
Kuramoto Orientation Diffusion Models

Yue Song¹, T. Anderson Keller², Sevan Brodjian¹, Takeru Miyato^{3,4},
Yisong Yue¹, Pietro Perona¹, and Max Welling^{4,5}

¹ Caltech ² Harvard University ³ University of Tübingen
⁴ University of Amsterdam ⁵ CuspAI

Abstract

Orientation-rich images, such as fingerprints and textures, often exhibit coherent angular directional patterns that are challenging to model using standard generative approaches based on isotropic Euclidean diffusion. Motivated by the role of phase synchronization in biological systems, we propose a score-based generative model built on periodic domains by leveraging stochastic Kuramoto dynamics in the diffusion process. In neural and physical systems, Kuramoto models capture synchronization phenomena across coupled oscillators – a behavior that we re-purpose here as an inductive bias for structured image generation. In our framework, the forward process performs *synchronization* among phase variables through globally or locally coupled oscillator interactions and attraction to a global reference phase, gradually collapsing the data into a low-entropy von Mises distribution. The reverse process then performs *desynchronization*, generating diverse patterns by reversing the dynamics with a learned score function. This approach enables structured destruction during forward diffusion and a hierarchical generation process that progressively refines global coherence into fine-scale details. We implement wrapped Gaussian transition kernels and periodicity-aware networks to account for the circular geometry. Our method achieves competitive results on general image benchmarks and significantly improves generation quality on orientation-dense datasets like fingerprints and textures. Ultimately, this work demonstrates the promise of biologically inspired synchronization dynamics as structured priors in generative modeling. Code is available at: <https://github.com/KingJamesSong/OrientationDiffusion>.

1 Introduction

Synchronization phenomena, characterized by coordinated rhythmic activity across coupled oscillators, are ubiquitous in nature. Such patterns are fundamental in biological neural networks, where synchronous neural firing supports critical cognitive functions including attention, sensory integration, and memory [9]. Similarly, in physical and engineering systems, synchronization underpins the collective behavior of coupled oscillatory circuits, chemical reactions, and mechanical structures [36, 62]. Central to understanding these phenomena is the Kuramoto model [36], a canonical framework arising from nonlinear dynamics that elegantly describes how global coherence emerges spontaneously from interactions among oscillatory units.

In this work, we explore how these principles of synchronized coherence can be harnessed to tackle a persistent challenge in generative modeling for *orientation-rich data*, such as fingerprints, textures, and directional fields. These data appear in numerous applications, including fingerprint generation for biometric security [17], material characterization through crystal orientation analysis [43], and fiber orientation modeling for improved medical diagnostics [47, 64]. The defining structures of these data types are characterized primarily by the orientations of local features (*i.e.*, piece-wise constant

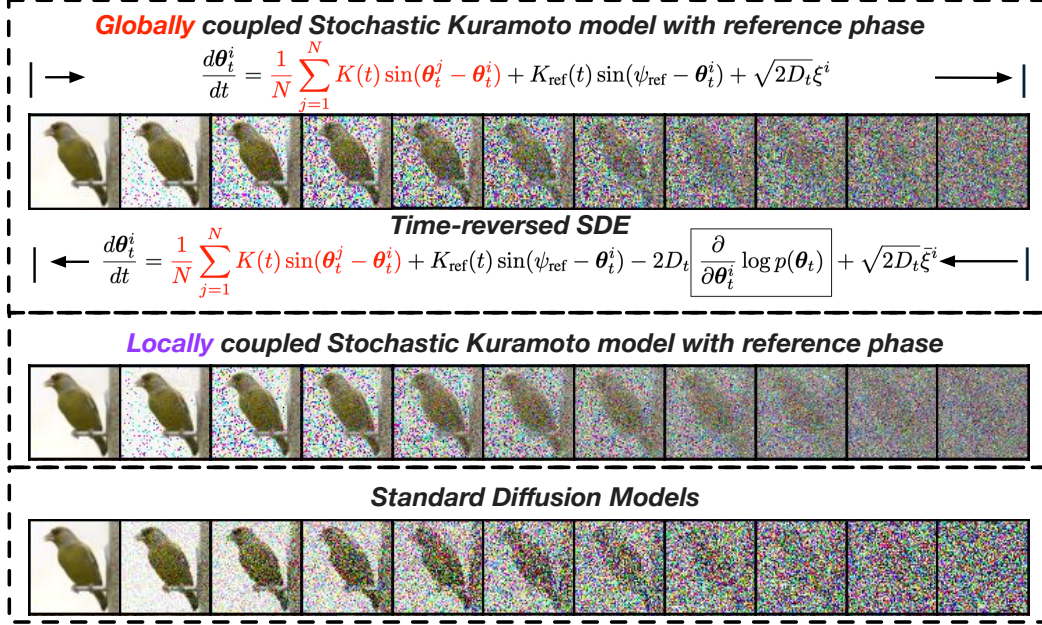


Figure 1: Governing stochastic differential equations (SDEs) and representative image samples from our globally and locally coupled Kuramoto orientation diffusion models. Pixels are mapped onto periodic domains as angular phase variables. In the globally coupled model, each pixel interacts with all other pixels via Kuramoto sinusoidal coupling (highlighted in red). The locally coupled variant corresponds to a similar SDE but restricts this sinusoidal coupling to a local neighborhood around each pixel. Unlike standard diffusion models, our approach introduces non-isotropic noise dynamics via pulling similar phases together, enabling a more structured destruction process. These dynamics help preserve the global structure in the early stages of diffusion (*e.g.*, the overall shape of the bird), while allowing for faster adaptation to noise as the process progresses. The forward SDE **synchronizes** phase variables through oscillator interactions and a global reference phase. The reverse process **desynchronizes** these variables using learned score functions to synthesize images.

signals with sharp transitions) rather than by raw pixel intensities. Crucially, such orientation patterns exist on periodic domains where angular discontinuities are problematic for conventional models that do not explicitly account for periodicity. Early attempts to handle these issues date back to the seminal work on orientation diffusions for image denoising [48], which highlighted how treating angular data without considering its circular nature can lead to artifacts and loss of coherence. Those findings underscore the need for specialized generative frameworks that explicitly account for the periodic structure inherent in directional data.

In this paper, we introduce a novel nonlinear score-based generative framework that leverages stochastic Kuramoto dynamics to operate directly on periodic domains. Synchronization offers a powerful inductive bias here: it encourages local patterns and orientations to reinforce one another – edges to align, ridges to remain coherent, and flows to stay smooth – before noise erodes these features. In our framework, pixel values are first mapped to angular phase variables, which enables natural compatibility with circular geometry and allows the diffusion process to evolve through interactions among phase coupling. We present our governing SDEs and illustrative examples of diffusion dynamics in Fig. 1. Our forward diffusion process performs *structured destruction* by progressively synchronizing angular phase variables through phase coupling and attraction to a common reference phase, driving the data distribution toward a low-entropy von Mises state. The Kuramoto interactions induce non-isotropic dynamics by pulling similar phases together, helping preserve local orientation and making the model particularly suitable for orientation-dense images.

The reverse generative process then performs desynchronization, leveraging learned periodic score functions to gradually reintroduce variability and reconstruct the image. The generation thus follows a hierarchical process, where the global structure is established first, and finer details are subsequently introduced, following a coarse-to-fine paradigm. We also propose a locally coupled variant of the

model that aligns with the spatial correlations of image data. Due to the coherence imposed by synchronization, our model converges faster to the terminal distribution during the forward process, which in turn enables the reverse process to generate high-quality samples within fewer diffusion steps. To handle periodicity robustly, we employ periodicity-aware neural networks with sinusoidal embeddings and define forward transitions using wrapped Gaussian kernels. The score function is then estimated by sampling from these local transitions derived from the forward dynamics.

Experiments on orientation-rich datasets, such as fingerprints, textures, and terrains, demonstrate that our Kuramoto orientation diffusion model consistently produces higher-fidelity images compared to standard diffusion baselines, often requiring fewer diffusion steps. Furthermore, our method remains competitive even on general CIFAR-10 benchmarks. Beyond images, we also report results on Earth/climate datasets on the 2D sphere and on Navier–Stokes fluid velocity fields, where the synchronization prior aligns with natural periodicity and angular structure, delivering consistent gains. Overall, this work bridges neural oscillation theory and modern score-based generative models, underscoring the potential of biologically inspired synchronization dynamics as structured priors.

2 Kuramoto Orientation Diffusion Models

2.1 Preliminary: Score-based Generative Models

In score-based generative models [3, 60], the forward and reverse processes of diffusion models are formulated as a pair of coupled SDEs:

$$\begin{aligned} \text{Forward-SDE: } d\mathbf{x} &= \mathbf{f}(\mathbf{x}, t) dt + g(t) d\mathbf{w} \\ \text{Reverse-SDE: } d\mathbf{x} &= [\mathbf{f}(\mathbf{x}, t) - g^2(t) \boxed{\nabla_{\mathbf{x}} \log p(\mathbf{x})}] dt + g(t) d\bar{\mathbf{w}} \end{aligned} \quad (1)$$

where $\mathbf{f}(\mathbf{x}, t)$ represents the vector-valued drift function, $g(t)$ denotes the scalar function of the diffusion coefficient, and $\mathbf{w}$ and $\bar{\mathbf{w}}$ are the standard Wiener processes. The boxed term represents the score function, which corresponds to the gradient of the log-density. A neural network $s(\mathbf{x}_t, t)$ is trained to approximate the score function by minimizing the following objective:

$$\mathcal{L} = \mathbb{E}_{t \sim U(1, T)} \mathbb{E}_{\mathbf{x}_0} \mathbb{E}_{\mathbf{x}_t \sim p(\mathbf{x}_t | \mathbf{x}_0)} \left[g^2(t) \left\| s(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_0) \right\|_2^2 \right] \quad (2)$$

Common choices for the forward SDE include the Variance-Preserving (VP) and Variance-Exploding (VE) formulations [59, 60, 23]. Typically, the drift function is chosen to be linear, ensuring that the conditional distribution $p(\mathbf{x}_t | \mathbf{x}_0)$ remains analytically Gaussian and the corresponding score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_0)$ can be computed in closed form.

2.2 Stochastic Kuramoto Models with Reference Phase

Fig. 2 illustrates the forward and reverse processes in our Kuramoto orientation diffusion model. During the forward process (left-to-right), we progressively destroy image information through synchronization, modeled by the following stochastic Kuramoto dynamics [36]:

$$\frac{d\theta_t^i}{dt} = \frac{1}{N} \sum_{j=1}^N K(t) \sin(\theta_t^j - \theta_t^i) + K_{\text{ref}}(t) \sin(\psi_{\text{ref}} - \theta_t^i) + \sqrt{2D_t} \xi^i \quad (3)$$

where each colored circle represents an oscillator with periodic phase $\theta_t^i \in [-\pi, \pi]$, $K(t)$ is the time-varying coupling strength among oscillators, $K_{\text{ref}}(t)$ is the coupling strength to a reference phase ψ_{ref} , $2D_t$ denotes the variance of Gaussian noise, and N denotes the number of oscillators. The white rectangle represents the global reference phase ψ_{ref} , serving as an attractor to guide the population of oscillators toward a synchronized target. To measure the level of synchronization, we compute the complex ordering parameter:

$$r(t) e^{i\psi(t)} = \frac{1}{N} \sum_{j=1}^N e^{i\theta_t^j} \quad (4)$$

where $\psi(t) \in [-\pi, \pi]$ denotes the mean phase, and $r(t) \in [0, 1]$ measures the coherence of these phases. They correspond to the angle and length of the arrow in Fig. 2, respectively. When $r = 0$,

the system is completely *desynchronized*. When $r = 1$, the oscillators are *perfectly synchronized*. For intermediate values $0 < r < 1$, the system exhibits *partial synchronization*, where phases are clustered around the mean phase $\psi(t)$ with dispersion induced by noise.

Phase Wrapping. After each Kuramoto update step, we apply phase wrapping to ensure that all phase variables remain within the interval $[-\pi, \pi]$. Specifically, the wrapping is performed as $\theta = (\theta + \pi) \bmod (2\pi) - \pi$ where $\bmod$ denotes the modulo operation. This guarantees that the periodicity of the phase variables is consistently maintained throughout the forward process.

Quasi-equilibrium. In the thermodynamic limit ($N \rightarrow \infty$), the collective dynamics of the oscillator population can be effectively described by the mean-field evolution of a single representative oscillator with phase θ . The mean-field dynamics of Eq. (3) follow the Fokker–Planck equation:

$$\frac{\partial p(\theta, t)}{\partial t} = -\frac{\partial}{\partial \theta} \left[(K(t)r(t) \sin(\psi(t) - \theta) + K_{\text{ref}}(t) \sin(\psi_{\text{ref}} - \theta)) p(\theta, t) \right] + D_t \frac{\partial^2 p(\theta, t)}{\partial \theta^2} \quad (5)$$

In our framework, akin to standard diffusion models, the coupling strengths and noise variance gradually increase over time. Consequently, the system does not reach a true stationary distribution. Instead, after an initial transient phase, the evolution of $p(\theta, t)$ becomes sufficiently slow such that, at each moment, the system can be approximated well by an instantaneous steady state – a regime we refer to as *quasi-equilibrium*, characterized by $\partial p_{\text{st}}(\theta)/\partial t \approx 0$.

Under quasi-equilibrium, the phase distribution approximately satisfies:

$$p_{\text{st}}(\theta) \approx \frac{1}{Z} \exp \left(\frac{K(T)r(T)}{D_T} \cos(\psi(T) - \theta) + \frac{K_{\text{ref}}(T)}{D_T} \cos(\psi_{\text{ref}} - \theta) \right) \quad (6)$$

where Z is the normalization constant, and T denotes the final timestep. The proof is given in the Appendix. In the long-time limit, as the average phase $\psi(T)$ synchronizes toward the reference phase ψ_{ref} , the distribution further simplifies to a well-known von Mises form (Gaussian on a circle):

$$\text{von Mises Distribution: } p_{\text{st}}(\theta) \approx \frac{1}{Z} \exp \left(\frac{K(T)r(T) + K_{\text{ref}}(T)}{D_T} \cos(\psi_{\text{ref}} - \theta) \right) \quad (7)$$

Local Coupling. A natural variant of the Kuramoto model is to introduce local coupling, where each oscillator interacts only with its neighboring oscillators:

$$\frac{d\theta_t^i}{dt} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} K(t) \sin(\theta_t^j - \theta_t^i) + K_{\text{ref}}(t) \sin(\psi_{\text{ref}} - \theta_t^i) + \sqrt{2D_t} \xi^i \quad (8)$$

where $\mathcal{N}_i$ denotes the set of neighbors for the i -th oscillator. In contrast to the globally coupled case, local interactions introduce spatial inhomogeneity and break the mean-field approximation. As a result, the system exhibits diffusion-like behavior, with oscillatory phases progressively smoothing out over space and time. This manifests as blurring effects during the forward process, analogous to phenomena observed in heat dissipation models [51] and blurring diffusion models [24].

Interestingly, while local coupling prevents analytical simplification of the dynamics, the strong reference phase (via K_{ref}) still guides the system toward global synchronization. After sufficiently long evolution, the terminal distribution concentrates around the reference phase and can be effectively approximated by a von Mises distribution, despite the absence of mean-field symmetry.

Coupling Strength and Noise Variance. In the forward process, we maintain the relationship $K_{\text{ref}}(t) > D_t > K(t)$ to balance structure and noise. $D_t > K(t)$ ensures that the stochastic noise is strong enough to disrupt image details, while $K_{\text{ref}}(t) > D_t$ guarantees that the stronger reference coupling dominates over noise and local/global coupling to guide the system toward synchronization.

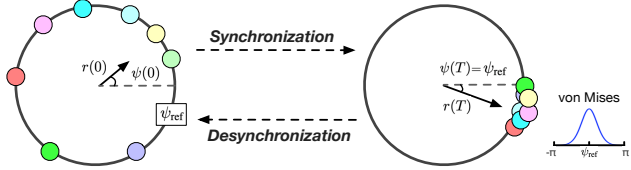


Figure 2: Illustration of our Kuramoto orientation diffusion model. In the forward process (left-to-right), angular phase variables (colored circles) synchronize toward a low-entropy von Mises distribution, guided by attraction to a reference phase (white rectangle). The reverse process (right-to-left) uses learned score functions to desynchronize these phases.

Algorithm 1 Training algorithm for Kuramoto orientation diffusion models.

Require: Score prediction network $s(\cdot, \cdot)$, noise variance schedule $\{2D_t\}_{t=0}^{T-1}$, forward Kuramoto drift $f(\cdot, \cdot)$, noise $\epsilon \sim \mathcal{N}(0, I)$, number of MC samples M .

```
1: repeat
2:   Sample initial phase variable:  $\theta_0 \sim p(\theta_0)$ 
3:   Sample a timestep:  $t \sim U(1, T)$ 
4:   Simulate the forward Markov chain to obtain  $\theta_{t-1} \sim p(\theta_{t-1}|\theta_0)$ 
5:   Initialize sample counter  $m = 0$ 
6:   while  $m < M$  do
7:     Sample:  $\theta_t^m = \theta_{t-1} + f(\theta_{t-1}, t-1) + \sqrt{2D_{t-1}}\epsilon$ 
8:     Wrap phase:  $\theta_t^m = (\theta_t^m + \pi) \bmod (2\pi) - \pi$ 
9:     Compute local score:  $\nabla_{\theta_t^m} \log p(\theta_t^m|\theta_{t-1})$ 
10:     $m \leftarrow m + 1$ 
11:  end while
12:  Compute training loss:  $\mathcal{L} = \frac{1}{M} \sum_{m=0}^{M-1} \left( 2D_t \left\| s(\theta_t^m, t) - \nabla_{\theta_t^m} p(\theta_t^m|\theta_{t-1}) \right\|^2 \right)$ 
13: until converged
```

2.3 Learning the Score Function

In the reverse process (right-to-left in Fig. 2), we perform desynchronization guided by learned score functions, progressively restoring image complexity from a synchronized von Mises distribution back to diverse angular states. Below we illustrate how the score function is learned.

Local Score Matching. In our stochastic Kuramoto models, the presence of the nonlinear drift renders the marginal distribution $p(\theta_t)$ intractable, and consequently the score function $\nabla_{\theta_t} \log p(\theta_t)$ is also unavailable in closed form. As a result, typical score-matching algorithms that rely on direct access to the marginal density cannot be applied. Nevertheless, we can still train a score network by exploiting the local Markov transition kernel $p(\theta_t|\theta_{t-1})$, based on the following general identity:

$$\nabla_{\theta_t} \log p(\theta_t) = \mathbb{E}_{\theta_{t-1} \sim p(\theta_{t-1}|\theta_t)} \left[\nabla_{\theta_t} \log p(\theta_t|\theta_{t-1}) \right] \quad (9)$$

where the expectation is taken over the reverse transition $p(\theta_{t-1}|\theta_t)$. A detailed derivation of this identity is provided in the Appendix. Although the reverse transition is intractable in practice, Denoising Score Matching (DSM) [65] shows that the score function can be learned by sampling from the forward transition $p(\theta_t|\theta_{t-1})$ instead. Specifically, the training objective can be written as:

$$\mathbb{E}_{t \sim U(1, T)} \mathbb{E}_{\theta_0} \mathbb{E}_{\theta_{t-1} \sim p(\theta_{t-1}|\theta_0)} \mathbb{E}_{\theta_t \sim p(\theta_t|\theta_{t-1})} \left[2D_t \left\| s(\theta_t, t) - \nabla \log p(\theta_t|\theta_{t-1}) \right\|^2 \right] \quad (10)$$

where $s(\theta_t, t)$ denotes the score network being optimized. At each training step, we simulate the forward Markov chain to obtain θ_{t-1} , and then sample multiple θ_t from the local transition kernel $p(\theta_t|\theta_{t-1})$ to estimate the training loss via Monte Carlo (MC) approximations.

Wrapped Gaussian Transition. Due to the phase wrapping on a periodic domain, the local transition probability $p(\theta_t|\theta_{t-1})$ follows a wrapped Gaussian distribution:

$$\begin{aligned} p(\theta_t|\theta_{t-1}) &= \mathcal{WN}(\theta_{t-1} + f(\theta_{t-1}, t-1), 2D_{t-1}I) \\ &\approx \frac{1}{\sqrt{4\pi D_{t-1}}} \sum_{k=-K}^K \exp \frac{-\left(\theta_t - \theta_{t-1} - f(\theta_{t-1}, t-1) + 2\pi k\right)^2}{4D_{t-1}} \end{aligned} \quad (11)$$

where $f(\theta_t, t)$ denotes the forward drift of the Kuramoto model, *i.e.*, the coupling and reference terms weighted by the coupling strength. Since the wrapped Gaussian involves an infinite series, neither the transition density nor its score function admits a simple closed-form. We thus approximate the transition by truncating the summation to a small finite number of terms K .

Periodicity-aware Networks. To incorporate the inherent periodicity of the phase variables into the score network, we embed the input phases using sinusoidal features $[\sin(\theta), \cos(\theta)]$ as input to the

Table 1: FID Scores ($\downarrow$) on SOCOFing fingerprint dataset [56].

Diffusion Steps	100	300	1000
SGM [60]	104.92	62.66	23.84
Kuramoto Orientation Diffusion (Globally Coupled)	74.41	47.93	20.64
Kuramoto Orientation Diffusion (Locally Coupled)	67.49	43.57	18.75

Table 2: FID Scores ($\downarrow$) on Brodatz texture dataset [1, 8].

Steps	100	300	1000
SGM [60]	38.33	22.40	20.37
Kuramoto Orientation Diffusion (Globally Coupled)	20.26	18.51	15.42
Kuramoto Orientation Diffusion (Locally Coupled)	18.47	15.93	14.19

network. The network predicts two outputs $[s_1(\theta, t), s_2(\theta, t)]$ corresponding to the two Cartesian components. We then project the output back onto the angular domain via:

$$s(\theta, t) = s_1(\theta, t) \cos(\theta) + s_2(\theta, t) \sin(\theta) \quad (12)$$

This operation ensures that the score prediction respects the circular geometry of the phase space.

Training Algorithms. We summarize the training algorithms in Alg. 1. To adapt non-periodic image data to the periodic phase domain, input pixels in the range $[-1, 1]$ are linearly mapped to $[-0.9\pi, 0.9\pi]$. This margin near the boundaries helps distinguish values near -1 and 1 , which would otherwise collapse under phase wrapping. Throughout both training and inference, we enforce the periodic geometry by wrapping all phase variables into the interval $[-\pi, \pi]$ after each SDE step. Our training procedure relies on estimating local scores using Monte Carlo samples from truncated wrapped Gaussian transitions. We find that using $K=3$ and $M=5$ samples per step provides a good balance between training stability and computational efficiency.

3 Experiments

3.1 Setup

Datasets, Baselines, and Metrics. We evaluate our proposed method across both general-purpose and orientation-rich image generation tasks. For standard benchmarking, we first test on CIFAR10 [35]. To specifically assess performance on orientation-dense data, we then apply our model to the SOCOFing fingerprint dataset [56], the Brodatz texture dataset [1, 8], and the ground terrain dataset [68]. The input image resolutions are $3 \times 32 \times 32$ for CIFAR10 and $1 \times 96 \times 96$ for SOCOFing. The Brodatz

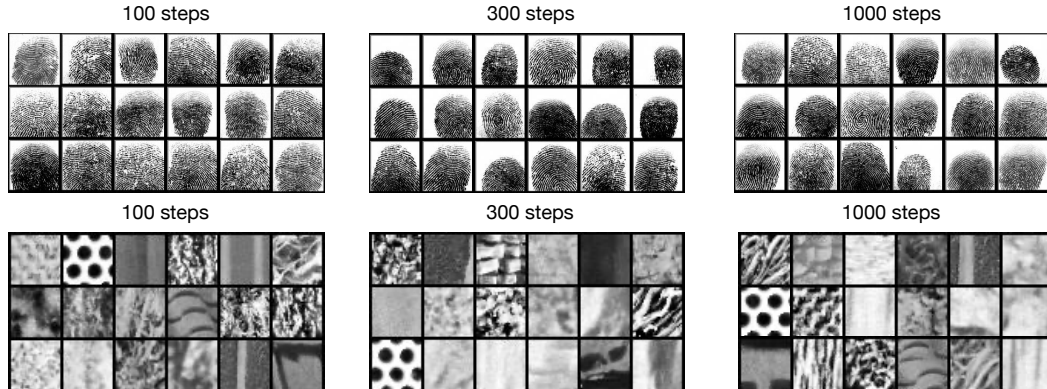


Figure 3: Samples generated by our Kuramoto orientation diffusion model on SOCOFing fingerprint and Brodatz texture datasets under varying denoising steps.

Table 3: FID Scores ($\downarrow$) on the ground terrain dataset [68].

Diffusion Steps	100	300	1000
SGM [60]	114.90	56.72	33.79
Kuramoto Orientation Diffusion (Globally Coupled)	101.65	54.17	33.56
Kuramoto Orientation Diffusion (Locally Coupled)	92.86	49.68	30.62



Figure 4: Samples generated by our Kuramoto model on the ground terrain dataset.

texture dataset originally consists of high-resolution samples depicting various textures (*e.g.*, grass, water, sand, wool) with sizes of $1 \times 512 \times 512$ or $1 \times 1024 \times 1024$. Due to the limited number of available images, we increase the dataset size by dividing these high-resolution textures into smaller patches of $1 \times 32 \times 32$. Compared to Brodatz textures, the ground terrain dataset [68] contains a broader set of orientation-rich material textures (*e.g.*, plastic, turf, steel, asphalt, leaves, stone, and brick) at the higher resolution of 128×128 . We primarily compare with the standard variance-preserving (VP) score-based generative model (SGM) [60]. To evaluate the quality of generated images, we use the Fréchet Inception Distance (FID) [21], a widely used metric for assessing visual fidelity and diversity. All models are trained until the FID score converges to ensure fair comparison.

In the Supplementary, we extend evaluation beyond images to (i) Earth and climate science datasets on the 2D Sphere [46, 45, 5, 18], and (ii) Navier-Stokes fluid velocity fields [6]. The former suite of datasets is defined on *naturally periodic grids*, while the latter one constitutes *intrinsically angular data*. Together with the image datasets, the benefit of the synchronization inductive bias built into our model will be consistently validated across three distinct domains – standard orientation-dense images (with pixels mapped to phases), spherical geophysical fields, and fluid velocity phases.

Timesteps. Due to the phase coupling in our Kuramoto diffusion model, the forward process leads to faster convergence toward the terminal distribution compared to conventional diffusion models. To explore this advantage, we evaluate our model under varying diffusion step counts: 100, 300, and 1000 steps. These comparisons reveal how efficiently our model captures structural information with fewer steps. Details of the noise and coupling schedules are provided in the Supplementary.

3.2 Results

Fingerprints and Textures. Tables 1 and 2 report FID scores on the SOCOFing fingerprint and Brodatz texture datasets, respectively. Across all diffusion step settings, both globally and locally coupled variants of our Kuramoto orientation diffusion model consistently outperform the standard SGM [60]. The improvement is especially notable on the Brodatz texture dataset, where our 100-step Kuramoto model achieves performance comparable to or better than SGM using 1000 steps – demonstrating a substantial gain in sampling efficiency. These results underscore the advantage of incorporating structured coupling dynamics for modeling orientation-dense data. The synchronization-based inductive bias helps preserve coherent directional patterns, which are critical for the perceptual quality of textures and fingerprints. Notably, the locally coupled variant offers further improvements by aligning with the local spatial correlations. Fig. 3 provides qualitative examples, illustrating that our method produces sharp and consistent patterns under varying diffusion steps.

Ground Terrain. Table 3 reports FID on the ground terrain dataset [68], comparing our Kuramoto diffusion model with SGM. The trend mirrors fingerprints and textures: across denoising step counts, our method consistently attains lower FID. This indicates that the synchronization prior continues to help at orientation-dense scenes of higher resolutions. Qualitative results in Fig. 4 illustrate that,

Table 4: FID Scores ($\downarrow$) on CIFAR10 [35].

Diffusion Steps	100	300	1000
SGM [60]	38.04	25.76	3.17
Kuramoto Orientation Diffusion (Globally Coupled)	29.96	25.83	11.58
Kuramoto Orientation Diffusion (Locally Coupled)	28.17	24.86	10.79



Figure 5: Samples generated by our Kuramoto model on CIFAR10 under varying denoising steps.

even under fewer denoising steps, our samples exhibit coherent directional structure and material appearance, while additional steps further refine detail realism and texture alignment.

CIFAR10. Table 4 presents the FID scores for our Kuramoto orientation diffusion models against SGM on CIFAR10. At 100 diffusion steps, both Kuramoto models substantially outperform SGM, highlighting the effectiveness of structured synchronization as an inductive bias in low-step regimes. At 300 steps, Kuramoto models achieve comparable or slightly better performance than SGM, demonstrating their ability to maintain sample quality as diffusion progresses. At 1000 steps, SGM achieves the best overall score, though both Kuramoto models remain competitive.

Fig. 5 displays generated samples under different step counts. As diffusion steps increase, the benefits of structured synchronization become less prominent. We expect that in longer trajectories, the structured reverse dynamics may inhibit the flexibility to fully capture fine-grained details, especially in datasets that lack strong orientation priors. These results suggest that Kuramoto-based synchronization dynamics are particularly advantageous in generating high-quality images within limited steps, even on general-purpose datasets. However, this structured bias may slightly limit expressiveness under excessive denoising steps on natural images lacking strong orientation patterns.

3.3 Discussions

Structured Destruction. Our Kuramoto forward process introduces a very structured destruction process, leveraging either global or local coupling mechanisms to achieve more controlled noising dynamics. Unlike conventional isotropic diffusion which quickly loses object structure, our model preserves the objects in the early stages through synchronized coupling (see also Fig. 1). This structured noising progressively aligns similar phase variables while maintaining orientation consistency, allowing the model to better retain structural information. As noise levels increase, the coupling interactions expedite the convergence to the noise distribution, enabling a faster transition compared to standard diffusion models. This unique dual-phase dynamic – **initial structured synchronization followed by rapid noise adaptation** – allows our model to converge faster while maintaining orientation coherence. The empirical SNR plot in Fig. 6 clearly demonstrates the advantage described above, highlighting how the coupling-driven dynamics allow for efficient and structured noise progression.

Hierarchical Generation Process. Figs. 7 and 8 depict the hierarchical generation process of our Kuramoto orientation diffusion model. The generation starts from a synchronized state sampled from the von Mises distribution, representing a low-entropy configuration with highly aligned phase variables. In the reverse process, the large-scale structure is established first, as global coherence from the initial synchronized state is preserved. This occurs because the forward diffusion inherently maintains global coherence in the later stages, which correspond to the early stages in the reverse process. Once the primary structure is set, finer-scale details gradually emerge through successive diffusion steps, driven by anti-coupling dynamics that introduce local variability, allowing the image to evolve into more complex and nuanced patterns. Our model follows a coarse-to-fine paradigm, where the global structure remains consistent while localized, intricate features evolve flexibly.

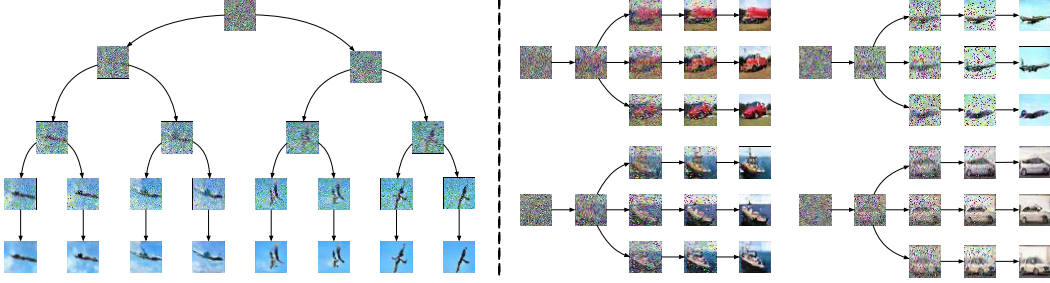


Figure 7: Hierarchical generation process of our locally coupled Kuramoto diffusion model applied to CIFAR-10. The generation follows a coarse-to-fine progression, starting from a structured von Mises sample and progressively adding fine-scale details. These results correspond to 100 diffusion steps.

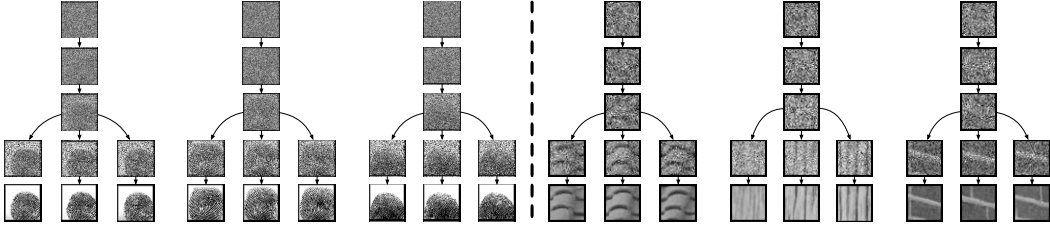


Figure 8: Hierarchical generation of our locally coupled Kuramoto diffusion model on the SOCOFing fingerprint dataset (left) and Brodatz texture dataset (right). The model first establishes large-scale orientation patterns, followed by finer texture details. These results correspond to 300 diffusion steps.

The hierarchical nature of the Kuramoto diffusion enhances interpretability by providing a clear generative pathway from **global coherence to local variability**. This hierarchical generation approach aligns with the spectral biases observed in coarse-to-fine generation of diffusion models [34, 51]. In typical score-based diffusion models, this hierarchy arises implicitly from the progressive noise attenuation, preserving low-frequency components longer while high-frequency details emerge later. In contrast, Rissanen *et al.* [51] use heat equations for the forward process, resulting in isotropic blurring as high-frequency components dissipate. Our Kuramoto model explicitly encodes hierarchical progression through synchronization dynamics. This structured, non-isotropic phase alignment offers a more interpretable pathway from global to local patterns.

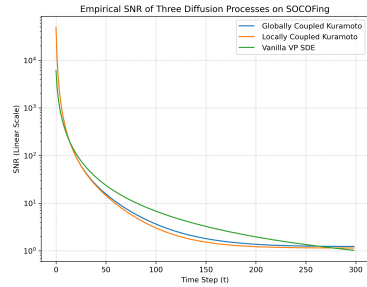


Figure 6: Empirical SNR over time.

4 Related Work

Generative Models. Deep generative models have made remarkable progress in the last decade, both in the creation of powerful foundation models [52] and in the design of sample-efficient optimization algorithms [66, 49]. Early advances in deep generative modeling, such as Variational Autoencoders (VAEs) [33] and Generative Adversarial Networks (GANs) [19], pushed the frontiers of generative capabilities but faced distinct limitations: VAEs often produced blurry samples, while GANs suffered from instability in training. Recently, diffusion models [57, 23, 60, 58, 13, 25, 32] have emerged as a powerful, principled alternative. These models learn to reverse a noising process by estimating the score function $\nabla_x \log p(x_t)$ of intermediate Gaussian-corrupted distributions. Building on the view of generative modeling as a continuous transformation from noise to data, Flow Matching [38] and Rectified Flow [39] directly learn velocity fields connecting noise and data without relying on stochastic diffusion, sidestepping the need for score estimation. Stochastic Interpolants [2] further generalize diffusion and flow models by learning stochastic trajectories interpolating between the data and prior, providing a unifying perspective across two modeling paradigms. Several recent

efforts extend flow matching and diffusion models to non-Euclidean geometries. For example, Riemannian Flow Matching [10] formulates generative models directly on Riemannian manifolds, respecting underlying geometric constraints. Other Riemannian diffusion models account for manifold curvature to enable generative modeling over geometries such as hyperspheres, tori, and hyperbolic spaces [11, 37, 26, 12, 69, 28, 14]. These advances highlight the growing interest in incorporating geometric inductive biases into generative modeling.

Our work contributes to this landscape by exploring nonlinear diffusion models grounded in stochastic Kuramoto dynamics. Instead of linear drifts in conventional diffusion models, we introduce structured phase dynamics driven by nonlinear Kuramoto coupling, evolving on the periodic domain. Unlike classical score-based models that assume a Gaussian forward process, our framework leads to non-Gaussian, wrapped distributions that evolve according to nonlinear SDEs. This extends recent interest in geometry-aware generative models and interpretable generation process.

Neural Oscillations in Machine Learning. Oscillatory dynamics are central to understanding biological neural systems, where phase-locked rhythms and traveling waves are believed to support functions such as sensory binding, working memory, and attention [9, 42, 16]. Motivated by their computational relevance, recent machine learning research has increasingly incorporated oscillatory and wave-based dynamics into neural representations, treating them as inductive biases that promote generalization and structured behaviors [54, 55, 15, 61, 30, 31]. A key focus has been on **synchronization**, where multiple units align their oscillatory phases to form stable or emergent patterns. This phenomenon has been studied through the lens of the Kuramoto model [36], a classical framework from nonlinear dynamics that describes how populations of coupled oscillators evolve toward synchronization. In computational neuroscience and machine learning, Kuramoto-inspired models have been applied to tasks such as modeling neural connectivity [7], clustering through emergent synchrony [50], and mitigating over-smoothing in graph neural networks [44]. A recent notable example is the Artificial Kuramoto Oscillatory Neuron (AKOrN) framework [41], which replaces thresholding units with oscillatory ones to bind neurons together through synchronization dynamics. These efforts reflect the growing interest in integrating principles of neural oscillations and synchronization into machine learning frameworks, offering novel perspectives and tools for modeling complex, dynamic systems.

5 Conclusion

This paper introduces a nonlinear score-based generative framework that incorporates stochastic Kuramoto dynamics to model orientation-rich data. By formulating the forward process as synchronization and the reverse process as desynchronization, our method brings biologically inspired inductive biases into the diffusion process. Through wrapped Gaussian transitions and periodicity-aware networks, the model naturally captures the geometry of angular data. Experiments show that our approach outperforms conventional baselines on orientation-dense datasets and remains competitive on general image generation tasks. This work highlights the potential of integrating biologically inspired synchronization dynamics as structured priors in generative modeling of orientation-dense data, paving the way for incorporating nonlinear dynamics into generative models.

Limitations and Future Work. A primary limitation lies in the training efficiency: each training step incurs an $\mathcal{O}(T)$ time cost due to the need to explicitly simulate the forward Markov chain. Actually, this cost actually can be nearly eliminated with **pre-computation & cache**: before training, we can run the forward SDE on the entire dataset, save all pairs to disk, and then load them directly during training. This makes the simulation cost effectively $\mathcal{O}(0)$ at each training step. Otherwise if disk space is limited, we can simply re-generate and cache one epoch’s worth of pairs at the start of each epoch. This still dramatically reduces the simulation overhead.

Beyond computational improvements, applying this framework to neural spiking data offers a compelling opportunity to further explore the biological plausibility of synchronization-driven generative models and to bridge the gap between theoretical neuroscience and machine learning.

Broader Impacts. Our approach may benefit socially relevant domains such as biometric security (e.g., synthetic fingerprint generation), medical imaging (e.g., fiber orientation modeling in MRI), and scientific visualization (e.g., materials analysis). As with other generative technologies, we advocate for responsible research and deployment in accordance with ethical and societal guidelines.

Acknowledgments

We would like to thank anonymous reviewers for their constructive suggestions and feedback. Yue Song was supported by the gift donation from Cisco. T. Anderson Keller was supported by the Kempner Institute Research Fellowship. Takeru Miyato was supported by the ERC Starting Grant LEGO-3D (850533) and the German Research Foundation (DFG): SFB 1233, Robust Vision: Inference Principles and Neural Mechanisms, project number: 276693517.

References

- [1] Usc-sipi image database website. <https://sipi.usc.edu/database/database.php?volume=textures>.
- [2] Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.
- [3] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 1982.
- [4] Heli Ben-Hamu, Samuel Cohen, Joey Bose, Brandon Amos, Maximillian Nickel, Aditya Grover, Ricky TQ Chen, and Yaron Lipman. Matching normalizing flows and probability paths on manifolds. In *ICML*. PMLR, 2022.
- [5] G. R. Brakenridge. Global Active Archive of Large Flood Events. <http://floodobservatory.colorado.edu/Archives/index.html>, 2017. Dartmouth Flood Observatory, University of Colorado.
- [6] Johannes Brandstetter, Rianne van den Berg, Max Welling, and Jayesh K Gupta. Clifford neural layers for pde modeling. *ICLR*, 2023.
- [7] Michael Breakspear, Stewart Heitmann, and Andreas Daffertshofer. Generative models of cortical oscillations: neurobiological implications of the kuramoto model. *Frontiers in human neuroscience*, 2010.
- [8] Phil Brodatz. Textures: a photographic album for artists and designers. 1966.
- [9] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford university press, 2006.
- [10] Ricky TQ Chen and Yaron Lipman. Riemannian flow matching on general geometries. *ICLR*, 2024.
- [11] Tim R Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M Tomczak. Hyperspherical variational auto-encoders. *UAI*, 2018.
- [12] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *NeurIPS*, 2022.
- [13] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *NeurIPS*, 2021.
- [14] Muskan Dosi, Chiranjeev Chiranjeev, Kartik Thakral, Mayank Vatsa, and Richa Singh. Harmonizing geometry and uncertainty: Diffusion with hyperspheres. *ICML*, 2025.
- [15] Felix Effenberger, Pedro Carvalho, Igor Dubinin, and Wolf Singer. A biology-inspired recurrent oscillator network for computations in high-dimensional state space. *bioRxiv*, 2022.
- [16] Felix Effenberger, Pedro Carvalho, Igor Dubinin, and Wolf Singer. The functional role of oscillatory dynamics in neocortical circuits: a computational perspective. *Proceedings of the National Academy of Sciences*, 2025.
- [17] Joshua J Engelsma, Debayan Deb, and Anil K Jain. Printsgan: Synthetic fingerprint generator. *arXiv preprint arXiv:2202.10571*, 2022.
- [18] EOSDIS. Active Fire Data. <https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data>, 2020. Land, Atmosphere Near real-time Capability for EOS (LANCE) system operated by NASA’s Earth Science Data and Information System (ESDIS).
- [19] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NeurIPS*, 2014.
- [20] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv*, 2016.

- [21] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NeurIPS*, 2017.
- [22] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop*.
- [23] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 2020.
- [24] Emiel Hoogetboom and Tim Salimans. Blurring diffusion models. *ICLR*, 2023.
- [25] Chin-Wei Huang, Jae Hyun Lim, and Aaron C Courville. A variational perspective on diffusion-based generative models and score matching. *NeurIPS*, 2021.
- [26] Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron C Courville. Riemannian diffusion models. *NeurIPS*, 2022.
- [27] Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 1989.
- [28] Yesukhei Jagvaral, Francois Lanusse, and Rachel Mandelbaum. Unified framework for diffusion generative models in so (3): applications in computer vision and astrophysics. In *AAAI*, 2024.
- [29] Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking fid: Towards a better evaluation metric for image generation. In *CVPR*, 2024.
- [30] T Anderson Keller and Max Welling. Neural wave machines: learning spatiotemporally structured representations with locally coupled oscillatory recurrent neural networks. In *ICML*. PMLR, 2023.
- [31] T Anderson Keller, Lyle Muller, Terrence Sejnowski, and Max Welling. Traveling waves encode the recent past and enhance sequence learning. *ICLR*, 2024.
- [32] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *NeurIPS*, 2021.
- [33] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *ICLR*, 2013.
- [34] Karsten Kreis, Ruiqi Gao, and Arash Vahdat. Tutorial on denoising diffusion-based generative modeling: Foundations and applications. In *CVPR*, 2022.
- [35] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
- [36] Yoshiki Kuramoto. *Chemical turbulence*. Springer, 1984.
- [37] Adam Leach, Sebastian M Schmon, Matteo T Degiacomi, and Chris G Willcocks. Denoising diffusion probabilistic models on so (3) for rotational alignment. In *ICLR 2022 Workshop*, 2022.
- [38] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *ICLR*, 2023.
- [39] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *ICLR*, 2023.
- [40] Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. *NeurIPS*, 2020.
- [41] Takeru Miyato, Sindy Löwe, Andreas Geiger, and Max Welling. Artificial kuramoto oscillatory neurons. *ICLR*, 2025.
- [42] Lyle Muller, Frédéric Chavane, John Reynolds, and Terrence J Sejnowski. Cortical travelling waves: mechanisms and computational principles. *Nature Reviews Neuroscience*, 19(5):255–268, 2018.
- [43] H et al. Murgas. Generative adversarial networks for virtual microstructure generation of cold spray aluminum alloys. *npj Computational Materials*, 2024.
- [44] Tuan Nguyen, Hirotada Honda, Takashi Sano, Vinh Nguyen, Shugo Nakamura, and Tan Minh Nguyen. From coupled oscillators to graph neural networks: Reducing over-smoothing via a kuramoto model-based approach. In *AISTATS*. PMLR, 2024.
- [45] NOAA National Centers for Environmental Information. Global Significant Earthquake Database. <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.hazards:G012153>, 2020. National Geophysical Data Center / World Data Service (NGDC/WDS): NCEI/WDS Global Significant Earthquake Database.

- [46] NOAA National Centers for Environmental Information. Global Significant Volcanic Eruptions Database. <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.hazards:G10147>, 2020. National Geophysical Data Center / World Data Service (NGDC/WDS): NCEI/WDS Global Significant Volcanic Eruptions Database.
- [47] Clement et al. Papadacci. Ultrasound backscatter tensor imaging (bti): analysis of fiber orientation. *Scientific Reports*, 7, 2017.
- [48] Pietro Perona. Orientation diffusions. *IEEE Transactions on Image processing*, 1998.
- [49] Yun Qu, Qi Wang, Yixiu Mao, Vincent Tao Hu, Björn Ommer, and Xiangyang Ji. Can prompt difficulty be online predicted for accelerating rl finetuning of reasoning models? *arXiv preprint arXiv:2507.04632*, 2025.
- [50] Matthew Ricci, Minju Jung, Yuwei Zhang, Mathieu Chalvidal, Aneri Soni, and Thomas Serre. Kuranet: systems of coupled oscillators that learn to synchronize. *arXiv preprint arXiv:2105.02838*, 2021.
- [51] Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. *ICLR*, 2023.
- [52] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022.
- [53] Noam Rozen, Aditya Grover, Maximilian Nickel, and Yaron Lipman. Moser flow: Divergence-based generative modeling on manifolds. *NeurIPS*, 2021.
- [54] T Konstantin Rusch and Siddhartha Mishra. Coupled oscillatory recurrent neural network (cornn): An accurate and (gradient) stable architecture for learning long time dependencies. *ICLR*, 2021.
- [55] T Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *ICML*. PMLR, 2022.
- [56] Yahaya Isah Shehu, Ariel Ruiz-Garcia, Vasile Palade, and Anne James. Sokoto coventry fingerprint dataset. *arXiv preprint arXiv:1807.10609*, 2018.
- [57] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*. pmlr, 2015.
- [58] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *ICLR*, 2021.
- [59] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *NeurIPS*, 2019.
- [60] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *ICLR*, 2021.
- [61] Yue Song, T Anderson Keller, Nicu Sebe, and Max Welling. Latent traversals in generative models as potential flows. *ICML*, 2023.
- [62] Steven H Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 2000.
- [63] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 2017.
- [64] Sebastian et al. Vellmer. Diffusion mri-based fiber orientation distribution synthesis using gan-vae hybrid models. *Communications Biology*, 2025.
- [65] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.
- [66] Qi Cheems Wang, Zehao Xiao, Yixiu Mao, Yun Qu, Jiayi Shen, Yiqin Lv, and Xiangyang Ji. Model predictive task sampling for efficient and robust adaptation. *arXiv preprint arXiv:2501.11039*, 2025.
- [67] Yuxin Wu and Kaiming He. Group normalization. In *ECCV*, 2018.
- [68] Jia Xue, Hang Zhang, and Kristin Dana. Deep texture manifold for ground terrain recognition. In *CVPR*, 2018.
- [69] Sebastian Zeng, Florian Graf, and Roland Kwitt. Latent sdes on homogeneous spaces. *NeurIPS*, 2023.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#) .

Justification: The main claims are validated in the method and experiment sections.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: We explicitly have a paragraph about limitations at the beginning of the Appendix.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#) .

Justification: The assumptions are listed and the proofs are in the Appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#) .

Justification: All the implementation details are released to guarantee the reproducibility.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes] .

Justification: The code is released in a GitHub link. All used datasets are publicly available.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes] .

Justification: The key implementation details are given in the main paper, while the rest details are given in the Appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No] .

Justification: The main evaluation metric is the FID score, which usually does not vary much for different training runs.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [\[Yes\]](#) .

Justification: We report the used GPUs in the implementation details of the Appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [\[Yes\]](#) .

Justification: The research respect the NeurIPS code of ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [\[Yes\]](#) .

Justification: We have discussed the broader impacts in the Appendix.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA] .

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes] .

Justification: All used existing assets get cited properly.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset’s creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA] .

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA] .

Justification: The paper does not involve any research with crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA] .

Justification: The paper does not involve any research with crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA] .

Justification: The development of the core methodology does not involve any usages of LLMs.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

A	Math Derivations and Intuitions	21
A.1	Mean-field Dynamics of Fokker-Planck Equation for Stochastic Kuramoto Models	21
A.2	Local Score Matching	22
A.3	Fourier Interpretation of Kuramoto Coupling	22
A.4	Why Kuramoto Diffusion Falls Short on General Natural Images	22
B	More Experimental Results	22
B.1	Implementation Details	22
B.2	Ablation Study and Alternative Metrics	23
B.3	Earth and Climate Science Datasets on Spheres	24
B.4	Navier-Stokes Fluid Velocity Field	25
B.5	More Examples of Generative Samples	26

A Math Derivations and Intuitions

A.1 Mean-field Dynamics of Fokker-Planck Equation for Stochastic Kuramoto Models

In the thermodynamic limit of infinite oscillators ($N \rightarrow \infty$), the collective dynamics of the oscillator population can be effectively described by the mean-field evolution of a single representative oscillator with phase variable θ . The corresponding probability density $p(\theta, t)$ evolves according to the Fokker-Planck equation:

$$\frac{\partial p(\theta, t)}{\partial t} = -\frac{\partial}{\partial \theta} \left[(K(t)r(t) \sin(\psi(t) - \theta) + K_{\text{ref}}(t) \sin(\psi_{\text{ref}} - \theta)) p(\theta, t) \right] + D_t \frac{\partial^2 p(\theta, t)}{\partial \theta^2} \quad (13)$$

where $r(t)$ and $\psi(t)$ denote the magnitude and phase of the complex order parameter, respectively. In the long-time limit, the quasi-equilibrium satisfies $\partial p_{\text{st}}(\theta)/\partial t \approx 0$. At this point, the stationary solution $p(\theta, T)$ approximately follows:

$$\begin{aligned} D_T \frac{\partial^2 p(\theta, T)}{\partial \theta^2} &\approx \frac{\partial}{\partial \theta} \left[(K(T)r(T) \sin(\psi(T) - \theta) + K_{\text{ref}}(T) \sin(\psi_{\text{ref}} - \theta)) p(\theta, T) \right] \\ D_T \frac{\partial p(\theta, T)}{\partial \theta} &\approx (K(T)r(T) \sin(\psi - \theta) + K_{\text{ref}}(T) \sin(\psi_{\text{ref}} - \theta)) p(\theta, T) + C_1 \\ \frac{1}{p(\theta, T)} \frac{\partial p(\theta, T)}{\partial \theta} &\approx \frac{K(T)r(T) \sin(\psi(T) - \theta) + K_{\text{ref}}(T) \sin(\psi_{\text{ref}} - \theta)}{D_T} \\ \log p(\theta, T) &\approx \frac{K(T)r(T)}{D_T} \cos(\psi(T) - \theta) + \frac{K_{\text{ref}}(T)}{D_T} \cos(\psi_{\text{ref}} - \theta) + C_2 \\ p(\theta, T) &\approx \frac{1}{Z} \exp\left(\frac{K(T)r(T)}{D_T} \cos(\psi(T) - \theta) + \frac{K_{\text{ref}}(T)}{D_T} \cos(\psi_{\text{ref}} - \theta)\right) \end{aligned} \quad (14)$$

where C_1, C_2 are constants, and Z is the normalization constant. We assume $C_1 = 0$ because the distribution is periodic, and C_2 is absorbed into Z . Since the average phase $\psi(T)$ will synchronize to the reference ψ_{ref} , the quasi-stationary solution is given by:

$$p_{\text{st}}(\theta) \approx \frac{1}{Z} \exp\left(\frac{K(T)r(T) + K_{\text{ref}}(T)}{D_T} \cos(\psi_{\text{ref}} - \theta)\right) \quad (15)$$

This final form follows a von Mises distribution, reflecting the low-entropy steady state induced by synchronized interactions and reference phase attraction.

A.2 Local Score Matching

At any timestep t of the Markov chain, the marginal distribution $p(\theta_t)$ can be expressed as:

$$\begin{aligned} p(\theta_t) &= \int p(\theta_{t-1})p(\theta_t|\theta_{t-1}) d\theta_{t-1} \\ \nabla_{\theta_t} p(\theta_t) &= \int p(\theta_{t-1})\nabla_{\theta_t} p(\theta_t|\theta_{t-1}) d\theta_{t-1} \end{aligned} \quad (16)$$

For any score $\nabla_{\theta_t} \log p(\theta_t)$, we have $\nabla_{\theta_t} \log p(\theta_t) = \nabla_{\theta_t} p(\theta_t)/p(\theta_t)$. Then we can rewrite the above equation leveraging this identity:

$$\begin{aligned} p(\theta_t)\nabla_{\theta_t} \log p(\theta_t) &= \int p(\theta_{t-1})p(\theta_t|\theta_{t-1})\nabla_{\theta_t} \log p(\theta_t|\theta_{t-1}) d\theta_{t-1} \\ p(\theta_t)\nabla_{\theta_t} \log p(\theta_t) &= \int p(\theta_{t-1})p(\theta_t|\theta_{t-1})\nabla_{\theta_t} \log p(\theta_t|\theta_{t-1}) d\theta_{t-1} \\ \nabla_{\theta_t} \log p(\theta_t) &= \mathbb{E}_{\theta_{t-1} \sim p(\theta_{t-1}|\theta_t)} \left[\nabla_{\theta_t} \log p(\theta_t|\theta_{t-1}) \right] \end{aligned} \quad (17)$$

The above results indicate that the score can be effectively approximated by the local transition kernel.

A.3 Fourier Interpretation of Kuramoto Coupling

From a spectral standpoint, local phase coupling behaves like an angular low-pass filter. In the small-angle approximation ($\sin(\theta_j - \theta_i) \approx \theta_j - \theta_i$) and without a global reference phase, the forward SDE reduces to a stochastic heat equation (or a graph Laplacian form equivalently):

$$\frac{d\theta_t^i}{dt} = K(t)\nabla^2 \theta_t^i + \sqrt{2D_t}\xi^i \quad (18)$$

where ∇^2 denotes the Laplacian operator. Each Fourier component decays like $e^{-K(t)k^2t}$, where k denotes the spatial frequency. Modes with very high spatial frequency k (pixel-scale noise) are damped almost instantly, while moderate-frequency modes (coherent, edge-defining structures) decay much more slowly. Consequently, the Kuramoto drift effectively filters away pixel-scale noise while preserving the smooth, orientation-rich patterns that constitute edges.

A.4 Why Kuramoto Diffusion Falls Short on General Natural Images

Our Kuramoto coupling is foremost a synchronization mechanism: it excels at pulling similar phases together, thereby preserving edges and repetitive patterns. On orientation-rich data where the data are dominated by the **simple outlines and repeating ridge structures**, this yields sharp samples in fewer steps. By contrast, natural images (*e.g.*, CIFAR-10) demand modeling *e.g.*, **complex, global semantics (object shapes, color gradients, backgrounds), often characterized by higher-order and longer-range correlations**. In this setting, the local synchronization bias becomes less relevant and potentially even detrimental; an isotropic diffusion process with a global drift may be better suited to capture these large-scale, non-repetitive features. As a result, our method trades some global fidelity (reflected in higher CIFAR-10 FID) in exchange for strong local coherence.

B More Experimental Results

B.1 Implementation Details

Alg. 2 outlines the inference procedure for our model. We adopt the SDE formulation by default, but it can be optionally replaced with an ODE solver for improved efficiency by modifying Line 4 to: $\theta_{t-1} = \theta_t - f(\theta_t, t) + D_t \cdot s(\theta_t, t)$. We use AdamW optimizer with a learning rate of $1e-4$ and apply exponential moving average (EMA) updates with decay rate 0.995. A single NVIDIA A100 GPU is used for all the training and inference processes.

Network Architectures. We use a U-Net architecture following the design of [13, 60], equipped with three self-attention layers [63]. These are applied at spatial resolutions of 16, 8, and 4 for CIFAR-10

Algorithm 2 Inference algorithm for Kuramoto orientation diffusion models.

Require: Trained score network $s(\cdot, \cdot)$, noise variance schedule $\{2D_t\}_{t=1}^T$, forward Kuramoto drift $\mathbf{f}(\cdot, \cdot)$, noise $\epsilon \sim \mathcal{N}(0, I)$.

- 1: Sample initial phase: $\theta_T \sim p(\theta_T)$
 - 2: Initialize timestep counter $t = T$
 - 3: **while** $t > 0$ **do**
 - 4: Update: $\theta_{t-1} = \theta_t - \mathbf{f}(\theta_t, t) + 2D_t \cdot s(\theta_t, t) + \sqrt{2D_t}\epsilon$
 - 5: Wrap phase: $\theta_{t-1} = (\theta_{t-1} + \pi) \bmod (2\pi) - \pi$
 - 6: $t \leftarrow t - 1$
 - 7: **end while**
-

and Brodatz, and 24, 12, and 6 for SOCOFing. Timestep conditioning is implemented via sinusoidal positional embeddings. Each block uses group normalization [67] and GELU activations [20] throughout. The same architecture is shared across both our method and SGM for fair comparison.

Table 5: Linear schedules of noise variance and coupling strength under varying diffusion steps.

Steps	Global Coupling			Local Coupling		
	$2D$	K	K_{ref}	$2D$	K	K_{ref}
100	[1e-4,0.1]	[3e-5,0.03]	[4.5e-5,0.045]	[1e-4,0.1]	[5e-5,0.05]	[5e-5,0.05]
300	[1e-4,0.07]	[3e-5,0.02]	[4.5e-5,0.03]	[1e-4,0.07]	[5e-5,0.03]	[5e-5,0.03]
1000	[1e-4,0.015]	[3e-5,0.0045]	[4.5e-5,0.00675]	[1e-4,0.025]	[5e-5,0.01]	[5e-5,0.01]

Schedules of Noise and Coupling Strength. Table 5 summarizes the linear schedules for the noise variance $2D_t$, internal coupling strength K , and reference phase coupling K_{ref} . In the globally coupled variant, we maintain the relation $K_{\text{ref}} > 2D > K$ to ensure synchronization toward a reference phase while allowing sufficient noise injection. In the locally coupled variant, we increase the internal coupling strength K to compensate for its restricted spatial influence, which is limited to a 5×5 neighborhood around each pixel. We keep $K_{\text{ref}} = K$ in this case, since the reference coupling term inherently has a broader impact by acting across the entire image domain. For the SGM baseline, we adopt the VP-SDE formulation with linear variance schedules: [1e-4, 0.1] for 100 steps, [1e-4, 0.07] for 300 steps, and [1e-4, 0.02] for 1000 steps.

B.2 Ablation Study and Alternative Metrics

Table 6: FID Scores ($\downarrow$) on Brodatz texture dataset [1, 8].

Steps	100	300	1000
SGM [60]	38.33	22.40	20.37
Reference-only Process ($K(t) = 0$)	33.76	20.54	19.01
Kuramoto Orientation Diffusion (Globally Coupled)	20.26	18.51	15.42
Kuramoto Orientation Diffusion (Locally Coupled)	18.47	15.93	14.19

Ablation of the Kuramoto Coupling. To isolate the impact of Kuramoto coupling, we include a model variant by setting $K(t) = 0$ in Eq. (3), which keeps only the phase-reference drift in the SDE:

$$\frac{d\theta_t^i}{dt} = K_{\text{ref}}(t) \sin(\psi_{\text{ref}} - \theta_t^i) + \sqrt{2D_t}\xi^i \quad (19)$$

Table 6 presents the evaluation results on Brodatz textures. Removing the non-linear Kuramoto coupling raises 100-step FID by over 15 points, which confirms that non-isotropic phase synchronization is the key driver of both rapid convergence and the performance gain on orientation-rich data.

Alternative Metric. The FID metric is known for several limitations – its dependence on Inception embeddings, its Gaussian-moment matching assumption, and occasional misalignment with human judgment. As a more robust alternative, we adopt CLIP-MMD (CMMD) [29], which computes a nonparametric Maximum Mean Discrepancy in the pretrained CLIP feature space, which combines the benefits of distribution-free sample efficiency with the strong embedding power of CLIP.

Table 7: CMMD Scores ($\downarrow$) on Brodatz texture dataset [1, 8].

Steps	100	300	1000
SGM [60]	0.183	0.165	0.141
Kuramoto Orientation Diffusion (Locally Coupled)	0.072	0.045	0.030

Table 7 evaluates our local Kuramoto model versus SGM on the Brodatz textures dataset using the CMMD score. Our model achieves substantially lower CMMD at every step count. **Remarkably, the 100-step Kuramoto model outperforms the 1000-step SGM by a large margin.** These results mirror and amplify our FID improvements, demonstrating the superior sample fidelity even under a distribution-free, CLIP-based evaluation.

B.3 Earth and Climate Science Datasets on Spheres

To further evaluate our Kuramoto orientation diffusion model on naturally periodic data, we consider four real-world datasets of Earth and climate events: significant volcanic eruptions, earthquakes, floods, and wildfires [46, 45, 5, 18]. These datasets capture empirical spatial distributions of geophysical events on the surface of the Earth and are inherently defined over a 2D spherical domain. We map longitude to $[-\pi, \pi]$ and linearly scale latitude from $[-\frac{\pi}{2}, \frac{\pi}{2}]$ to the same interval to ensure compatibility with the periodic modeling framework. We compare our approach against a suite of Riemannian geometry-aware generative models, including Riemannian CNFs [40], Moser flows [53], CNF matching [4], and Riemannian score-based or flow-based generative models [26, 12, 10].

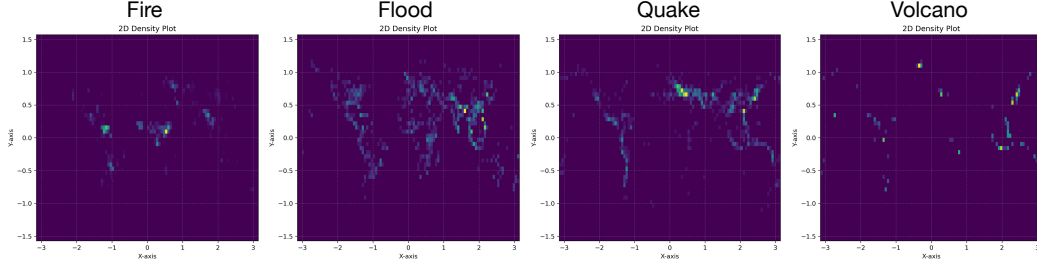


Figure 9: Learned density plot of our method on each Earth and climate science dataset. The X-axis denotes the longitude while the Y-axis represents the latitude.

Table 8: Test NLL on Earth and climate science datasets averaged across 5 runs.

Dataset	Volcano	Earthquake	Flood	Fire
Riemannian CNF [40]	-6.05 $\pm$ 0.61	0.14 $\pm$ 0.23	1.11 $\pm$ 0.19	-0.80 $\pm$ 0.54
Moser Flow [53]	-4.21 $\pm$ 0.17	-0.16 $\pm$ 0.06	0.57 $\pm$ 0.10	-1.28 $\pm$ 0.05
CNF Matching [4]	-2.38 $\pm$ 0.17	-0.38 $\pm$ 0.01	0.25 $\pm$ 0.02	-1.40 $\pm$ 0.02
Riemannian score-based [12]	-4.92 $\pm$ 0.25	-0.19 $\pm$ 0.07	0.48 $\pm$ 0.17	-1.33 $\pm$ 0.06
Riemannian diffusion model [26]	-6.61 $\pm$ 0.96	-0.40 $\pm$ 0.05	0.43 $\pm$ 0.07	-1.38 $\pm$ 0.05
Riemannian flow matching [10]	-7.93 $\pm$ 1.67	-0.28 $\pm$ 0.08	0.42 $\pm$ 0.05	-1.86 $\pm$ 0.11
Our Kuramoto orientation diffusion model	-5.18 $\pm$ 0.17	-0.18 $\pm$ 0.06	0.49 $\pm$ 0.18	-1.44 $\pm$ 0.05

Fig. 9 visualizes the learned densities produced by our method, capturing both highly concentrated regions (e.g., volcanic and fire clusters) and dispersed patterns. Table 8 presents the test negative log-likelihood (NLL) on each dataset. Our method achieves comparable performance against these baselines. We compute NLL using the change-of-variables formula under the time-reversed ODE solver: $\log p(\theta_0) = \log p(\theta_T) + \sum_{t=T}^1 \text{Tr}(\mathcal{J}_{b(\theta_t, t)})$ where $p(\theta_T)$ denotes the von Mises prior and $b(\theta_t, t)$ denotes the backward drift, i.e., $b(\theta_t, t) = -f(\theta_t, t) + D_t \cdot s(\theta_t, t)$. The Jacobian trace is estimated via Hutchinson’s stochastic estimator [27].

While our method achieves competitive NLL scores across datasets, we note that direct comparison across models can be influenced by the choice of different priors. In particular, von Mises distributions

with higher concentration can artificially boost log-likelihood scores, whereas broader priors can deflate them. Thus, while our results confirm the effectiveness of our model, qualitative evaluations remain critical for comprehensive assessments.

B.4 Navier-Stokes Fluid Velocity Field

To demonstrate the applicability to real angular data, we evaluate our Kuramoto diffusion model on 2D incompressible Navier–Stokes (NS) velocity fields from [6]. Each velocity frame (v_x, v_y) is converted to polar form:

$$r = \sqrt{v_x^2 + v_y^2}, \quad \theta = \arctan2(v_y, v_x) \quad (20)$$

where $r \geq 0$ is the amplitude and $\theta \in (-\pi, \pi]$ is the phase. Since r is positive, we work in log-magnitude $z = \log r$ and apply a VP SDE there in the log space, while phases evolve via the locally coupled Kuramoto SDE. The forward processes are defined as:

$$\begin{aligned} \text{Amplitude: } \frac{dz_t^i}{dt} &= -\frac{1}{2}\beta_t z_t^i + \boxed{\frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} K(t) \cos(\theta_t^j - \theta_t^i)} + \sqrt{\beta_t} \xi^i, \quad z_0 = \log r_0 \\ \text{Phase: } \frac{d\theta_t^i}{dt} &= \boxed{r_t^i} \left[\frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} K(t) \sin(\theta_t^j - \theta_t^i) + K_{\text{ref}}(t) \sin(\psi_{\text{ref}} - \theta_t^i) \right] + \sqrt{2D_t} \xi^i \end{aligned} \quad (21)$$

where the boxed terms couple the two channels: (i) local phase coherence accelerates amplitude growth (via $\cos$ in the z -SDE), and (ii) larger amplitude strengthens phase synchronization (amplitude factor r_t^i in the θ -SDE). To avoid over-coupling late in diffusion (when noise dominates), we enable coupling only in the early stage, *i.e.*, for $t < T/\alpha$ with a constant $\alpha > 1$; for $t \geq T/\alpha$ the two SDEs evolve independently under their native log-VP/Kuramoto dynamics.

In the reverse process we train two score models, one in log-magnitude space and one on phases, to solve the corresponding reverse SDEs and jointly synthesize (r, θ) .

Unconditional Generation. Following fluid-dynamics practice, we assess realism in the spectral domain at the final time step. From each 2D velocity field, we compute the radial energy spectrum $E(k)$ and then: (i) fit a line to $\log E(k)$ vs. $\log k$ over the mid-wavenumber band $k \in [0.1, 0.4] \cdot k_{max}$ (with $k_{max} = \min(H, W)/2$) and report the absolute slope difference, and (ii) compute the 1D Wasserstein distance between the *normalized* mean spectra of real and generated samples.

Table 9: Spectral evaluation of generated Navier-Stokes fluid velocity fields.

Metrics	Slope Difference ($\downarrow$)	Wasserstein Distance ($\downarrow$)
SGM (Amplitude-Phase Decomposition)	0.7435	0.0015
SGM (Cartesian Coordinates)	0.5590	0.0029
Ours (Naive Kuramoto Phase + Log-VP Amplitude)	0.6954	0.0011
Ours (Coupled Kuramoto Phase + Log-VP Amplitude)	0.3343	0.0005

Table 9 shows that the Coupled Kuramoto Phase + Log-VP Amplitude model achieves the best spectral realism among all methods, with the lowest slope error and Wasserstein distance. Relative to the SGM (Amplitude–Phase Decomposition) baseline, this corresponds to a 55% reduction in slope error and a 67% reduction in Wasserstein distance. It also improves over SGM (Cartesian Coordinates) by 40% in slope error and 83% in Wasserstein, and over the Naive Kuramoto + Log-VP variant by 52% (slope) and 55% (Wasserstein). These gains indicate that explicitly coupling phase synchronization with amplitude evolution yields more physically plausible velocity spectra than treating angle and magnitude independently or modeling them without coupling. Fig. 10 plots the average energy spectra with fitted lines, where the coupled model shows a visibly tighter fit in the mid-frequency band and a closer overall spectral shape.

Conditional Forecasting. Building on the unconditional results above, we next consider a setting that is arguably even more useful in PDE modeling: forecasting future states from history. Unlike unconditional generation which tests realism in distribution, forecasting probes whether the model has learned the dynamics well enough to extrapolate in time.

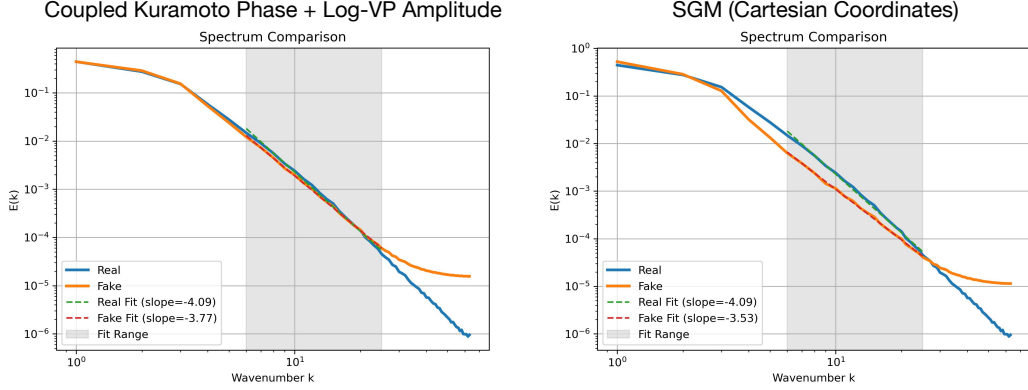


Figure 10: The average energy spectra with fitted lines.

We keep the coupled representations and train two conditioned score models s_c – one for phases θ and one for log-magnitudes z – together with their unconditioned counterparts s_u via condition dropout. At inference, each pair is fused with classifier-free guidance (CFG) [22]:

$$\begin{aligned} \text{Amplitude: } \tilde{s}_\omega(z) &= s_u(z) + \omega(s_c(z|c) - s_u(z)) \\ \text{Phase: } \tilde{s}_\omega(\theta) &= s_u(\theta) + \omega(s_c(\theta|c) - s_u(\theta)) \end{aligned} \quad (22)$$

where c denotes the two-step Navier–Stokes history, and ω is a guidance temperature that controls adherence to the condition (larger ω denotes stronger conditioning). We evaluate accuracy using Mean Squared Error (MSE) between predicted and ground-truth velocity fields.

Table 10: MSE of Navier-Stokes fluid velocity predictions.

Methods	MSE ($\downarrow$)
SGM (Cartesian Coordinates)	0.0260
Ours (Coupled Kuramoto Phase + Log-VP Amplitude)	0.0188

As shown in Table 10, the Coupled Kuramoto Phase + Log-VP Amplitude model reduces MSE from 0.0260 to 0.0188 compared with the SGM (Cartesian Coordinates) baseline, indicating more accurate conditional forecasts. Fig. 11 illustrates two examples; predictions closely match the ground truth in both horizontal and vertical velocity components.

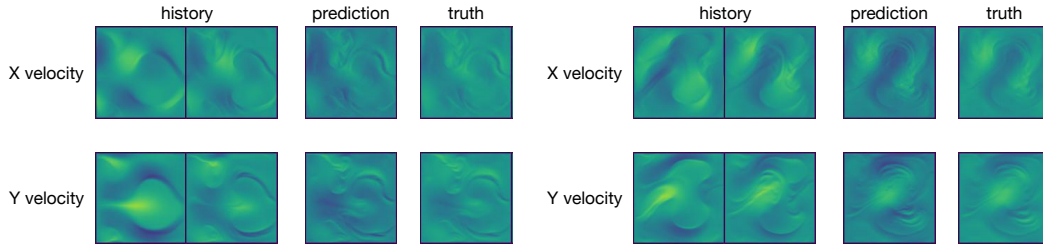


Figure 11: Conditional one-step forecasts of our method on Navier–Stokes velocity: history (left), predictions (middle), and ground truth (right).

B.5 More Examples of Generative Samples

Figs. 12, 13, 14, and 15 present additional randomly generated samples from our 1000-step locally coupled Kuramoto orientation diffusion model on the SOCOFing fingerprint, Brodatz texture, ground terrain, and CIFAR10 datasets, respectively.

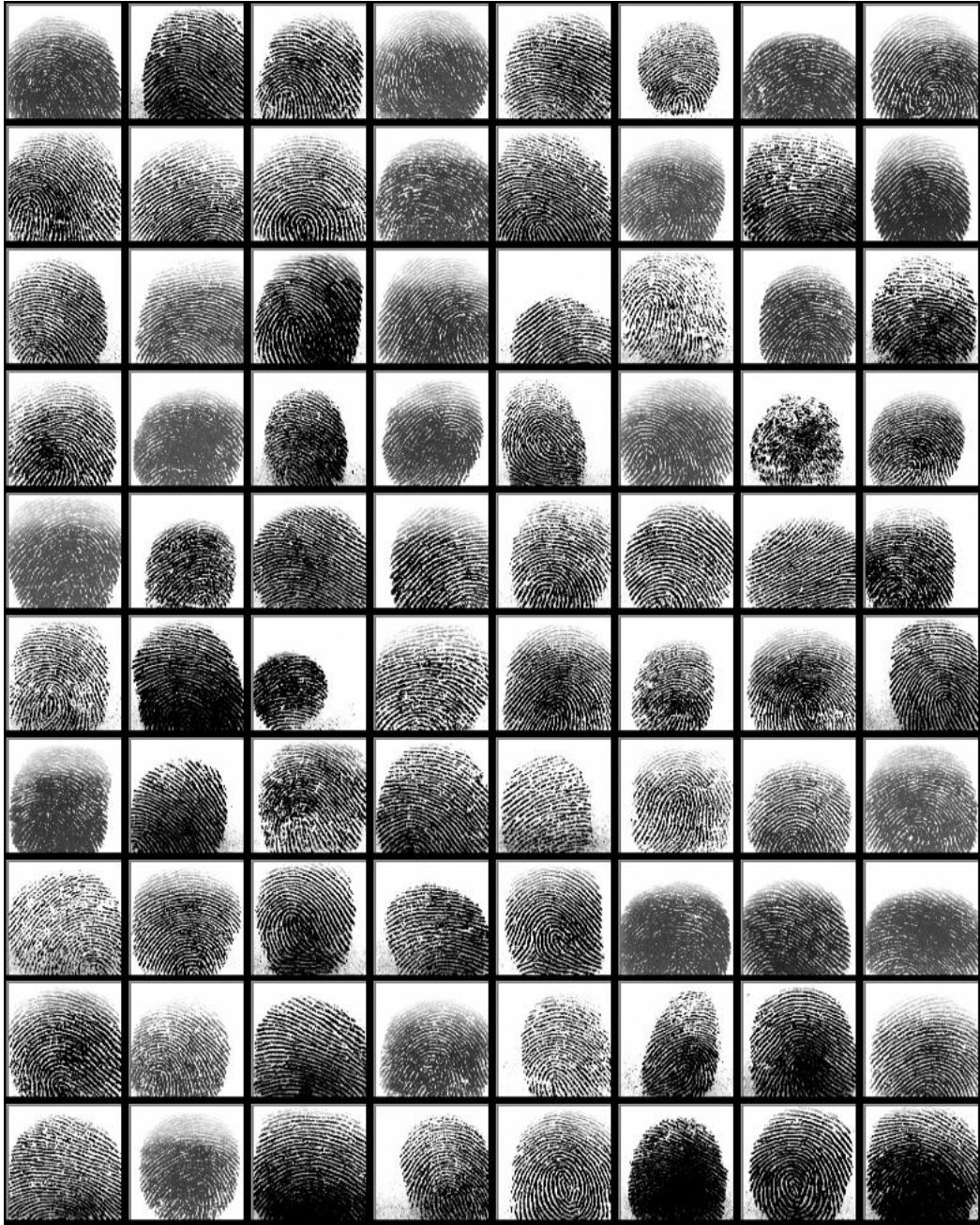


Figure 12: Randomly generated samples on SOCOFing fingerprint dataset.

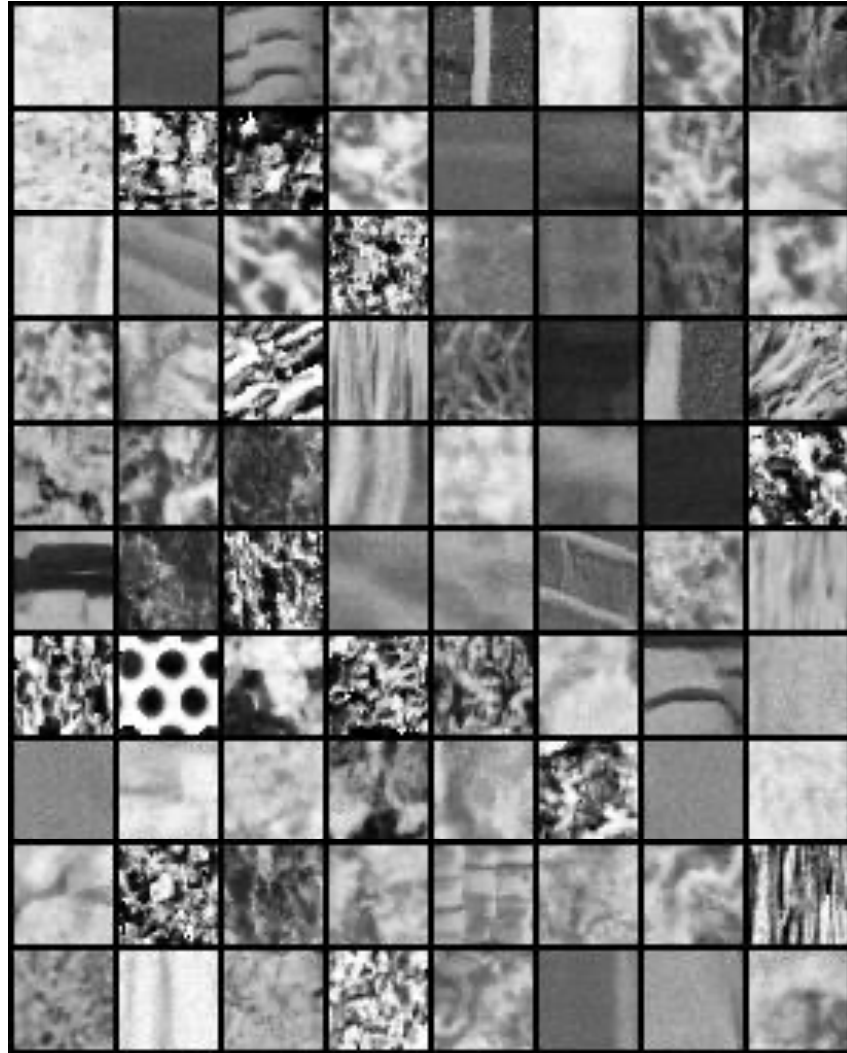


Figure 13: Randomly generated samples on Brodatz textures dataset.

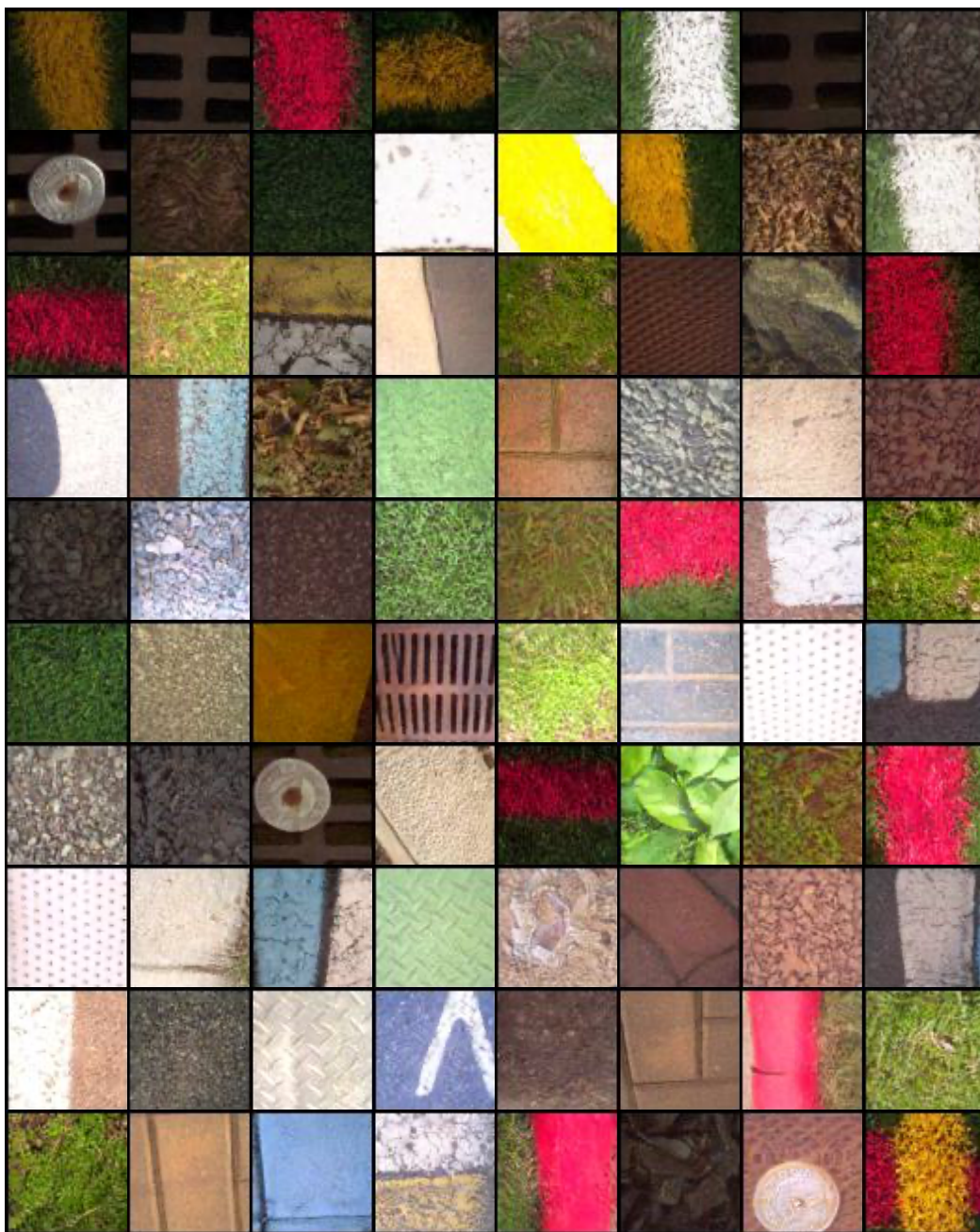


Figure 14: Randomly generated samples on the ground terrain dataset.



Figure 15: Randomly generated samples on CIFAR10 dataset.