Novelty Assessment Report

Paper: Forward-Learned Discrete Diffusion: Learning how to noise to denoise faster

PDF URL: https://openreview.net/pdf?id=45EtKUdgbJ

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-07

Abstract

Discrete diffusion models are a powerful class of generative models that demonstrate strong performance across many domains. However, for efficiency, discrete diffusion typically parameterizes the generative (reverse) process with factorized distributions, which makes it difficult for the model to learn a target process in a small number of steps and necessitates a long, computationally expensive sampling procedure. To reduce the gap between the target and model distributions and enable few-step generation, we introduce a learnable noising (forward) process for discrete diffusion. Instead of fixing a Markovian forward chain, we adopt a non-Markovian formulation and introduce learnable marginal and posterior distributions. This allows the generative process to remain factorized while matching the target defined by the noising process. We train all parameters end-to-end under the standard variational objective.

Core Task Landscape

This paper addresses: **Learning Discrete Diffusion with Learnable Forward Process**

A total of **50 papers** were analyzed and organized into a taxonomy with **25 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Learnable Forward Process Architectures**
- **Discrete State Space Diffusion Models**
- **Hybrid and Unified Diffusion Frameworks**
- **Sampling and Inference Acceleration**
- **Domain-Specific Discrete Diffusion Applications**
- **Conditional and Guided Discrete Diffusion**
- **Theoretical Foundations and Analysis**
- **Specialized Continuous-Space Diffusion Extensions**

Complete Taxonomy Tree

- Learning Discrete Diffusion with Learnable Forward Process Survey Taxonomy
- Learnable Forward Process Architectures
  - Non-Markovian and Adaptive Forward Processes ★ (4 papers)
  - [0] Forward-Learned Discrete Diffusion: Learning how to noise to denoise faster (Anon et al., 2026) View paper
  - [8] Neural Flow Diffusion Models: Learnable Forward Process for Improved Diffusion Modelling (Grigory Bartosh, 2024) View paper
  - [22] A flexible diffusion model (Du, 2023) View paper
  - [33] Adaptive Destruction Processes for Diffusion Samplers (Morozov, 2025) View paper
  - Structured and Hierarchical Forward Processes (2 papers)
  - [9] Improving Discrete Diffusion Models via Structured Preferential Generation (Rissanen, 2024) View paper
  - [27] A Versatile Diffusion Transformer with Mixture of Noise Levels for Audiovisual Generation (Agrim Gupta, 2024) View paper
  - Equivariant and Geometry-Aware Forward Processes (2 papers)
  - [14] Point Cloud Resampling With Learnable Heat Diffusion (Wenqiang Xu, 2024) View paper
  - [19] Equivariant Neural Diffusion for Molecule Generation (Cornet, 2024) View paper
- Discrete State Space Diffusion Models
  - Categorical and Token-Based Discrete Diffusion (6 papers)
  - [5] Glauber generative model: Discrete diffusion models via binary classification (Nagaraj, 2024) View paper
  - [12] Structured denoising diffusion models in discrete state-spaces (Jacob Austin, 2021) View paper
  - [25] Blackout Diffusion: Generative Diffusion Models in Discrete-State Spaces (Santos, 2023) View paper
  - [26] DiffusionBERT: Improving Generative Masked Language Models with Diffusion Models (HE Zhengfu, 2022) View paper
  - [28] A Cheaper and Better Diffusion Language Model with Soft-Masked Noise (Chen, 2023) View paper
  - [49] Discrete Modeling via Boundary Conditional Diffusion Processes (Xiaocheng Feng, 2024) View paper
  - Graph and Combinatorial Structure Diffusion (3 papers)
  - [11] Autoregressive Diffusion Model for Graph Generation (Kong, 2023) View paper
  - [20] Symmetricdiffusers: Learning discrete diffusion on finite symmetric groups (Zhang Yongxing, 2024) View paper
  - [37] A Prior-based Discrete Diffusion Model for Social Graph Generation (Shu Yin, 2025) View paper
  - Bit-Level and Binary Discrete Diffusion (1 papers)
  - [10] Bit-Level Discrete Diffusion with Markov Probabilistic Models: An Improved Framework with Sharp Convergence Bounds under Minimal Assumptions (Le-Tuyet-Nhi Pham, 2025) View paper

- Hybrid and Unified Diffusion Frameworks
  - Discrete-Continuous Unified Models (1 papers)
  - [3] Generating, reconstructing, and representing discrete and continuous data: Generalized diffusion with learnable encodingdecoding (Liu Guangyi, 2024) View paper
  - Diffusion-Flow and Diffusion-GAN Hybrids (2 papers)
  - [4] Diffusion-GAN: Training GANs with Diffusion (Wang, 2022) View paper
  - [23] Diffusion Augmented Flows: Combining Normalizing Flows and Diffusion Models for Accurate Latent Space Mapping (Soham Sajekar, 2024) View paper
- Sampling and Inference Acceleration
  - Learned Discretization and Step Reduction (2 papers)
  - [38] Learning to discretize denoising diffusion odes (Tong, 2024) View paper
  - [47] Learnable Sampler Distillation for Discrete Diffusion Models (Feiyang Fu, 2025) View paper
  - Quantized and Transition-Based Acceleration (2 papers)
  - [18] Almost Linear Convergence under Minimal Score Assumptions: Quantized Transition Diffusion (Huang, 2025) View paper
  - [43] Learning Flexible Forward Trajectories for Masked Molecular Diffusion (Seo Hyun-jin, 2025) View paper
- Domain-Specific Discrete Diffusion Applications
  - Molecular and Protein Design (1 papers)
  - [7] Protein Design with Guided Discrete Diffusion (Gruver, 2023) View paper
  - Layout and Spatial Structure Generation (1 papers)
  - [30] LayoutDiffusion: Improving Graphic Layout Generation by Discrete Diffusion Probabilistic Models (Junyi Zhang, 2023) View paper
  - Recommendation and Reranking Systems (3 papers)
  - [15] Discrete Conditional Diffusion for Reranking in Recommendation (Xiao Lin, 2024) View paper
  - [16] Continuous-time Discrete-space Diffusion Model for Recommendation (Chengyi Liu, 2025) View paper
  - [45] Fading to Grow: Growing Preference Ratios via Preference Fading Discrete Diffusion for Recommendation (Hu Guoqing, 2025) View paper
  - Motion and Temporal Sequence Generation (2 papers)
  - [35] M2D2M: Multi-Motion Generation from Text with Discrete Diffusion Models (Chi Hyung-gun, 2024) View paper
  - [39] DiffusionPhase: Motion Diffusion in Frequency Domain (Wan Weilin, 2023) View paper
- Conditional and Guided Discrete Diffusion
  - Classifier-Free and Reinforcement-Based Guidance (2 papers)
  - [2] Adding conditional control to diffusion models with reinforcement learning (Zhao, 2024) View paper
  - [24] DFRL-DS: A Diffusion-based Reinforcement Learning Algorithm in Discrete Actions for Base Station Energy-saving Control (Jun Li, 2024) View paper
  - Energy-Based and Score-Based Guidance (1 papers)
  - [46] Sampling from Energy distributions with Target Concrete Score Identity (Vargas, 2025) View paper
- Theoretical Foundations and Analysis (3 papers)
  - [29] A Unified Approach to Analysis and Design of Denoising Markov Models (Ren, 2025) View paper
  - [34] Non-Denoising Forward-Time Diffusions (Peluchetti, 2023) View paper
  - [40] Trans-dimensional generative modeling via jump diffusion models (Campbell, 2023) View paper
- Specialized Continuous-Space Diffusion Extensions
  - Dense Prediction and Discriminative Tasks (2 papers)
  - [1] D3RM: A Discrete Denoising Diffusion Refinement Model for Piano Transcription (Kwon, 2025) View paper
  - [6] Ddp: Diffusion model for dense visual prediction (YUANFENG, 2023) View paper
  - Temporal Dynamics and Video Modeling (1 papers)
  - [13] Dynamical diffusion: Learning temporal dynamics with diffusion models (Guo, 2025) View paper
  - Medical Imaging and Super-Resolution (2 papers)
  - [17] Discrete residual diffusion model for high-resolution prostate MRI synthesis (Zhitao Han, 2024) View paper
  - [32] Universal Network for Image Registration and Generation Using Denoising Diffusion Probability Model (Huizhong Ji, 2024) View paper
  - Flow-Based and Momentum Diffusion (1 papers)
  - [21] Flow Diverse and Efficient: Learning Momentum Flow Matching via Stochastic Velocity Field Sampling (Ma, 2025) View paper
  - Representation Learning and Autoencoders (2 papers)
  - [31] Exploring Diffusion Time-steps for Unsupervised Representation Learning (Yue Zhongqi, 2024) View paper
  - [48] On Designing Diffusion Autoencoders for Efficient Generation and Representation Learning (Proszewska, 2025) View paper
  - Deblurring and Image Restoration (1 papers)
  - [44] BlurDM: A Blur Diffusion Model for Image Deblurring (Jin-Ting He, 2025) View paper
  - Predictor-Corrector and Sampling Schemes (2 papers)
  - [36] Evaluating the design space of diffusion-based generative models (Ye He, 2024) View paper
  - [41] Discrete predictor-corrector diffusion models for image synthesis (J Lezama, 2022) View paper
  - Specialized Network Architectures (2 papers)
  - [42] Toward Diffusion-Based Deep Reinforcement Learning for Discrete Decision-Making: Methods and Evaluations (Zhen Chen, 2025) View paper
  - [50] Trainable Diffusion Network Based on Morphological Laplacian (Gouki Okada, 2021) View paper

Narrative

Core task: learning discrete diffusion with learnable forward process. The field of discrete diffusion models has evolved into a rich taxonomy spanning multiple complementary directions. At the highest level, researchers explore Learnable Forward Process Architectures that adapt the corruption mechanism itself, Discrete State Space Diffusion Models that handle categorical or structured data, and Hybrid and Unified Diffusion Frameworks that bridge continuous and discrete settings. Parallel branches address practical concerns such as Sampling and Inference Acceleration, Domain-Specific Discrete Diffusion Applications (e.g., D3RM Piano Transcription[1], Guided Protein Design[7]), and Conditional and Guided Discrete Diffusion for controlled generation. Theoretical Foundations and Analysis provide rigorous underpinnings, while Specialized Continuous-Space Diffusion Extensions adapt ideas from the continuous domain. Representative works like Glauber Generative Model[5] and Structured Discrete Denoising[12] illustrate how different branches tackle the challenge of defining appropriate noise processes for non-Euclidean spaces.

Within the Learnable Forward Process Architectures branch, a particularly active line of work focuses on Non-Markovian and Adaptive Forward Processes, where the corruption schedule is not fixed but learned or conditioned on data. Forward-Learned Discrete Diffusion[0] sits squarely in this cluster, emphasizing end-to-end optimization of the forward trajectory alongside the reverse denoising network. Nearby, Flexible Diffusion[22] and Adaptive Destruction Processes[33] explore related themes of flexibility and data-dependent corruption, though they differ in whether adaptivity is global or instance-specific. In contrast, Neural Flow Diffusion[8] leans toward continuous-time formulations with learnable dynamics. The central trade-off across these methods is between expressiveness—allowing richer forward processes that better match data structure—and tractability, as non-standard schedules can complicate training and sampling. Forward-Learned Discrete Diffusion[0] addresses this by jointly learning both directions, positioning itself as a holistic approach within the adaptive forward process paradigm.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Neural Flow Diffusion Models: Learnable Forward Process for Improved Diffusion Modelling

Authors: Grigory Bartosh, Dmitry Vetrov | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Conventional diffusion models typically relies on a fixed forward process, which implicitly defines complex marginal distributions over latent variables. This can often complicate the reverse process' task in learning generative trajectories, and results in costly inference for diffusion models. To address these limitations, we introduce Neural Flow Diffusion Models (NFDM), a novel framework that enhances diffusion models by supporting a broader range of forward processes beyond the standard Gau...

Relationship Analysis

Both papers belong to the Non-Markovian and Adaptive Forward Processes category, focusing on learning flexible forward processes in discrete/continuous diffusion models to improve generation efficiency. They overlap in addressing the limitation of fixed forward processes by introducing learnable, non-Markovian forward dynamics that adapt to the reverse process during training. The key difference is that the original paper (FLDD) specifically targets discrete diffusion with factorized reverse processes and uses Maximum Coupling for posterior parameterization, while the candidate paper (NFDM) focuses on continuous diffusion with neural flow-based transformations and supports both generative modeling and bridge learning between distributions.

---

2. A flexible diffusion model

Authors: Du, Weitao, Yang Tao, Weitao Du, Zhang He, et al. (9 authors total) | **Year/Venue**: 2023 | **URL**: View paper

Abstract

Diffusion (score-based) generative models have been widely used for modeling various types of complex data, including images, audios, and point clouds. Recently, the deep connection between forward-backward stochastic differential equations (SDEs) and diffusion-based models has been revealed, and several new variants of SDEs are proposed (e.g., sub-VP, critically-damped Langevin) along this line. Despite the empirical success of the hand-crafted fixed forward SDEs, a great quantity of proper for...

Relationship Analysis

Both papers belong to the Non-Markovian and Adaptive Forward Processes category, focusing on learning flexible forward diffusion dynamics rather than using fixed noising schedules. They overlap in their core motivation to learn forward processes that better align with reverse processes for improved generation, but differ fundamentally in their approach: the original paper (FLDD) learns discrete non-Markovian forward processes with learnable marginals and posteriors using Maximum Coupling for few-step discrete generation, while the candidate paper parameterizes continuous forward SDEs using Riemannian geometry and symplectic structures to create spatially-adaptive diffusion processes on data manifolds.

---

3. Adaptive Destruction Processes for Diffusion Samplers

Authors: Morozov, Nikita, Timofei Gritsaev, Nikita Morozov, Tiapkin, et al. (20 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

This paper explores the challenges and benefits of a trainable destruction process in diffusion samplers -- diffusion-based generative models trained to sample an unnormalised density without access to data samples. Contrary to the majority of work that views diffusion samplers as approximations to an underlying continuous-time model, we view diffusion models as discrete-time policies trained to produce samples in very few generation steps. We propose to trade some of the elegance of the underly...

Relationship Analysis

Both papers belong to the Non-Markovian and Adaptive Forward Processes category, exploring learnable forward processes in diffusion models to improve generation efficiency. They overlap in addressing the challenge of reducing sampling steps through adaptive forward dynamics with trainable parameters. However, the original paper (FLDD) focuses on discrete diffusion with a non-Markovian formulation using learnable marginals and Maximum Coupling posteriors for few-step generation, while the candidate paper (Adaptive Destruction Processes) targets continuous-space diffusion samplers for unnormalized density sampling, decoupling generation and destruction variances and employing reinforcement learning objectives with techniques like prioritized replay buffers.

Contributions Analysis

Overall novelty summary. The paper introduces a learnable forward noising process for discrete diffusion, enabling end-to-end training of both corruption and generation dynamics. It resides in the 'Non-Markovian and Adaptive Forward Processes' leaf, which contains four papers total, including the original work. This leaf sits within the broader 'Learnable Forward Process Architectures' branch, indicating a moderately active but not overcrowded research direction. The taxonomy shows that while learnable forward processes are an established theme, the specific combination of non-Markovian formulation with learnable marginals and posteriors occupies a relatively focused niche.

The taxonomy reveals neighboring leaves addressing 'Structured and Hierarchical Forward Processes' (two papers) and 'Equivariant and Geometry-Aware Forward Processes' (two papers), suggesting that learnable forward process research branches into specialized structural constraints. The sibling papers in the same leaf explore related adaptive dynamics but differ in scope: some focus on continuous-time formulations or instance-specific adaptivity, while this work emphasizes joint optimization of marginals and posteriors. The broader 'Discrete State Space Diffusion Models' branch (thirteen papers across three leaves) provides context for the discrete setting, though those works typically assume fixed forward processes.

Among twenty-six candidates examined, the contribution-level analysis shows varied novelty profiles. The core FLDD framework (ten candidates examined, zero refutable) appears relatively novel within the limited search scope. The end-to-end simulation-free training procedure (eight candidates examined, one refutable) has at least one overlapping prior work among the examined papers, suggesting this aspect may be less distinctive. The non-Markovian parameterization with learnable marginals and posteriors (eight candidates examined, zero refutable) shows no clear refutation in the examined set, indicating potential novelty in this specific formulation.

Based on the limited search of twenty-six semantically related papers, the work appears to occupy a moderately explored area with some novel aspects. The analysis does not cover exhaustive literature review or papers outside the top-K semantic matches, so conclusions about absolute novelty remain tentative. The taxonomy structure suggests the field is diversifying into specialized branches, and this work contributes to the adaptive forward process direction with a particular emphasis on joint learning of corruption and generation.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Forward-Learned Discrete Diffusion (FLDD) framework

Description: The authors propose FLDD, a discrete diffusion framework that introduces learnable forward (noising) processes with non-Markovian formulation. This allows the generative process to remain factorized while better matching the target distribution, enabling few-step generation without changing the reverse parameterization or adding inference overhead.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Frequency Domain Diffusion Model with Scale-Dependent Noise Schedule

URL: View paper

Brief Assessment

Frequency Domain Schedule[75] operates in the frequency domain for image generation with fixed noise schedules, while FLDD introduces learnable forward processes for discrete diffusion across multiple domains (text, molecules, images). These are fundamentally different approaches to diffusion modeling.

---

2. Text diffusion model with encoder-decoder transformers for sequence-to-sequence generation

URL: View paper

Brief Assessment

Text Encoder-Decoder[77] focuses on continuous diffusion in embedding space for sequence-to-sequence tasks with encoder-decoder transformers, not on learnable discrete forward processes or non-Markovian formulations for few-step generation.

---

3. Few-Shot Learner Parameterization by Diffusion Time-Steps

URL: View paper

Brief Assessment

Few-Shot Time-Steps[74] focuses on using diffusion model time-steps as an inductive bias for few-shot image classification, not on learning forward noising processes for discrete generative modeling. The candidate addresses a completely different problem domain (discriminative few-shot learning vs. generative discrete diffusion).

---

4. Reschedule Diffusion-based Bokeh Rendering

URL: View paper

Brief Assessment

Reschedule Bokeh Rendering[71] focuses on continuous image-to-image bokeh rendering with adaptive noise scheduling, not discrete diffusion with learnable forward processes for general generation tasks.

---

5. Efficient diffusion policies for offline reinforcement learning

URL: View paper

Brief Assessment

Efficient Offline Policies[69] focuses on continuous action spaces in offline RL with fixed forward processes, not discrete diffusion with learnable forward processes. The candidate uses action approximation for efficiency in RL settings, while the original proposes learning the forward noising process itself for discrete data generation.

---

6. Noise Estimation for Generative Diffusion Models

URL: View paper

Brief Assessment

Noise Estimation[73] focuses on learning noise schedules for continuous diffusion models to improve few-step generation, whereas FLDD introduces learnable forward processes specifically for discrete diffusion with non-Markovian formulation and factorized distributions.

---

7. Seqdiffuseq: Text diffusion with encoder-decoder transformers

URL: View paper

Brief Assessment

SeqDiffuSeq[72] focuses on continuous diffusion models for sequence-to-sequence text generation with encoder-decoder transformers, using fixed forward processes with adaptive noise schedules. It does not propose learnable forward (noising) processes or non-Markovian formulations for discrete diffusion.

---

8. Ambient diffusion posterior sampling: Solving inverse problems with diffusion models trained on corrupted data

URL: View paper

Brief Assessment

Ambient Diffusion Posterior[70] focuses on solving inverse problems using diffusion models trained on corrupted data in continuous/image domains, not on discrete diffusion with learnable forward processes for few-step generation.

---

9. Score-Optimal Diffusion Schedules

URL: View paper

Brief Assessment

Score-Optimal Schedules[78] focuses on optimizing discretization schedules for continuous diffusion models through cost-based analysis of the diffusion path, not on learning forward processes for discrete diffusion models.

---

10. DINOISER: Diffused Conditional Sequence Learning by Manipulating Noises

URL: View paper

Brief Assessment

DINOISER[76] focuses on manipulating noise scales during training and inference for discrete sequence learning, but does not propose a learnable forward process with non-Markovian formulation as in FLDD. DINOISER[76] uses fixed noise schedules and adapts them through clipping strategies rather than learning the forward marginals and posteriors.

Contribution 2: End-to-end simulation-free training procedure

Description: The authors develop a training method that optimizes both forward and reverse process parameters jointly under the standard variational objective. They use REINFORCE for unbiased gradient estimation and introduce a continuous relaxation warm-up strategy to stabilize training from scratch.

This contribution was assessed against **8 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Safe, Efficient, and Robust Reinforcement Learning for Ranking and Diffusion Models

URL: View paper

Brief Assessment

Safe Ranking Reinforcement[67] appears to be a thesis focused on reinforcement learning for ranking and diffusion models, but the provided context contains only title page and copyright information without technical content about training procedures or REINFORCE methods.

---

2. Training Diffusion Models with Reinforcement Learning

URL: View paper

Brief Assessment

Reinforcement Learning Training[59] focuses on optimizing diffusion models for downstream objectives using policy gradient methods (DDPO), not on simulation-free training of forward and reverse processes jointly under variational objectives with REINFORCE for discrete diffusion as in the original paper.

---

3. Large-scale Reinforcement Learning for Diffusion Models

URL: View paper

Brief Assessment

Large-scale Reinforcement[68] focuses on applying REINFORCE to diffusion models for reward optimization in text-to-image generation, not on developing a general simulation-free training procedure for discrete diffusion models with continuous relaxation warm-up as in the original paper.

---

4. Inference-time alignment control for diffusion models with reinforcement learning guidance

URL: View paper

Brief Assessment

Inference-time Alignment Control[62] focuses on inference-time guidance for already-trained diffusion models using RL fine-tuning, not on developing simulation-free training procedures with REINFORCE gradient estimation.

---

5. Amortizing intractable inference in diffusion models for vision, language, and control

URL: View paper

Brief Assessment

Amortizing Intractable Inference[61] focuses on posterior inference under diffusion priors using trajectory balance objectives, not on learning forward processes in discrete diffusion. The training methods differ fundamentally in their objectives and problem settings.

---

6. A simulation-free deep learning approach to stochastic optimal control

URL: View paper

Prior Art Analysis

Simulation-free Stochastic Control[66] demonstrates that simulation-free training using REINFORCE-style gradient estimation for stochastic optimal control was established prior to the original paper's work. The candidate paper presents a complete framework for end-to-end training that avoids differentiating through SDE solutions by using Girsanov theorem to compute gradients on-policy, which directly parallels the original paper's claim of developing a simulation-free training method with REINFORCE for unbiased gradient estimation. Both papers explicitly use REINFORCE or REINFORCE-like methods to obtain unbiased gradient estimators without backpropagating through stochastic trajectories.

Evidence

Evidence 1 - **Rationale**: Both papers claim to provide simulation-free training procedures. The candidate explicitly describes avoiding adjoint solutions and using on-policy gradient calculation, which is the core of simulation-free training. - **Original**: we provide an efficient, end-to-end, simulation-free training procedure. - **Candidate**: we propose a simulation-free algorithm for the solution of generic problems in stochastic optimal control (soc). unlike existing methods, our approach does not require the solution of an adjoint problem, but rather leverages girsanov theorem to directly calculate the gradient of the soc objective on...

Evidence 2 - **Rationale**: Both papers explicitly use REINFORCE (Williams, 1992) for unbiased gradient estimation. The candidate paper establishes this as a known approach in continuous-time settings, predating the original's claim of developing this method. - **Original**: unbiased optimization.the standard approach for computing gradients of this form is the reinforce method williams (1992). we can rewrite the gradients of the diffusion loss as follows: ∇φldiff =e u(t)qφ(zt|x) [∇φ qφ(zt|x) ⌊qφ(zt|x)⌋sg dkl qφ(zs|zt,x)∥p θ(zs|zt) !] - **Candidate**: the algorithm we propose can be viewed as a continuous-time variant of the well-known reinforce method (williams, 1992; 1988; sutton et al., 1999).

Evidence 3 - **Rationale**: Both papers describe using Monte Carlo methods to build unbiased gradient estimators for end-to-end training, with the candidate demonstrating this approach was already established. - **Original**: in this way, we can use the monte carlo method to build an unbiased gradient estimator and train the model end-to-end. importantly, as in conventional diffusion, we optimize a variational bound on the model's likelihood. - **Candidate**: equation (6) can be implemented to directly estimate the gradient of the objective l(θ) = j(uθ) by replacing the expectation exθ by an empirical expectation over an ensemble of independent realizations of the sde 7. alternatively, we can use automatic differentiation of an alternative objective

Evidence 4 - **Rationale**: Both papers acknowledge the high-variance and training instability issues with REINFORCE. The original paper's warm-up strategy addresses the same fundamental challenge that the candidate paper identifies, showing this is a known limitation requiring mitigation strategies. - **Original**: relaxed warm-up.unfortunately, reinforce is known to produce high-variance gradients, so training a model from scratch with reinforce alone often leads to instability. to obtain a better initialization, we adopt a different strategy. we begin training with a continuous relaxation of the categorical ... - **Candidate**: limitations. while our approach is simpler and more efficient than the vanilla method to minimize the soc objective, it eventually performs the same minimization of the soc objective. as a result, it is prone to the same issues: slow convergence of the training in absence of good warm-start and pote...

---

7. Simplified and generalized masked diffusion for discrete data

URL: View paper

Brief Assessment

Simplified Masked Diffusion[60] uses a continuous-time variational objective with simple cross-entropy losses and does not employ REINFORCE or continuous relaxation warm-up strategies. The training approaches are fundamentally different.

---

8. Diffusion Model as Representation Learner

URL: View paper

Brief Assessment

Representation Learner[64] focuses on using pre-trained diffusion models for knowledge transfer to recognition tasks via reinforcement learning for time-step selection, not on training diffusion models themselves with REINFORCE for gradient estimation of forward/reverse processes.

Contribution 3: Non-Markovian forward process parameterization with learnable marginals and posteriors

Description: The authors reformulate the forward process from a Markovian chain to a non-Markovian form with learnable factorized marginals and tractable posteriors constructed via Maximum Coupling. This parameterization enables efficient sampling during training while allowing each coordinate's trajectory to depend on the entire data point.

This contribution was assessed against **8 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Conditional Diffusion Model with Nonlinear Data Transformation for Time Series Forecasting

URL: View paper

Brief Assessment

Nonlinear Data Transformation[57] focuses on time series forecasting with nonlinear transformations in the forward process, while the original paper addresses discrete diffusion models for general discrete domains (images, molecules, text) with a different parameterization approach using Maximum Coupling for posteriors.

---

2. Generative fractional diffusion models

URL: View paper

Brief Assessment

Fractional Diffusion Models[51] focuses on replacing Brownian motion with fractional Brownian motion (FBM) in continuous-time score-based models, using a Markov approximation (MA-FBM) for tractability. The original paper proposes learning discrete diffusion forward processes with factorized marginals and Maximum Coupling posteriors for discrete data, which is a fundamentally different technical approach and domain.

---

3. Constrained Diffusion: Applications to Image Generation, Manifold Learning, and Motion Planning

URL: View paper

Brief Assessment

Constrained Diffusion[55] focuses on diffusion processes on Riemannian manifolds and reflected diffusion for bounded regions, not on learning forward process marginals and posteriors for discrete diffusion models.

---

4. Bernoulli Priors as Efficient Denoising Guides for Diffusion Models

URL: View paper

Brief Assessment

Bernoulli Priors[56] focuses on learning input-dependent binary latent variables to guide denoising in diffusion autoencoders, not on reformulating forward processes with learnable factorized marginals and Maximum Coupling posteriors as in the original paper.

---

5. Fast non-markovian diffusion model for weakly supervised anomaly detection in brain mr images

URL: View paper

Brief Assessment

Fast Non-Markovian Anomaly[54] applies a non-Markovian conditional diffusion model for weakly supervised anomaly detection in brain MR images, which is a specialized application domain. The candidate's limited context does not provide sufficient detail about learnable marginals, posteriors, or Maximum Coupling techniques to challenge the original paper's novelty in reformulating the forward process with these specific components.

---

6. Diffusion Bridge Implicit Models

URL: View paper

Brief Assessment

Diffusion Bridge Implicit[53] focuses on non-Markovian bridges for paired distribution interpolation in image translation, not on learning forward process marginals and posteriors for general discrete diffusion as in the original paper.

---

7. Remaining useful life prediction for multivariable stochastic degradation systems with nonâ□□Markovian diffusion processes

URL: View paper

Brief Assessment

Non-Markovian Degradation[52] applies fractional Brownian motion to industrial equipment degradation modeling for remaining useful life prediction, not to diffusion models for generative modeling. The domains and technical objectives are entirely different.

---

8. Generative Modeling in AI and Stochastic Processes

URL: View paper

Brief Assessment

Stochastic Processes Modeling[58] appears to be a general thesis on generative modeling and stochastic processes. The provided context contains only the title page with no technical content about non-Markovian forward processes, learnable marginals, or diffusion model architectures.

Appendix: Text Similarity Detection

Textual similarity detection checked 31 papers and found 2 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Neural Flow Diffusion Models: Learnable Forward Process for Improved Diffusion Modelling

Detected in: Core Task (sibling)

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Forward-Learned Discrete Diffusion: Learning how to noise to denoise faster View paper
- [1] D3RM: A Discrete Denoising Diffusion Refinement Model for Piano Transcription View paper
- [2] Adding conditional control to diffusion models with reinforcement learning View paper
- [3] Generating, reconstructing, and representing discrete and continuous data: Generalized diffusion with learnable encodingdecoding View paper
- [4] Diffusion-GAN: Training GANs with Diffusion View paper
- [5] Glauber generative model: Discrete diffusion models via binary classification View paper
- [6] Ddp: Diffusion model for dense visual prediction View paper
- [7] Protein Design with Guided Discrete Diffusion View paper
- [8] Neural Flow Diffusion Models: Learnable Forward Process for Improved Diffusion Modelling View paper
- [9] Improving Discrete Diffusion Models via Structured Preferential Generation View paper
- [10] Bit-Level Discrete Diffusion with Markov Probabilistic Models: An Improved Framework with Sharp Convergence Bounds under Minimal Assumptions View paper
- [11] Autoregressive Diffusion Model for Graph Generation View paper
- [12] Structured denoising diffusion models in discrete state-spaces View paper
- [13] Dynamical diffusion: Learning temporal dynamics with diffusion models View paper
- [14] Point Cloud Resampling With Learnable Heat Diffusion View paper
- [15] Discrete Conditional Diffusion for Reranking in Recommendation View paper
- [16] Continuous-time Discrete-space Diffusion Model for Recommendation View paper
- [17] Discrete residual diffusion model for high-resolution prostate MRI synthesis View paper
- [18] Almost Linear Convergence under Minimal Score Assumptions: Quantized Transition Diffusion View paper
- [19] Equivariant Neural Diffusion for Molecule Generation View paper
- [20] Symmetricdiffusers: Learning discrete diffusion on finite symmetric groups View paper
- [21] Flow Diverse and Efficient: Learning Momentum Flow Matching via Stochastic Velocity Field Sampling View paper
- [22] A flexible diffusion model View paper
- [23] Diffusion Augmented Flows: Combining Normalizing Flows and Diffusion Models for Accurate Latent Space Mapping View paper
- [24] DFRL-DS: A Diffusion-based Reinforcement Learning Algorithm in Discrete Actions for Base Station Energy-saving Control View paper
- [25] Blackout Diffusion: Generative Diffusion Models in Discrete-State Spaces View paper
- [26] DiffusionBERT: Improving Generative Masked Language Models with Diffusion Models View paper
- [27] A Versatile Diffusion Transformer with Mixture of Noise Levels for Audiovisual Generation View paper
- [28] A Cheaper and Better Diffusion Language Model with Soft-Masked Noise View paper
- [29] A Unified Approach to Analysis and Design of Denoising Markov Models View paper
- [30] LayoutDiffusion: Improving Graphic Layout Generation by Discrete Diffusion Probabilistic Models View paper
- [31] Exploring Diffusion Time-steps for Unsupervised Representation Learning View paper
- [32] Universal Network for Image Registration and Generation Using Denoising Diffusion Probability Model View paper
- [33] Adaptive Destruction Processes for Diffusion Samplers View paper
- [34] Non-Denoising Forward-Time Diffusions View paper
- [35] M2D2M: Multi-Motion Generation from Text with Discrete Diffusion Models View paper
- [36] Evaluating the design space of diffusion-based generative models View paper
- [37] A Prior-based Discrete Diffusion Model for Social Graph Generation View paper
- [38] Learning to discretize denoising diffusion odes View paper
- [39] DiffusionPhase: Motion Diffusion in Frequency Domain View paper
- [40] Trans-dimensional generative modeling via jump diffusion models View paper
- [41] Discrete predictor-corrector diffusion models for image synthesis View paper
- [42] Toward Diffusion-Based Deep Reinforcement Learning for Discrete Decision-Making: Methods and Evaluations View paper
- [43] Learning Flexible Forward Trajectories for Masked Molecular Diffusion View paper
- [44] BlurDM: A Blur Diffusion Model for Image Deblurring View paper
- [45] Fading to Grow: Growing Preference Ratios via Preference Fading Discrete Diffusion for Recommendation View paper
- [46] Sampling from Energy distributions with Target Concrete Score Identity View paper
- [47] Learnable Sampler Distillation for Discrete Diffusion Models View paper
- [48] On Designing Diffusion Autoencoders for Efficient Generation and Representation Learning View paper
- [49] Discrete Modeling via Boundary Conditional Diffusion Processes View paper
- [50] Trainable Diffusion Network Based on Morphological Laplacian View paper
- [51] Generative fractional diffusion models View paper
- [52] Remaining useful life prediction for multivariable stochastic degradation systems with nonâMarkovian diffusion processes View paper
- [53] Diffusion Bridge Implicit Models View paper
- [54] Fast non-markovian diffusion model for weakly supervised anomaly detection in brain mr images View paper
- [55] Constrained Diffusion: Applications to Image Generation, Manifold Learning, and Motion Planning View paper

- [56] Bernoulli Priors as Efficient Denoising Guides for Diffusion Models View paper
- [57] Conditional Diffusion Model with Nonlinear Data Transformation for Time Series Forecasting View paper
- [58] Generative Modeling in AI and Stochastic Processes View paper
- [59] Training Diffusion Models with Reinforcement Learning View paper
- [60] Simplified and generalized masked diffusion for discrete data View paper
- [61] Amortizing intractable inference in diffusion models for vision, language, and control View paper
- [62] Inference-time alignment control for diffusion models with reinforcement learning guidance View paper
- [63] Enhancing Diffusion Models with Text-Encoder Reinforcement Learning View paper
- [64] Diffusion Model as Representation Learner View paper
- [65] Reinforcing the Diffusion Chain of Lateral Thought with Diffusion Language Models View paper
- [66] A simulation-free deep learning approach to stochastic optimal control View paper
- [67] Safe, Efficient, and Robust Reinforcement Learning for Ranking and Diffusion Models View paper
- [68] Large-scale Reinforcement Learning for Diffusion Models View paper
- [69] Efficient diffusion policies for offline reinforcement learning View paper
- [70] Ambient diffusion posterior sampling: Solving inverse problems with diffusion models trained on corrupted data View paper
- [71] Reschedule Diffusion-based Bokeh Rendering View paper
- [72] Seqdiffuseq: Text diffusion with encoder-decoder transformers View paper
- [73] Noise Estimation for Generative Diffusion Models View paper
- [74] Few-Shot Learner Parameterization by Diffusion Time-Steps View paper
- [75] Frequency Domain Diffusion Model with Scale-Dependent Noise Schedule View paper
- [76] DINOISER: Diffused Conditional Sequence Learning by Manipulating Noises View paper
- [77] Text diffusion model with encoder-decoder transformers for sequence-to-sequence generation View paper
- [78] Score-Optimal Diffusion Schedules View paper