Bending b	Panel L2↓	L1↓
AIpparel (Nakayama et al., 2025)	–	6.10	–
Soft	1	31.50	33.32
Medium	100	7.87	**20.08**
Stiff (ours)	500	**4.58**	23.71

Efficiency of continuous residual diffusion. Comparing diffusion formulations and step counts, the ODE-based continuous residual diffusion achieves competitive panel-level accuracy with significantly fewer refinement steps. Increasing the number of discrete diffusion steps from 20 to 500 yields only marginal improvements in Panel L2 (from 5.40 cm to 5.18 cm), indicating rapidly diminishing returns from long diffusion chains. The discrete ablation model attains a comparable Panel L2 (5.13 cm) but does not offer a clear advantage once computational efficiency is considered. These results support the central claim of HD: that the dominant surface residual structure can be captured with short, geometry-aligned diffusion trajectories, enabling effective refinement within 5–10 steps—with almost no loss of quality, using fewer diffusion steps.

Impact of material stiffness. Reconstruction accuracy improves substantially with increased bending stiffness. HD achieves markedly lower Panel L2 error on stiff garments ($b = 500$, 4.58 cm) than on medium-stiffness garments ($b = 100$, 7.87 cm) or soft garments ($b = 1$, 31.50 cm), reflecting the more regular, near-isometric deformation behavior of stiff materials. In this regime, HD outperforms the AIpparel baseline (6.10 cm), demonstrating that diffusion in a surface-attached, low-curvature residual space is particularly effective when deformations are well approximated by smooth residual fields.

6 Analysis

6.1 Energy and PDE Perspective

Cloth simulation can be formulated as minimizing the total energy:

$$E(\mathbf{x}) = E_{\text{stretch}}(\mathbf{x}) + b\, E_{\text{bend}}(\mathbf{x}) + E_{\text{contact}}(\mathbf{x}), \tag{17}$$

where b controls bending stiffness. For **soft materials** ($b \approx 1$): E_{contact} dominates, resulting in a highly non-convex energy landscape with multiple valid local minima. For **stiff materials** ($b \gg 1$): E_{bend} dominates, deformations are close to near-isometric, and the energy landscape is smoother and more structured.

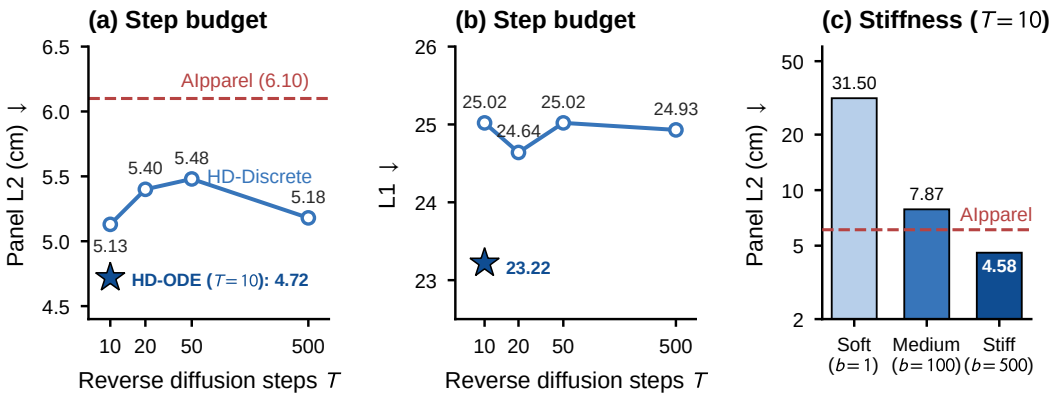

Figure 2: **Step budget and material stiffness.** (a,b) Panel L2 and L1 of HD-Discrete as the number of reverse steps grows from 10 to 500, together with HD-ODE at T=10 (star); values from Table 1. (c) Panel L2 across bending stiffness at T=10 (log scale); values from Table 2. The dashed line marks the AIpparel Panel L2 of 6.10 cm.

In the stiff regime, garment deformation is well approximated by a fourth-order bending PDE (e.g., thin-shell models), where solutions are smooth, globally constrained, and stable under small perturbations. HD performs iterative refinement over surface geometry, which can be interpreted as learning a diffusion-like operator:

$$\frac{\partial \mathbf{x}}{\partial t} \approx -\nabla_{\mathbf{x}} \mathcal{E}_{\text{geom}}, \tag{18}$$

which favors: (i) smooth normals, (ii) coherent curvature, and (iii) global structural consistency. Such behavior closely matches the optimization dynamics of bending-dominated cloth, but is fundamentally misaligned with contact-driven folding in soft materials. HD's diffusion process naturally approximates a descent direction in this regime, whereas in soft materials the governing dynamics resemble discontinuous contact PDEs, for which local diffusion is ineffective. This explains why HD performs best on stiff materials, where diffusion-like operators are effective, and why short-step refinement is particularly successful in this regime (see Figure 3 in Section B for a one-dimensional illustration). Section B.1 further analyzes how HD allocates high-frequency detail.

6.2 Limitations

HD assumes a valid coarse surface and a stable UV atlas. When the base geometry is incorrect (e.g., wrong panel topology) or chart boundaries are unstable, refinement can fail or introduce artifacts. Future work could integrate learned geometric priors and jointly optimize structure and refinement in a single framework. Moreover, HD is optimized per garment instance (2–3 hours on one H200 GPU), so its step reduction applies to the reverse process rather than to total compute, and our quantitative evaluation covers a single representative garment.

7 Conclusion

We present HD, a geometry-aligned neural diffusion framework for garment surface refinement built on a stable Nuvo UV parameterization. By operating in a surface-attached residual space and using a hash-accelerated representation, HD supports high-quality refinement with only 5–10 reverse steps. Our results suggest that a strong geometric coordinate system can substantially reduce the complexity of diffusion-based refinement and enable efficient, scalable garment finishing pipelines. On a representative stiff garment from GarmentCodeData v2, HD achieves 4.58–4.72 cm Panel L2, outperforming the AIpparel baseline (6.10 cm) while using an order of magnitude fewer diffusion steps (10 vs. hundreds to thousands). We focus on surface refinement on top of an existing geometric space constructed by Nuvo, demonstrating that the original 200–1000 steps of surface refinement optimization can be compressed into 10 steps through geometry-aligned, per-instance optimization with a diffusion solver.

AI USE STATEMENT

We used generative AI tools to assist with figure preparation, English editing, reference checks, interpretation of experimental results, and preparation and revision of the supplementary code. These tools did not generate the reported experimental measurements. The authors take responsibility for the final text, claims, figures, and code.

ETHICS STATEMENT

This work studies garment surface refinement. Potential concerns include limited generalization across garment designs, materials, and body shapes. Our evaluation does not establish suitability for safety-critical applications. Downstream use should respect applicable dataset and software licenses and intellectual property rights.

REPRODUCIBILITY STATEMENT

The Experiments section describes the dataset, hardware, evaluation metrics, and implementation settings. The Extended Experimental Details appendix provides hyperparameters, computational requirements, training commands, and input and output representations. Appendix A.1 specifies the HD refinement algorithm. The Analysis section and additional appendices discuss the method's interpretation and evaluation limitations. The reported quantitative results concern a single representative garment and should be interpreted within this scope.

REFERENCES

Søren Asmussen and Peter W. Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, 2007.

Association for Computing Machinery. ACM digital library. `https://dl.acm.org/`, 2026. Accessed September 2026.

Grigory Bartosh, Dmitry Vetrov, and Christian A. Naesseth. Neural diffusion models. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024.

Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics (SIGGRAPH)*, 30(4), 2011.

Nawaf Bou-Rabee and Victor de la Peña. Decoupling for Markov chains, 2025. arXiv:2512.19351.

Selim Emir Can, Jan Ackermann, Kiyohiro Nakayama, Ruofan Liu, Tong Wu, Yang Zheng, Hugo Bertiche, Menglei Chai, Thabo Beeler, and Gordon Wetzstein. Image2Garment: Simulation-ready garment generation from a single image, 2026. arXiv:2601.09658.

Minshuo Chen, Song Mei, Jianqing Fan, and Mengdi Wang. An overview of diffusion models: Applications, guided generation, statistical rates and optimization. *arXiv preprint arXiv:2404.07771*, 2024. URL `https://arxiv.org/abs/2404.07771`.

Lorenzo De Rebotti, Emanuele Giacomini, Giorgio Grisetti, and Luca Di Giammarino. Resolution where it counts: Hash-based GPU-accelerated 3D reconstruction via variance-adaptive voxel grids. *ACM Transactions on Graphics*, 2025. doi: 10.1145/3777909. arXiv:2511.21459.

Antoine Dumoulin et al. D-Garment: Physically grounded latent diffusion for dynamic garment deformations. *Transactions on Machine Learning Research*, 2026.

Kai He, Kaixin Yao, Qixuan Zhang, Jingyi Yu, Lingjie Liu, and Lan Xu. DressCode: Autoregressively sewing and generating garments from text guidance. *ACM Transactions on Graphics (SIGGRAPH)*, 43(4), 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Peter Holderrieth and Ezra Erives. An introduction to flow matching and diffusion models, 2025. MIT course notes (6.S184); arXiv:2506.02070.

Maria Korosteleva and Sung-Hee Lee. NeuralTailor: Reconstructing sewing pattern structures from 3D point clouds of garments. *ACM Transactions on Graphics*, 41(4), 2022.

Maria Korosteleva, Timur Levent Kesdogan, Fabian Kemper, Stephan Wenninger, Jasmin Koller, Yuhan Zhang, Mario Botsch, and Olga Sorkine-Hornung. GarmentCodeData: A dataset of 3D made-to-measure garments with sewing patterns. In *European Conference on Computer Vision (ECCV)*, 2024. URL `https://eccv.ecva.net/virtual/2024/poster/1326`.

Zorah Lähner, Daniel Cremers, and Tony Tung. DeepWrinkles: Accurate and realistic clothing modeling. In *European Conference on Computer Vision (ECCV)*, 2018.

Ren Li et al. Single view garment reconstruction using diffusion mapping via pattern coordinates. In *SIGGRAPH Conference Papers*, 2025. doi: 10.1145/3721238.3730651.

Lijuan Liu, Xiangyu Xu, Zhijie Lin, Jiabin Liang, and Shuicheng Yan. Towards garment sewing pattern reconstruction from a single image. *ACM Transactions on Graphics (SIGGRAPH Asia)*, 42 (6), 2023.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics (SIGGRAPH Asia)*, 34(6): 248:1–248:16, 2015.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.

Asit Mishra, Jorge Albericio Latorre, Jeff Pool, Darko Stosic, Dusan Stosic, Ganesh Venkatesh, Chong Yu, and Paulius Micikevicius. Accelerating sparse deep neural networks, 2021. arXiv:2104.08378.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022.

Kiyohiro Nakayama, Jan Ackermann, Timur Levent Kesdogan, Yang Zheng, Maria Korosteleva, Olga Sorkine-Hornung, Leonidas J. Guibas, Guandao Yang, and Gordon Wetzstein. AIpparel: A multimodal foundation model for digital garments. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

NVIDIA. CUDA Tile. `https://developer.nvidia.com/cuda/tile`, 2025.

Songyou Peng, Chiyu Max Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable Poisson solver. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations (ICLR)*, 2022.

Nicholas Sharp et al. Polyscope. `https://polyscope.run`, 2019.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pp. 2256–2265. PMLR, 2015.

Pratul P. Srinivasan, Stephan J. Garbin, Dor Verbin, Jonathan T. Barron, and Ben Mildenhall. Nuvo: Neural UV mapping for unruly 3D representations. *arXiv preprint arXiv:2312.05283*, 2023.

Raquel Vidaurre, Elena Garces, and Dan Casas. DiffusedWrinkles: A diffusion-based model for data-driven garment animation. In *British Machine Vision Conference (BMVC)*, 2024.

Daoye Wang, Prashanth Chandran, Gaspard Zoss, Derek Bradley, and Paulo Gotardo. MoRF: Morphable radiance fields for multiview neural head modeling. In *ACM SIGGRAPH 2022 Conference Proceedings*, 2022.

Wenbo Wang, Hsuan-I Ho, Chen Guo, Boxiang Rong, Artur Grigorev, Jie Song, Juan Jose Zarate, and Otmar Hilliges. 4D-DRESS: A 4D dataset of real-world human clothing with semantic annotations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. `https://github.com/eth-ait/4d-dress`.

Meng Zhang, Tuanfeng Wang, Duygu Ceylan, and Niloy J. Mitra. Deep detail enhancement for any garment. *Computer Graphics Forum*, 2021.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. Code: `https://github.com/richzhang/PerceptualSimilarity`.

Yuxiao Zhou, Menglei Chai, Alessandro Pepe, Markus Gross, and Thabo Beeler. GroomGen: A high-quality generative hair model using hierarchical latent representations. *ACM Transactions on Graphics (SIGGRAPH Asia)*, 42(6), 2023.

Silvia Zuffi, Angjoo Kanazawa, David W. Jacobs, and Michael J. Black. 3D menagerie: Modeling the 3D shape and pose of animals. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

A Extended Experimental Details

This appendix includes additional implementation details and experimental configurations.

A.1 HD Refinement Algorithm

Algorithm 1 HD refinement for one garment instance.

1: **Input:** coarse surface $\mathcal{S}$, conditioning $\mathbf{c}$ (patterns/materials), step budget T
2: **Stage 1 (Nuvo):** optimize UV map π on $\mathcal{S}$
3: **Stability gate:** proceed only if chart assignments stabilize; then freeze π
4: **Stage 2 (HD):** optimize hash residual field $\mathcal{R}$ and diffusion model θ using Equation (16)
5: **Inference:** sample $\mathcal{R}_T \sim \mathcal{N}(0, I)$ and run T reverse steps (discrete or ODE) to obtain $\mathcal{R}_0$
6: **Output:** refined surface via Equation (3)

A.2 Hyperparameters

- **Stage 1 (Nuvo):** iterations 0–5000
- **Stage 2 (Nuvo+HD):** iterations 5000–10000
- **Checkpoint frequency:** every 500 iterations
- **Default sample:** index 5 (rand_02VLBVOOHQ) from batch 31
- **Diffusion steps:** $T = 10$
- **Beta schedule:** linear, $\beta \in [0.1, 0.8]$
- **Residual scaling:** max 5% of base mesh size
- **Loss weights:** diffusion_weight = 1.0, lpips_weight = 0.5, l1_weight = 0.5

A.3 Residual Query as a Database Lookup

The hash lookup $H(u, v, k)$ of Section 4.2 can be viewed as a database query:

```
SELECT residual FROM garment_hash_table
WHERE uv_coords = (u, v)
  AND chart_id = k;
```

A.4 Computational Resources

All experiments are conducted on a single NVIDIA H200 GPU. Per-sample optimization typically requires 2–3 hours for the full two-stage pipeline.

A.5 Training Commands

Main experiments (stiff material, $b = 500$):

```
python train.py --config configs/icml.yaml \
  --sample_idx 5 --material stiff
```

Ablation studies:

```
# Soft material (b=1.0)
python train.py --config configs/icml.yaml \
  --sample_idx 5 --material soft

# Medium material (b=100)
python train.py --config configs/icml.yaml \
  --sample_idx 5 --material medium
```

A.6 Input Representation

Per-sample inputs include:

- Sewing patterns: panel outlines (2D), seam connectivity / stitching graph
- Material parameters: bending resistance (1, 100, or 500)
- Body template: default_body (neutral pose)

A.7 Output Representation

The output is a continuous surface refinement field (implicit):

$$\text{Refined} = \text{Nuvo}_{\text{base}}(\text{UV}) + \text{HD}_{\text{residual}}(\mathcal{R}), \tag{19}$$

where $\mathcal{R}(u, v)$ is generated by neural diffusion in 5–10 steps. The residual field $\mathcal{R}$ is aligned with the underlying surface geometry through the Nuvo parameterization.

A.8 Qualitative Results

Qualitative comparisons are intentionally omitted from this compact submission draft. In the final camera-ready version, this section should include side-by-side renderings of the base surface, HD-refined output, and reference garment for the same samples used in the quantitative tables.

B Additional Analysis

Figure 3 illustrates the energy perspective of Section 6.1 on a one-dimensional toy energy.

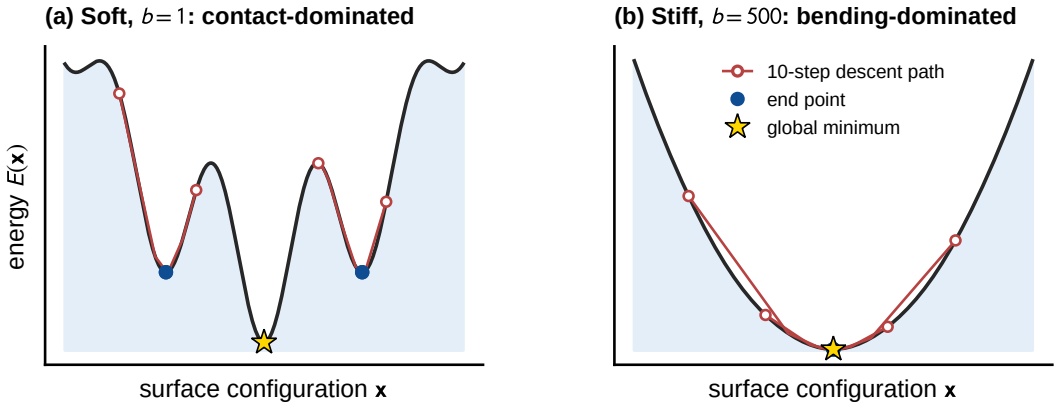

Figure 3: **Illustration of Section 6.1** (one-dimensional toy energy following Equation (17); not fitted to data). From identical starts, ten local descent steps stall in contact-induced local minima when b is small (a), but reach the minimum of the smooth, bending-dominated landscape when b is large (b).

B.1 Emergent Allocation of High-Frequency Detail

A natural question is how HD decides *where* to spend its limited high-frequency capacity, given that it uses no hand-coded rules and no explicit per-region detail labels. In HD the concentration of high-frequency detail is not prescribed; it *emerges* from the interaction of four tightly coupled mechanisms, each of which independently biases residual energy toward geometrically and perceptually salient regions.

(1) Non-uniform supervision from L_1+LPIPS. The training objective is $\mathcal{L} = \mathcal{L}_{\text{diff}} + \lambda_1 \mathcal{L}_1 + \lambda_p \mathcal{L}_{\text{LPIPS}}$ (Equation (16)). In flat, smoothly draping regions the coarse simulated surface $\mathbf{x}_{\text{base}}$ already agrees with the high-fidelity reference rendering, so the photometric residual is close to zero and back-propagation yields a vanishingly small gradient on the residual displacement; the diffusion

model consequently learns a near-zero noise/velocity prediction there. In crease, seam, and high-relief regions the situation is reversed. The L_1 term avoids the mean-mode collapse of L_2 and imposes a constant-magnitude gradient on sharp edges and pronounced deformations, while LPIPS—computed from local activations of a pretrained CNN feature pyramid—is highly sensitive to high-frequency texture, fold orientation, and fabric shadow boundaries. LPIPS therefore emits a large error signal *only* where genuine detail is present, driving the network to predict a significant residual displacement $\mathcal{R}(\mathbf{u}, k)$ at exactly those UV coordinates.

(2) Frequency-band decoupling in the multiresolution hash grid. The residual field $\mathcal{R}(\mathbf{u}, k)$ is queried through a multi-level hash encoding, which separates spatial scales by construction. Coarse levels capture panel-level, low-frequency deformation and large-area adjustments; flat regions are already explained by these levels, so the finer high-frequency entries remain effectively dormant and near-uniform. Fine levels index high-resolution UV cells: in densely wrinkled regions (e.g., the underarm, waist gathers, and seam junctions) the geometric residual varies sharply between neighbouring UV coordinates, so the high-resolution hash entries receive strong gradients, converge quickly, and activate a high-frequency response. This mitigates the spectral bias of a plain MLP and causes fine detail to concentrate naturally in localized high-frequency cells (illustrated in Figure 4).

(3) Physical guidance from the conditioning signal $\mathbf{c}$. When predicting the denoising direction, the network is conditioned on $\mathbf{c}$, comprising the material bending stiffness b, the panel layout, and the stitch graph. The sewing-pattern prior injects chart boundaries and stitching relations directly into the model, so it learns the prior that seam neighbourhoods carry higher residual energy. The material prior is equally informative: as discussed in Section 6.1 (Equation (17)), for stiff fabrics (b=500) deformation is dominated by a fourth-order thin-shell bending PDE and wrinkles form regular bands aligned with principal curvature. Using b together with the panel structure, the network can anticipate the stress-concentrated regions where wrinkles are most likely to appear.

(4) Surface-attached regularization in Nuvo space. Stage 1 builds a low-distortion, surface-attached multi-chart UV atlas, projecting the bumps, dents, and creases of the physical surface onto an unwrapped 2D manifold. Because the parameterization hugs the geometry, the high-frequency residual no longer has to search the full, unconstrained 3D Euclidean space; it is reduced to a local, low-curvature coordinate correction. Wherever surface normals vary sharply or curvature departs from the smooth base, the residual magnitude $\|\mathcal{R}(\mathbf{u}, k)\|$ necessarily deviates from zero in the corresponding UV region, yielding a natural, data-driven "detail activation map" rather than a hand-specified one.

Closed loop. These mechanisms form a self-reinforcing pipeline: the conditioning prior supplies structural topology and mechanically stress-concentrated regions; the Nuvo UV manifold restricts the otherwise global 3D search to a low-curvature local coordinate; the L_1+LPIPS rendering gradients localize the true creases and edges; and the multiresolution hash grid activates the corresponding fine-grained, high-frequency cells and completes the denoising generation. Detail placement is thus an emergent consequence of geometry-aligned diffusion rather than an explicitly engineered heuristic.

B.2 Metric Behavior and Perceptual Alignment

We find that vertex-wise L1 is not a reliable proxy for either panel-level correctness or perceptual quality. Small rigid misalignments or localized sharp folds can induce large L1 deviations even when the garment is visually plausible and panel structure is preserved. Interestingly, vertex-wise L1 does not correlate with Panel L2 or perceived garment correctness, particularly in the stiff regime. As bending stiffness increases, small rigid misalignments and sharp folds lead to disproportionately large L1 errors, despite improved global structure. This observation highlights the importance of panel-aware evaluation metrics for sewing-pattern–conditioned garment reconstruction, as purely vertex-wise measures may over-penalize visually minor discrepancies. This motivates our evaluation design: pattern-aware metrics are used as *sanity checks* (to ensure refinement does not harm structure), while image-level perceptual metrics (LPIPS) better capture refinement quality.

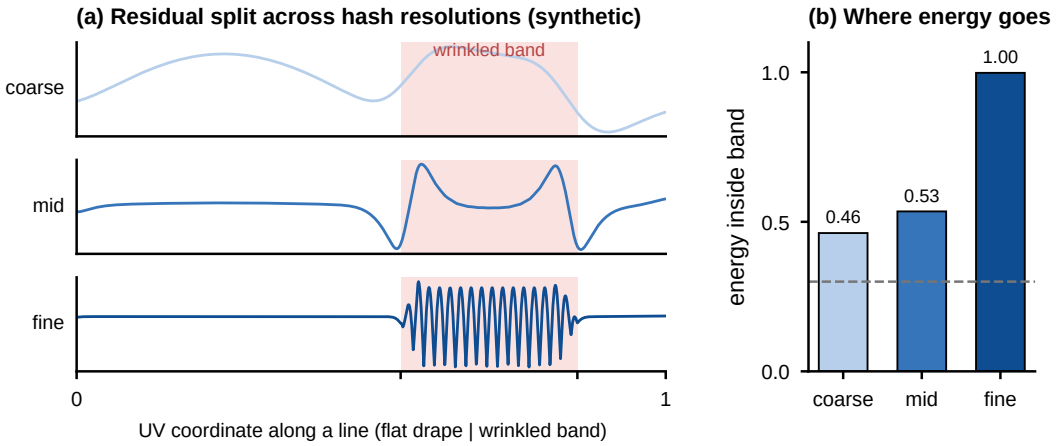

Figure 4: **Frequency-band decoupling across resolution levels** (synthetic signal; illustration of mechanism (2), not a measurement of HD). (a) A residual along a UV line, flat except for a wrinkled band (shaded), split into coarse, mid, and fine levels. (b) Fraction of each level's energy inside the band, which covers 30% of the line: the fine level concentrates almost all of its energy in the wrinkled band, while flat regions are explained by the coarse level. The dashed line marks the band's 30% area share.

C Flow Matching Comparison

We include this comparison to clarify conceptual differences. Flow matching learns a deterministic velocity field that transports samples between distributions without an explicit forward diffusion model:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{\mathbf{x}_0,\mathbf{x}_1,t}\left[\|v_\theta(\mathbf{x}_t,t) - (\mathbf{x}_1 - \mathbf{x}_0)\|^2\right], \tag{20}$$

whereas HD follows an explicit diffusion formulation with both stochastic (discrete) and deterministic (ODE) reverse-time interpretations. Our method can be viewed as complementary: HD focuses on *restricting the state space* via a geometry-aligned residual domain, while flow matching changes the training target and dynamics.

Table 3: **Conceptual comparison: HD vs. Flow Matching.** HD restricts the *state space* via a geometry-aligned residual domain, while Flow Matching modifies the *training target* and dynamics.

Aspect	HD (ours)	Flow Matching
Target of learning	Noise / score estimator	Deterministic velocity field
Forward model	Explicit diffusion	Not explicitly defined
Probabilistic interpretation	Likelihood-based	Implicit transport
Stochastic component	Injected diffusion noise	None (deterministic paths)
Reverse dynamics	Denoising process	Distribution transport