GENERATIVE FRACTIONAL INTERPOLANTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Mandelbrot–Van Ness fractional Brownian motion (Type I fBM) with Hurst index $H \in (0, 1)$ interpolates between Brownian motion, which drives continuous-time diffusion models, and random straight-line paths used in flow matching as $H \to 1$. The Hurst index controls the correlation of increments and thereby the temporal memory of the noise process. Away from $H = 1/2$, this memory makes fBM non-Markovian, motivating Markovian approximations for tractable computation. However, the Markovian approximation of fBM (MA-fBM) used in prior work exhibits increasing error as $H \to 1$ and fails to preserve the random straight-line limit. We therefore improve MA-fBM by augmenting the approximation process with the terminal value of fBM and derive the optimal approximation coefficients in closed form. The resulting ε-MA-fBM preserves this limiting behavior while reducing the approximation error. Building on this approximation, we introduce *Generative Fractional Interpolants (GFI)*, a generative model interpolating sub-diffusive dynamics ($H < 1/2$), standard diffusion ($H = 1/2$), superdiffusive dynamics ($H > 1/2$), and random straight line paths ($H \to 1$) used in flow matching. To train our generative model, we derive a KL-divergence of the reverse process and its parameterization and train over the full range of Hurst indicies, allowing the best operating regime to be selected at sampling time. Notably, unlike flow matching, GFI can achieve straight generative trajectories from noise to data, at the cost of trajectory crossings. Empirically, this design choice yields in our experiments lower FD-DINOv2 scores than diffusion models and flow matching on CIFAR-10, the imbalanced CIFAR-10-LT dataset, and CelebA-HQ-256 in latent space.

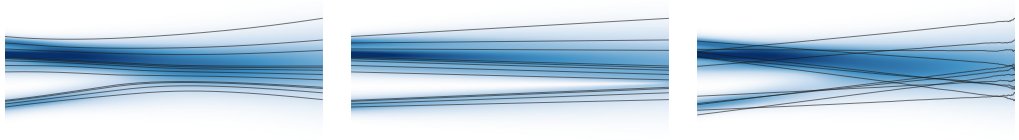
1 INTRODUCTION

Score-based diffusion models (Song et al., 2021) and flow matching (Lipman et al., 2023a) learn continuous-time generative processes that transport noise to data, with sampling performed by solving a stochastic differential equation (SDE) or an ordinary differential equation (ODE). When sampling via the ODE, the resulting trajectories are curved, even in flow matching which is trained by regression against random straight-line interpolations between data and noise (Gao et al., 2025). Recent efforts have been taken to straighten the trajectories of generative models via a mixture of ODE and SDE trajectories (Karras et al., 2022), optimal transport minibatching (Tong et al., 2024) or iterative refinement via rectified flow (Liu et al., 2023b; 2024).

On the other hand, the choice of driving noise in continuous-time diffusion models has been generalized beyond Brownian motion to heavy-tailed Lévy processes and Markov approximations of non-Markovian Type II fractional Brownian motion (fBM), with reported improvements in sample quality, the modeling of imbalanced data, and sampling efficiency (Yoon et al., 2023; Nobis et al., 2024; Paquet et al., 2024).

We show that these two research directions can be naturally combined by introducing *Generative Fractional Interpolants (GFI)*, a generative model driven by Mandelbrot–Van Ness fractional Brownian motion (Type I fBM) (Mandelbrot & Van Ness, 1968), which interpolates via its Hurst index $H \in (0, 1)$ between Brownian motion and the random straight-line paths used in flow matching. Notably, unlike in flow matching, straight generative trajectories from noise to data can be achieved, at the cost of allowing trajectory crossings in contrast to optimal transport visualized in Fig. 1.Our contributions are:

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



(a) Flow Matching (b) Optimal Transport (c) GFI (ours)

Figure 1: Comparison of closed-form transport trajectories from a standard normal distribution on the RHS to a mixture of Gaussians on the LHS. (a) Flow matching with linear interpolation between data and noise yields curved, non-crossing trajectories. (b) Optimal transport yields straight, non-crossing trajectories. (c) GFI with $H = 0.999$ yields straighter trajectories than flow matching, while trajectories do cross. See Figs. 8 and 9 for the full range of trajectories attainable by GFI and selectable at sampling time.

- We show that a generative model driven by Type I fractional Brownian motion with Hurst index $H \in (0, 1)$ interpolates between a continuous time score-based diffusion model ($H = 0.5$) and a flow model ($H \to 1$).
- We improve the Markovian approximation of Type I fractional Brownian motion to preserve this interpolating behaviour.
- We design a generative model driven by type I fBM with a continuum of memory levels in the driving noise, jointly available during training and yielding a free modeling choice at sampling time.

2 Related work

Noise processes of continuous-time diffusion models. Score-based diffusion models (SBDM), introduced in the seminal work of Song et al. (2021), formulate generative diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020) in continuous time by connecting score matching (Hyvärinen, 2005; Song & Ermon, 2019) with the time-reversal theory of stochastic differential equations (SDEs) (Stratonovich, 1960; Anderson, 1982; Föllmer, 1986). Instead of injecting noise at a finite number of fixed noise levels via a Markov chain, the continuous-time formulation provides a continuum of noise levels, with noise continuously injected through the driving Brownian motion, a centered Gaussian Markov process with independent and stationary increments. Recent work has begun to relax this choice of driving noise by considering heavy-tailed processes (Yoon et al., 2023; Paquet et al., 2024; Shariatian et al., 2025) and Markov approximations of non-Markovian Type II fractional Brownian motion (fBM) (Nobis et al., 2024; Liang et al., 2025; Nobis et al., 2025). These alternatives have been reported to improve sample quality (Yoon et al., 2023; Paquet et al., 2024; Nobis et al., 2024), better capture imbalanced data (Yoon et al., 2023; Shariatian et al., 2025), and reduce the number of sampling steps or model evaluations required during generation (Yoon et al., 2023; Shariatian et al., 2025; Nobis et al., 2024).

Diffusion and flow models. While continuous-time diffusion models parameterize the score function associated with a solution to an SDE, flow matching (FM) (Lipman et al., 2023b; Liu et al., 2023b), parameterizes a velocity field, approximating a flow from noise to data that solves an ordinary differential equation (ODE). From the perspective of Gaussian probability paths from data to noise, diffusion and flow models provide two different realizations of the same marginal evolution (Karras et al., 2022; Lipman et al., 2023a). A further unification is provided by stochastic interpolants (SI) (Albergo & Vanden-Eijnden, 2022), which show that the same interpolating density satisfies a transport equation and a family of Fokker–Planck equations, yielding deterministic ODE and stochastic SDE generative dynamics with identical time-dependent marginals. This connection is further illustrated in Gao et al. (2025). Efficient few step sampling methods for diffusion and flow models mostly rely on sampling from ODE trajectories (Karras et al., 2022; Lipman et al., 2023b). As shown by Karras et al. (2022), the quality and speed of generation process is influenced by the curvature of the integrated paths with straight paths lowering the discretization error and allowing for bigger step sizes. Flow Matching (Lipman et al. (2023b)) makes straight paths an explicit design goal by regressing against straight-line interpolation between data and noise. Since the pairing of noise and data are drawn independently during training, the linear interpolations used for training cross and the learned marginal sampling paths remain curved (Lee et al., 2023). Rectified Flow (Liu et al., 2023b; 2024)) and Optimal Transport Minibatching (Tong et al., 2024) both aim to straighten

the paths during sampling. While the former adds a post-training step by retraining on pairs of noise and sampled data, the latter computes the OT coupling for each minibatch during model training. Alternative lines of research add curvature penalties to the training objective (Lee et al., 2023) or distill trained models into one-step samplers that directly map noise to data samples (Song et al., 2023; Salimans & Ho, 2022).

3 PRELIMINARIES

We begin with a review of the concepts underlying our generative model: continuous-time diffusion models driven by Brownian motion, flow matching based on straight-line paths, and fractional Brownian motion, which interpolates between Brownian motion and random straight-line paths.

Score-based diffusion models. Continuous-time score-based diffusion models describe the gradual transformation of a data distribution p_0 on $\mathbb{R}^d$ into a tractable noise distribution p_1 through a stochastic process $X = (X(t))_{t\in[0,1]}$, whose marginal density at time t is denoted by p_t, satisfying the stochastic differential equation (SDE) (Song et al., 2021)

$$\mathrm{d}X(t) = f(X(t), t)\,\mathrm{d}t + g(t)\,\mathrm{d}B(t), \quad X(0) \sim p_0, \tag{1}$$

where the drift $f : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ and the diffusion coefficient $g : [0,1] \to \mathbb{R}$ are chosen such that the terminal law p_1 of X_1 is approximately Gaussian. The driving noise process B is a d-dimensional Brownian motion, i.e. a centered Gaussian process with independent, stationary increments, uniquely characterized in law by its covariance function $Cov(B(t), B(s)) = \min\{s,t\}I_d$. The generative process reverses this transformation, traversing from noise to data. Whenever $X = (X(t))_{t\in[0,1]}$ is a stochastic process, we denote its time reversal by $\overline{X}(t) = X(1-t)$. Similarly, we write $\overline{g}(t) = g(1-t)$, $\overline{f}(x,t) = f(x,1-t)$, and $\overline{p}_t = p_{1-t}$. For any $\eta \in C([0,1],[0,\infty))$, the family of reverse-time dynamics (Song et al., 2021; Karras et al., 2022; Gao et al., 2025)

$$\mathrm{d}\overline{X}(t) = \big[-\overline{f}(\overline{X}(t),t) + \frac{1+\eta(t)^2}{2}\overline{g}^2(t)\,\nabla_x \log \overline{p}_t(\overline{X}(t))\big]\,\mathrm{d}t + \eta(t)\overline{g}(t)\,\mathrm{d}\overline{B}_t, \quad \overline{X}(0) \sim p_1 \tag{2}$$

has marginals $\overline{p}_t = p_{1-t}$, where $\overline{B}$ is a d-dimensional Brownian motion. The choice $\eta \equiv 1$ recovers the time reversal of the forward SDE in Eq. (1) (Stratonovich, 1960; Anderson, 1982; Föllmer, 1986), whereas $\eta \equiv 0$ yields the probability flow ODE (PF ODE) (Song et al., 2021), whose trajectories are deterministic conditional on their initialization. When p_0 is unknown, so is the score function $\nabla_x \log p_t$, which is approximated by a time-dependent score model s_θ trained via score matching (Hyvärinen, 2005).

Flow matching. Flow models (Lipman et al., 2023a) instead parameterize the velocity field of deterministic dynamics

$$\mathrm{d}X(t) = u_t(X(t))\,\mathrm{d}t, \qquad X(0) \sim p_0, \tag{3}$$

where $u : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ is a time-dependent vector field whose flow ϕ_t, defined by $\partial_t \phi_t(x) = u_t(\phi_t(x))$ and $\phi_0 = \mathrm{id}$, transports the data distribution to the noise distribution, $X(1) = \phi_1(X(0)) \sim p_1$. Since the marginal vector field u_t is generally intractable, flow matching learns it by regression against tractable probability paths conditioned on independently sampled data and noise endpoints. A common choice for these conditional probability paths is the straight-line interpolation $X^\star(t) = (1-t)X(0) + tX(1)$, which solves the random ODE from noise to data

$$\mathrm{d}X^\star(t) = \frac{X^\star(t) - X(0)}{t}\,\mathrm{d}t, \quad (X^\star(0), X(1)) \sim p_0 \otimes \mathcal{N}(0, \mathrm{I}). \tag{4}$$

For every fixed pair (x_0, x_1), the solution of Eq. (4) is the straight-line displacement interpolation between its endpoints, also referred to as a conditional optimal transport (OT) path. The vector field in Eq. (4) depends on its initial endpoint $X(0)$ and therefore cannot be evaluated from any state $X^\star(t), t \in (0,1]$ alone. Flow matching avoids this dependence by approximating the marginal vector field $v_t(x) = \mathbb{E}\,[X(1) - X(0) | X^\star(t) = x]$ by a time-dependent neural network s_θ, trained by regressing against the tractable conditional velocity (Lipman et al. (2023a)). The resulting marginal ODE

$$\mathrm{d}\widehat{X}(t) = v_t(\widehat{X}(t))\,\mathrm{d}t, \qquad \widehat{X}(0) \sim p_0, \tag{5}$$

shares the same one-time marginals as $X^\star$, while generally inducing a different distribution over trajectories. Data generation is performed by solving this ODE backward in time, starting from

$\widehat{X}(1) \sim p_1$ and replacing v with the learned model s_θ. Thus, although flow matching is trained using conditional straight-line paths, it generally neither learns the optimal transport coupling between p_0 and p_1 nor produces straight-line trajectories under the learned marginal vector field visible in Fig. 1a and Fig. 1b.

Fractional Brownian motion. Fractional Brownian motion (fBM) generalizes the driving process of continuous time diffusion models, the Brownian motion by relaxing the independence of increments: increments become temporally correlated, the dynamics non-Markovian, and the sample paths smoother or rougher than those of Brownian motion, as depicted in Fig. 3. Two variants appear in the literature, that differ by the stationarity of their increments and by their limiting behavior as $H \to 1$, where Type I fBM converges to a random straight-line.

Definition 1 (Type I Fractional Brownian motion) *Denote with Γ the Gamma function. Mandelbrot–Van Ness fractional Brownian motion (Type I fBM) (Mandelbrot & Van Ness, 1968) with Hurst index $H \in (0,1)$ is the centered Gaussian process $B_H^{(I)} = (B_H^{(I)}(t))_{t\in[0,1]}$ uniquely characterized in law by its covariances*

$$\mathrm{Cov}\big(B_H^{(I)}(t), B_H^{(I)}(s)\big) = \tfrac{1}{2}\big(t^{2H} + s^{2H} - |t-s|^{2H}\big), \quad t,s \in [0,T]. \tag{6}$$

For $H = \frac{1}{2}$, the covariance reduces to $\min\{t,s\}$, so that Brownian motion is recovered. For $H \neq \frac{1}{2}$, fBM is non-Markovian and fails to be a semimartingale (Biagini et al., 2008), with positively correlated increments for $H > \frac{1}{2}$ (superdiffusion) and negatively correlated increments for $H < \frac{1}{2}$ (subdiffusion). Type I fBM possesses stationary increments, whereas the increments of Type II fBM as defined in Dfn. 4 are non-stationary. Importantly, Type I fBM converges weakly to a random straight-line path as $H \to 1$, thereby interpolating between Brownian motion and random straight-line paths via the Hurst index.

Proposition 1 (Smooth endpoint collapse of Type I fBM) *Assume that the standard normal random variable $\varepsilon \sim \mathcal{N}(0,1)$ satisfies the common-terminal coupling assumption (A1). As $H \uparrow 1$,*

$$B^H \Longrightarrow (t\varepsilon)_{t\in[0,1]} \qquad \text{in } C([0,1]). \tag{7}$$

We provide the proof in Sec. C.1, where we also show convergence of Type I fBM to random straight-line paths in probability and in L^p. In contrast, we show in Prop. 10 that Type II fBM converges to a stochastic process that is almost surely *not* a straight line, i.e., that retains stochasticity along its paths. As a consequence of Prop. 1 and Prop. 10, we can construct a generative model driven by Type I fBM that interpolates between a diffusion model and a flow model, as made precise in the next section, whereas Type II fBM does not exhibit this interpolating behavior. To improve readability, throughout the remainder of this manuscript, we use fBM to refer to Type I fBM and explicitly specify Type II fBM only where needed.

4 Interpolating between diffusion and flow models via memory

Although the time reversal of an SDE driven by fBM does not admit the structure required to train a score-based diffusion model (Darses & Saussereau, 2007), a forward process driven by fBM is well defined for all $H \in (0,1)$ (Biagini et al., 2008). We define the forward noising process of our generative model as a family of continuous-time processes over the memory continuum $H \in (0,1)$. The mathematical details of this section and proofs are provided in Sec. C.

Family of forward processes with memory continuum. Fix any Hurst index $H \in (0,1)$, a time-dependent drift coefficient $\mu \in L^1([0,1])$ and a diffusion function $g : [0,1]\times[0,1] \to \mathbb{R}$ such that $t \mapsto g(t,H)\exp(-\int_0^t \mu(s)\,\mathrm{d}s)$ is continuous and of bounded variation. For an arbitrary Borel data distribution p_0, we define the forward process X_H of *generative fractional interpolants (GFI)* as a solution to the SDE

$$\mathrm{d}X_H(t) = \mu(t)X_H(t)\,\mathrm{d}t + g(t,H)\,\mathrm{d}B_H(t), \quad X_H(0) = X_0 \sim p_0, \tag{8}$$

where the data generating random variable X_0 is assumed to be independent of the driving fBM B_H. For $H = 1/2$, the forward process of GFI recovers the noising process of SBDM and we show in Sec. C.5, as $H \to 1$, weak convergence in path space to the solution of an ODE.

Proposition 2 (Path space convergence towards flow) *Assume that the standard normal random variable $\varepsilon \sim \mathcal{N}(0,1)$ satisfies the common terminal coupling assumption (A1) as well as independence of ε and $X_0 \sim p_0$. Define*

$$\Phi(t) := \exp\left(\int_0^t \mu(s)\right), \qquad q(t) := \frac{g(t)}{\Phi(t)} \qquad \textit{and} \qquad X^\star(t) := \Phi(t)\left(X_0 + \varepsilon \int_0^t q(s)ds\right). \tag{9}$$

Then as $H \to 1$

$$X_H \Longrightarrow X^\star \qquad \textit{in } C([0,T]). \tag{10}$$

We define in the following two variants of GFIs forward noising process.

Fractional generalized Ornstein-Uhlenbeck forward dynamics. For GFI paired with the variance preserving (VP) dynamics (GFI_{VP}) from (Song et al., 2021) we have $\mu(t) = -\frac{1}{2}\beta(t)$, $g(t) = \sqrt{\beta(t)}$ with $\beta(t) = \beta_{min} + t(\beta_{max} - \beta_{min})$, $\beta_{min} = 0.1$ and $\beta_{max} = 20.0$ such that by Prop. 2 and the derivation in Sec. C.7, the corresponding limiting random path as $H \to 1$ is given by

$$X(t)^\star = \Phi(t)\left(X_0 + \varepsilon \int_0^t \Phi(s)^{-1}\sqrt{\beta(s)}\, ds\right), \quad \varepsilon \sim \mathcal{N}(0,1). \tag{11}$$

Note that GFI_{VP} is no longer variance preserving for $H \neq 1/2$, since the relation between drift and diffusion coefficients to preserve the variance is specific to the Brownian covariance structure.

Flow matching inherited forward dynamics. We now construct a forward process that converges for $H \to 1$ to the linear interpolation path between data and noise used in flow matching. Fix $\delta > 0$ and define the drift coefficient $\mu : [0,1] \to \mathbb{R}$ with $t \mapsto -(1+\delta-t)^{-1}$. For the diffusion function choose

$$q_H(t) := c_H t^{1-H}(1-t)^{-H-1}, \qquad g(t,H) := (1-t)q_H(t) = c_H t^{1-H}(1-t)^{-H}, \tag{12}$$

where c_H is given in Eq. (55). The so defined forward process converges in path space to the straight-line interpolation paths between data and noise used in Flow matching and we denote this schedule by GFI_{FM}.

Proposition 3 (Convergence to straight-line interpolation paths) *For the constructed forward process X_H following the dynamics*

$$\mathrm{d}X_H(t) = \mu(t)X_H(t)\,\mathrm{d}t + g(t,H)\,\mathrm{d}B_H(t), \qquad X(0) \sim p_0 \tag{13}$$

and

$$X^\star(t) = (1-t)X_0 + t\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1) \tag{14}$$

we find

$$\big((X_H(t))_{t\in[0,1]}, B_1^H\big) \Longrightarrow \big((X^\star)_{t\in[0,1]}, \varepsilon\big) \qquad \textit{in } C([0,1]) \times \mathbb{R}. \tag{15}$$

Unlike Liu et al. (2023a), who obtain $X^\star$ in the limit of a vanishing diffusion coefficient of a Brownian-motion-driven process, we obtain it in the limit $H \to 1$ of an fBM-driven proces. Consequently, GFI with the GFI_{FM} dynamics defined above interpolates between the paths of an SBDM and the straight-line interpolation paths used to train flow matching models.

Even for $H \in (1/2, 1)$, the time reversal of GFI's forward process does not admit the structure required to train a score-based diffusion model (Darses & Saussereau, 2007). We therefore follow generative fractional diffusion models (GFDM) (Nobis et al., 2024) and approximate the fractional driver by a Markov process. However, the Markovian approximation of fBM (MA-fBM) used in GFDM and defined in Dfn. 6 is not well suited to Type I fBM, particularly in the regime $H \to 1$: as shown in Sec. 4, the relative approximation error increases as H approaches 1. This is also reflected in the sample paths in Sec. 4, where the MA-fBM paths are no longer aligned with those of fBM. Before defining our approximate GFI model, we therefore introduce an improved approximation that preserves the limiting behavior of fBM as $H \to 1$, by adding a sample the terminal value $B_H(1)$ of fBM to the linear combination of OU processes.

Definition 2 (ε-MA-fBM) *Define for $\gamma > 0$ the family of Ornstein–Uhlenbeck (OU) processes Y_γ solving $\mathrm{d}Y_\gamma(t) = -\gamma Y_\gamma(t)\,\mathrm{d}t + \mathrm{d}B(t)$ with starting value $Y_\gamma(0) = \int_{-\infty}^0 e^{\gamma s}\,\mathrm{d}B(s)$. Set $c(H) := \sqrt{1/H + 1/(1-H)}$ and define for $\tilde{K} \in \mathbb{N}$ the grid $0 < \gamma_1 < \cdots < \gamma_{\tilde{K}}$ via*

$$\gamma_{min} = \exp\left(-\frac{\pi\sqrt{K}}{(1-H)C(H)}\right), \qquad \gamma_{max} = \exp\left(-\frac{\pi\sqrt{K}}{(1-H)C(H)}\right), \qquad \gamma_k = \gamma_{min}^{\,1-\frac{(k-1)}{K-1}}\,\gamma_{max}^{\,\frac{(k-1)}{K-1}}. \tag{16}$$

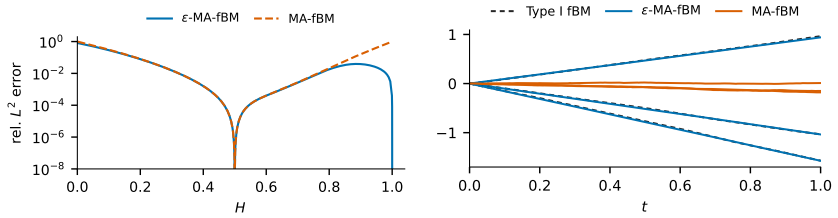


Figure 2: Comparison of MA-fBM and ε-MA-fBM (**ours**) as approximations of Type I fBM. (LHS) For MA-fBM, the relative approximation error deteriorates as $H \to 1$ (orange), whereas for ε-MA-fBM it approaches zero (blue). (RHS) Sample paths of ε-MA-fBM (blue) at $H = 0.999$ are visually close to fBM (black), while the paths of MA-fBM (orange) are not aligned with fBM.

We now set for $K \in \mathbb{N}_{\geq 2}$ approximating processes

$$Y(t) := (Y_1(t), Y_2(t), Y_3(t), \dots, Y_K(t)) := (B(t), tB_H(1), Y_{\gamma_1}(t), \dots Y_{\gamma_{\tilde{K}}}(t)) \tag{17}$$

and define for approximation coefficients $\omega_1, \dots, \omega_K \in \mathbb{R}$ the ε-Markov-approximate fBM (ε-MA-fBM) by the linear combination

$$\hat{B}_H(t) = \sum_{k=1}^{K} \omega_k \left(Y_k(t) - Y_k(0)\right). \tag{18}$$

Following the approach of Daems et al. (2024), we derive the L^2-optimal approximation coefficients in closed form. The derivation is more involved due to the terminal value of fBM and the resulting correlation structure.

Proposition 4 (Optimal approximation coefficients of ε-MA-fBM) *The approximation coefficients that minimize the L^2 error*

$$\mathcal{E}(\boldsymbol{\omega}) := \int_0^1 \mathbb{E}\left[\left(B_H(t) - \hat{B}_H^{\varepsilon}(t)\right)^2\right] \mathrm{d}t \tag{19}$$

are given by the closed-form expression $\boldsymbol{A}^{\varepsilon}\boldsymbol{\omega} = \boldsymbol{b}^{\varepsilon}$, with $\boldsymbol{A}^{\varepsilon}$ and $\boldsymbol{b}^{\varepsilon}$ defined in Eq. (85).

Our choice of $\gamma_1, \dots, \gamma_K$ is motivated by Bayer & Breneis (2023) and McLean (2018). Up to this choice, we recover MA-fBM proposed in Daems et al. (2024) with $\tilde{K} = K - 2$ approximating processes by setting $(\omega_1, \omega_2) = (0, 0)$. In Sec. 4, we observe that the relative approximation error of MA-fBM deteriorates as $H \to 1$, whereas that of ε-MA-fBM approaches zero. Correspondingly, Sec. 4 shows that for $H = 0.999$, the fBM path is well approximated by ε-MA-fBM, while the MA-fBM and fBM paths are no longer aligned. Sampling paths are depicted in Fig. 3.

5 Generative modeling with ε-MA-fBM

In this section, we define an generative model driven by ε-MA-fBM that approximates the GFI dynamics of Sec. 4, approximately inheriting the interpolating behaviour between a diffusion model and a flow model. Let p_0 denote a data distribution on $\mathbb{R}^d$ from which we aim to sample. We define the approximate forward process of GFI as a solution to

$$\mathrm{d}X(t) = \mu(t)X(t)\,\mathrm{d}t + g(t, H)\,\mathrm{d}\hat{B}_H(t), \quad X(0) \sim p_0, \quad t \in [0, T] \tag{20}$$

with drift scaling function $\mu : [0, T] \to \mathbb{R}$ and diffusion function $g : [0, 1] \times (0, 1] \to \mathbb{R}$. The forward process subsumes noising of different generative models in the sense that we see from the evolution of the optimal approximation coefficients in Fig. 4 for $K = 2$

$$dX_t \approx \begin{cases} \mu(t)X_t\,\mathrm{d}t + \omega_1 g(t, H)\,\mathrm{d}B(t), & H = 0.5 \quad \textit{(Diffusion)} \\ \mu(t)X_t\,\mathrm{d}t + \omega_2 \varepsilon g(t, H)\,\mathrm{d}t, & H = 1 \quad \textit{(Flow)} \\ \mu(t)X_t\,\mathrm{d}t + \omega_1 g(t, H)\,\mathrm{d}B(t) + \omega_2 \varepsilon g(t, H)\,\mathrm{d}t, & H \in (0.5, 1) \quad \textit{(Interpolation)} \end{cases} \tag{21}$$

where $H \in (0.5, 1)$ interpolates between the noising process of SBDM and a flow. For $K > 2$, $Y(0) = 0$ and $(\omega_0, \omega_1) = (0, 0)$ we recover GFDM driven by MA-fBM. For all $K > 0$ the forward

process X solving Eq. (20) approximately interpolates between the four regimes: a sub-diffusion ($H < 1/2$), a regular diffusion $H = 1/2$, a super-diffusion $H > 1/2$ and a flow as $H \to 1$. To reverse this forward process, we need to jointly reverse the augmented forward process $Z = (X, Y)$. Towards that goal we first rewrite the augmented forward process as a linear SDE with detailed derivations in Sec. G.

Proposition 5 *The augmented forward process $Z = (X, Y)$ follows the dynamics*

$$\mathrm{d}Z(t) = F(t, H)Z(t)\,\mathrm{d}t + G(t, H)\,\mathrm{d}B_t, \quad Z(0) = (X(0), Y(0)) \tag{22}$$

with $F : [0, 1] \times (0, 1) \to \mathbb{R}^{d(K+1),d(K+1)}$ and $G : [0, 1] \times (0, 1) \to \mathbb{R}^{d(K+1),d}$ given in **??**.

Since Z satisfies a linear SDE, its conditional law given $X(0) = x_0$ is Gaussian. In particular, $Z(t) \mid X(0) = x_0 \sim \mathcal{N}(\mu(t, x_0), \Sigma(t))$. We approximate the covariance $t \mapsto \Sigma(t)$ by numerically solving the ODE given in Eq. (162), which enables simulation-free training as detailed in Sec. E. Due to the Markovian approximation of fBM, the augmented forward process Z is Markovian and the reverse-time dynamics of Eq. (22) are given by (Stratonovich, 1960; Anderson, 1982; Föllmer, 1986)

$$\mathrm{d}\overline{Z}(t) = \left[-\overline{F}(t, H)\overline{Z}(t) + \overline{G}(t, H)\overline{G}(t, H)^T \nabla_z \log \overline{p}_t^a(\overline{Z}(t))\right] \mathrm{d}t + \overline{G}(t, H)\,\mathrm{d}\overline{B}_t, \tag{23}$$

where $\overline{p}_t^a$ denotes the density of the marginal law of the augmented reverse process $\overline{Z}$, defined by $\overline{Z}(t) = Z(1 - t)$. To train GFI over the full range of Hurst indices $H \in (0, 1)$ we derive a KL-divergence of the reverse process and its parameterization. The proof is provided in Sec. E.

Proposition 6 *Denote by $\mathbb{P}^\star$ the path measure associated to the reverse process $\overline{Z}$. For a time-dependent neural network $s_\theta : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ define the augmented score-model $S_\theta : \mathbb{R}^{d(K+3)} \times [0, 1] \to \mathbb{R}^{d(K+3)}$ via*

$$(x, y, t) = (z, t) \mapsto A(t)s_\theta(u, t) + (0, \nabla_y \log q_t(y))^\mathsf{T}, \tag{24}$$

where q_t denotes the known density of Y and $A(t) \in \mathbb{R}^{d(K+3),d(K+1)}$ is given in closed-form (Prop. 14). Replacing the score-function in Eq. (23) with S_θ yields the parameterized reverse process with associated path measure $\mathbb{P}^\theta$ and we find

$$D_{KL}\left(\mathbb{P}_H^\star | \mathbb{P}^\theta\right) \leq \mathcal{L}_{prior}(H, X_0) + \frac{1}{2}\int_0^T \lambda^2(t) \underbrace{\mathbb{E}_{X_0|X_t,Y_t}\left[\|\nabla_x \log p_{0t}(X_t|Y_t, X_0) - s_\theta(U_t, t)\|^2\right]}_{:=\mathcal{L}_t(\theta)} \mathrm{d}t \tag{25}$$

with weighting function $\lambda : [0, T] \to \mathbb{R}$ satisfying $G(t, H)^\mathsf{T} A(t) = \lambda(t) I_d$, neural-network input $U_t = X_t - \sum_{k=0}^{K+3} [A(t)]_{1,k+1} Y_k$ and $\mathcal{L}_{prior}(H, x_0) = \frac{\exp(\int_0^1 \mu(s)ds) x_0^\mathsf{T} x_0}{\mathbb{V}[X_1|Y_1]}$ is the kl-divergence between the conditional prior distribution and the law of $X_1|Y_1$.

When minimizing Eq. (25), several design choices appear, as detailed in Sec. F.2. In particular, we may predict the noise, the signal, or a combination of both, leading to different parameterizations of the regression target and the score model. Empirically, we find that the neural-network parameterization introduced by Karras et al. (2022) and generalized to our setting in Sec. F.2 yields the best performance.

Prior work on diffusion-based generative models with fractional noise fixes the Hurst index H a priori, before training (Nobis et al., 2024; Tong et al., 2022). Consequently, evaluating different levels of memory requires retraining the model for each choice of H. We propose to provide the Hurst index as an additional condition to the neural network and train over the range $\mathcal{I}_H = [0.05, 0.995]$. Since the prior loss term does not depend on the weights θ we minimze during training

$$\mathcal{L}_{\text{GFI}}(\theta) := \int_{\mathcal{I}_H} \int_0^1 \lambda^2(t, H)\mathcal{L}_t(\theta, H)\,\mathrm{d}t\,\mathrm{d}H. \tag{26}$$

For sampling we can now choose any $H \in \mathcal{I}_H$, selecting from trajectories illustrated in Fig. 9 for GFI_{VP} and Fig. 8 for GFI_{FM}. We now optimize over the family of augmented forward processes $(Z_H)_{H\in(0,1)}$, whose covariance matrices depend on the Hurst index. To avoid repeatedly solving

and differentiating through the covariance ODEs during training, we precompute $(t, H) \mapsto \Sigma_H(t)$ on interpolation nodes and approximate it with Chebyshev polynomials in Sec. H, enabling gradient flow with respect to H. Training over the range of Hurst indices moves the task of choosing an optimal Hurst index from training to sampling time and allows for optimization according to different criteria post-hoc. We find that the best performance is achieved by selecting in each subintervall the Hurst index with the lowest distance to the prior.

6 EXPERIMENTS

Our experimental setup is described in detail in Section I. For all methods under consideration, we normalize both the input and output of the neural network to unit variance, following Karras et al. (2022). The derivations for combining our framework with the preconditioning and neural-network parameterization of Karras et al. (2022) are provided in Sec. F.2. We use two variants of the Diffusion Transformer (DiT) architecture (Peebles & Xie, 2023) as the time-dependent backbone neural network, providing a scalable architecture for latent diffusion (Rombach et al., 2022) and large-scale datasets. For small-scale ablations, we use DiT-S/4, whereas for the final experiments on each dataset under consideration, we use DiT-B/2. We follow the advice of Stein et al. (2023) and use FD-DINOv2 as our primary image quality metric to select the best configuration, but we also report FID to allow comparison with previous work. Since the FID often deteriorates while the FD-DINOv2 further falls during training, we also report the lowest achieved FID as FID*.

Comparing fractional drivers. We first show empirically on MNIST and CIFAR-10 that the proposed (Type I) ε-MA-fBM yields better performance compared to the (Type II) MA-fBM driver used in generative fractional diffusion models (GFDM) (Nobis et al., 2024) when paired with their best-performing fractional variance preserving (FVP) noise schedule. Additionally, we compare against score-based diffusion models (SBDMs) (Song et al., 2021) with the corresponding variance preserving (VP) dynamics driven by Brownian motion, using the best-performing configuration for this schedule from Tab. 5 and Tab. 6. All methods are trained for 400k steps using DiT-S/4 with a fixed random seed and evaluated using 1000 Euler–Maruyama (Cohen & Elliott, 2015) sampling steps. In Tab. 4 we observe that on MNIST, the best ε-MA-fBM configuration achieves an FD-DINOv2 score of 7.20, compared with 10.27 for MA-fBM; on CIFAR-10, ε-MA-fBM achieves 614.17, compared with 687.03 for MA-fBM. This trend is further reflected in Tab. 7 and Tab. 8: among all 80 tested configurations (40 with MA-fBM and 40 with ε-MA-fBM), ranked by FD-DINOv2, the best MA-fBM configuration attains only rank 19 on MNIST and rank 22 on CIFAR-10, with 18 and 21 ε-MA-fBM configurations, respectively, achieving better performance. In our view, these results provide empirical evidence in favor of using ε-MA-fBM rather than MA-fBM as the driver for generative models. We therefore focus on ε-MA-fBM in the remaining experiments, where we train across the full range of Hurst indices, and compare it against Brownian-motion-driven diffusion models and flow matching.

Balanced data in pixel space. For the comparison of GFI against Brownian driven baselines on CIFAR-10, we first train and evaluate DiT-S/4 paired with 39 baseline configurations of diffusion models with different noise schedules, loss weightings, and target parameterizations, including noise prediction, signal prediction, and interpolations between the two as introduced by Karras et al. (2022). The four best performing configurations reported in Tab. 6 are then trained in a second step for 400k steps with DiT-B/2. For the comparison of GFI against flow matching (FM) with linear interpolation between data and noise we train FM directly with DiT-B/2 uniform as well as variance loss weighting. We train GFI over the numerical stable Hurst range $H \in [0.01, 0.995]$, which removes H from the training-time design space, leaving only $K \in \mathbb{N}$. The hurst index H is chosen at inference by prior loss.

Tab. 1 reports the three strongest baselines and GFI with the VP and FM schedules trained on CIFAR-10. GFI$_{VP}$ reduces FD-DINOv2 by 13% relative to the best diffusion baseline and attains the highest precision, at the cost of a lower recall in the table (0.51 vs. 0.55). GFI$_{FM}$ attains the lowest FID (4.77, 18% below the best baseline) and the highest recall (0.59), but its FD-DINOv2 is 4% worse than the best baseline. Diversity as measured by the Vendi score is unchanged across all models. Both schedules select H in the smooth regime (0.73 and 0.93).

Unbalanced data in pixel space. Yoon et al. (2023) and Shariatian et al. (2025) argue that Brownian-driven diffusion models tend to mode collapse or generate samples of worse quality on heavy tailed datasets. We therefore benchmark our method against baselines on the heavy-tailed dataset CIFAR-10-LT, introduced by Cao et al. (2019). While we train on this subsampled version of CIFAR-10 that contains $[5000, 2997, 1796, 1077, 645, 387, 232, 139, 83, 50]$ samples for its 10 classes, we evaluate on the full CIFAR-10 dataset by measuring image quality in conjunction with Improved Precision and Recall. To detect memorization we additionally report the memorization penalty introduced by Bai et al. (2021). In table 3 we compare the best three baselines in terms of FD-DINOv2 against GFI with FVP and FM schedules. Over all reported configurations GFI increases the Improved Precision, while having an Improved Recall on par with FM but below VP. While the baselines don't memorize the training data, GFI with both schedules memorizes the training data at low K. For higher approximation quality $K > 2$ GFIs with the FM schedule don't memorize the data anymore and still reach up to a 20% lower FD-DINOv2 and a higher precision than the best baseline.

Latent space celeba. To show that the advantages of GFIs found in pixel datasets transfer to higher-resolution, latent space setting, we train GFI on the 256×256-pixel CelebA-HQ dataset in the latent space of the SDXL encoder and compare it against the best three baselines. Results are in Table 2. GFI improves over the best baseline by 8% in FD-DINOv2 (176.6 vs. 192.2) and 9% in FID (7.21 vs. 7.92) for all $K \geq 6$, while matching the baselines in improved precision, recall and Vendi score. Performance plateaus for $K \geq 6$ since the FD-DINOv2 of $K = 6, \ldots, 10$ lies within 176–180, a range we consider within run-to-run variation. The selected Hurst index is consistently smooth ($H \approx 0.92$–0.95), whereas on CIFAR-10-LT the non-memorizing configurations preferred a rough regime ($H \approx 0.3$). We hypothesize that latent embeddings form a smoother manifold and therefore profit from smoother sampling paths. We leave a systematic study to future work.

Table 1: Quantitative results on CIFAR10, where we train GFI over the Hurst range $(0.01, 0.995)$ and evaluate at $H_{\text{rough}} \in [0.01, 0.4]$, $H_{\text{smooth-1}} \in [0.5, 0.79]$ and $H_{\text{smooth-2}} \in (0.8, 0.99)$ with the smallest prior loss.

		CIFAR-10				
Model	H	FID* ↓	$\text{FD}_{\text{DINOv2}}$ ↓	Precision ↑	Recall ↑	Vendi ↑
Diffusion FM	-	5.80	365.16	0.67	0.55	17.42
VP	-	6.15	394.78	0.66	0.55	**17.60**
Flow matching	-	6.87	399.17	0.65	0.55	17.11
$\text{GFI}_{VP(K=4)}$	0.73	5.30	**316.65**	**0.69**	0.51	17.44
$\text{GFI}_{FM(K=8)}$	0.93	**4.77**	380.81	0.65	**0.59**	17.19

Table 2: Quantitative results on CelebA-HQ-256, where we train GFI over the Hurst range $(0.01, 0.995)$ and evaluate at $H_{\text{rough}} \in [0.01, 0.4]$, $H_{\text{smooth-1}} \in [0.5, 0.79]$ and $H_{\text{smooth-2}} \in (0.8, 0.99)$ with the smallest prior loss.

		CelebA-HQ				
Schedule	H	FID* ↓	$\text{FD}_{\text{DINOv2}}$ ↓	Precision ↑	Recall ↑	Vendi ↑
Diffusion FM	-	8.02	192.23	0.78	0.33	7.25
VP	-	8.18	193.93	**0.80**	0.32	7.07
FM	-	7.92	213.45	0.75	**0.34**	**7.54**
$\text{GFI}_{FM(K=7)}$	0.92	7.23	**176.63**	**0.80**	**0.34**	7.33
$\text{GFI}_{FM(K=10)}$	0.95	**7.21**	177.46	**0.80**	0.31	7.21

7 Conclusion

We propose Generative Fractional Interpolants that inherit the interpolation of Type I fBM between Brownian motion and random straight line paths by introducing ε-MA-fBM, an improved fBM approximation that preserves this limiting behavior while reducing the overall approximation error. The forward process interpolates between four regimes: subdiffusive dynamics, regular diffusion, superdiffusive dynamics, and, as $H \to 1$, random straight-line paths used in flow matching. We train all regimes at once by randomly sampling the corresponding Hurst index in each training step, leaving the regime a modeling choice for sampling time. Notably, unlike flow matching, GFI can achieve straight generative trajectories from noise to data, at the cost of trajectory crossings. Across CIFAR-10, CIFAR-10-LT, and latent-space CelebA-HQ-256, this mdodeling choice achieves lower FD-DINOv2 scores than diffusion models and flow matching in our experiments.

Limitations and future work. We emphasize three limitations of our work. First, our framework does not learn an optimal transport coupling between the data and noise distributions, and crossings between generative trajectories can occur, which may be viewed as a drawback relative to flow matching. Second, due to computational constraints, we were unable to repeat all experiments sufficiently many times to report statistically robust estimates of the observed performance differences. Third, we do not establish convergence of the approximate forward process to the corresponding solution of the fBM-driven SDE. In future work, we aim to investigate how optimal transport can be incorporated into GFI and to close the remaining theoretical gaps regarding convergence.

AI USE STATEMENT

In this work, we used generative AI tools for assistance in writing, assistance in theory development and assistance in proving claims the authors came up with. We have not used generative AI tools for generating synthetic data sets, formulating mathematical claims, providing critical ingredients for proving mathematical claims, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, assisting with translation, cleaning and reformatting datasets, supporting qualitative and thematic data analysis or interpreting results. Additionally, we used generative AI tools for the implementation of research code. All LLM-generated code was verified and tested by at least two authors. We take full responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149.

Ching-Yuan Bai, Hsuan-Tien Lin, Colin Raffel, and Wendy Chi-wen Kan. On training sample memorization: Lessons from benchmarking generative modeling with a large-scale competition. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 2534–2542. ACM, August 2021. doi: 10.1145/3447548.3467198. URL http://dx.doi.org/10.1145/3447548.3467198.

Christian Bayer and Simon Breneis. Markovian approximations of stochastic volterra equations with the fractional kernel. *Quantitative Finance*, 23(1):53–70, 2023. doi: 10.1080/14697688.2022.2139193. URL https://doi.org/10.1080/14697688.2022.2139193.

Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*. Springer-Verlag London Limited 2008, 2008. doi: 10.1007/978-1-84628-797-8.

Denis Blessing, Julius Berner, Lorenz Richter, and Gerhard Neumann. Underdamped diffusion bridges with applications to sampling. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Q1QTxFm0Is.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss, 2019. URL https://arxiv.org/abs/1906.07413.

Samuel N. Cohen and Robert J. Elliott. *Stochastic Calculus and Applications*. Probability and Its Applications. Birkhäuser, New York, NY, 2st edition, 2015. ISBN 978-1-4939-2866-8.

Rembert Daems, Manfred Opper, Guillaume Crevecoeur, and Tolga Birdal. Variational inference for SDEs driven by fractional noise. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=rtx8B94JMS.

Rembert Daems, Manfred Opper, Guillaume Crevecoeur, and Tolga Birdal. Efficient training of neural sdes using stochastic optimal control. In *ESANN*, 2025.

Sebastien Darses and Bruno Saussereau. Time Reversal for Drifted Fractional Brownian Motion with Hurst Index $H > 1/2$. *Electronic Journal of Probability*, 12(none):1181 – 1211, 2007. doi: 10.1214/EJP.v12-439.

Mostafa Dehghani, Josip Djolonga, Basil Mustafa, Piotr Padlewski, Jonathan Heek, Justin Gilmer, Andreas Steiner, Mathilde Caron, Robert Geirhos, Ibrahim Alabdulmohsin, Rodolphe Jenatton, Lucas Beyer, Michael Tschannen, Anurag Arnab, Xiao Wang, Carlos Riquelme, Matthias Minderer, Joan Puigcerver, Utku Evci, Manoj Kumar, Sjoerd van Steenkiste, Gamaleldin F. Elsayed, Aravindh Mahendran, Fisher Yu, Avital Oliver, Fantine Huot, Jasmijn Bastings, Mark Patrick Collier, Alexey Gritsenko, Vighnesh Birodkar, Cristina Vasconcelos, Yi Tay, Thomas Mensink,

Alexander Kolesnikov, Filip Pavetić, Dustin Tran, Thomas Kipf, Mario Lučić, Xiaohua Zhai, Daniel Keysers, Jeremiah Harmsen, and Neil Houlsby. Scaling vision transformers to 22 billion parameters, 2023. URL https://arxiv.org/abs/2302.05442.

Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.

Dan Friedman and Adji Bousso Dieng. The Vendi Score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856.

Hans Föllmer. Time reversal on Wiener space. *Stochastic Processes - Mathematic and Physics, volume 1158 of Lecture Notes in Math*, pp. 119–129, 1986.

Ruiqi Gao, Emiel Hoogeboom, Jonathan Heek, Valentin De Bortoli, Kevin Patrick Murphy, and Tim Salimans. Diffusion models and gaussian flow matching: Two sides of the same coin. In *The Fourth Blogpost Track at ICLR 2025*, 2025. URL https://openreview.net/forum?id=C8Yyg9wy0s.

Philipp Harms and David Stefanovits. Affine representations of fractional processes with applications in mathematical finance. *Stochastic Processes and their Applications*, 129(4):1185–1228, 2019. ISSN 0304-4149.

Ulrich G. Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, 14(4):1188–1205, 1986.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.

Ioannis Karatzas and Steven E Shreve. *Brownian motion and stochastic calculus*. Graduate Texts in Mathematics. Springer, New York, NY, 2 edition, August 2004.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577. Curran Associates, Inc., 2022.

Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, and Samuli Laine. Analyzing and Improving the Training Dynamics of Diffusion Models . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24174–24184, Los Alamitos, CA, USA, June 2024. IEEE Computer Society. doi: 10.1109/CVPR52733.2024.02282. URL https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.02282.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Sangyun Lee, Beomsu Kim, and Jong Chul Ye. Minimizing trajectory curvature of ODE-based generative models. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 18957–18973. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/lee23j.html.

Paul Lévy. Random functions: general theory with special reference to Laplacian random functions. *University of California Publications in Statistics*, 1:331–390, 1953.

Xiao Liang, Wentao Ma, Eric Paquet, Herna Viktor, and Wojtek Michalowski. ProT-GFDM: A generative fractional diffusion model for protein generation. *Computational and Structural Biotechnology Journal*, 27:3464–3480, 2025. URL https://doi.org/10.1016/j.csbj.2025.07.045.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling, 2023a.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos Theodorou, Weili Nie, and Anima Anandkumar. I^2SB: Image-to-image schrödinger bridge. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22042–22062. PMLR, 2023a. URL https://proceedings.mlr.press/v202/liu23ai.html.

Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=XVjTT1nw5z.

Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation, 2024. URL https://arxiv.org/abs/2309.06380.

Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian Motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

William McLean. Exponential sum approximations for $t^{-\beta}$. In *Contemporary Computational Mathematics – A Celebration of the 80th Birthday of Ian Sloan*, pp. 911–930. Springer International Publishing, Cham, 2018. ISBN 978-3-319-72456-0. doi: 10.1007/978-3-319-72456-0_40. URL https://doi.org/10.1007/978-3-319-72456-0_40.

Gabriel Nobis, Maximilian Springenberg, Marco Aversa, Michael Detzel, Rembert Daems, Roderick Murray-Smith, Shinichi Nakajima, Sebastian Lapuschkin, Stefano Ermon, Tolga Birdal, Manfred Opper, Christoph Knochenhauer, Luis Oala, and Wojciech Samek. Generative fractional diffusion models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=B9qg3wo75g.

Gabriel Nobis, Maximilian Springenberg, Arina Belova, Rembert Daems, Christoph Knochenhauer, Manfred Opper, Tolga Birdal, and Wojciech Samek. Fractional diffusion bridge models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=Hhc5McwASX.

Eric Paquet, Farzan Soleymani, Herna Lydia Viktor, and Wojtek Michalowski. Annealed fractional lévy–itō diffusion models for protein generation. *Computational and Structural Biotechnology Journal*, 23:1641–1653, 2024. ISSN 2001-0370. doi: https://doi.org/10.1016/j.csbj.2024.04.009. URL https://www.sciencedirect.com/science/article/pii/S2001037024000928.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022. URL https://doi.org/10.1109/CVPR52688.2022.01042.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models, 2022. URL https://arxiv.org/abs/2202.00512.

Dario Shariatian, Umut Simsekli, and Alain Oliviero Durmus. Denoising levy probabilistic models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=SYmUS6qRub.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265. PMLR, 2015.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 2023.

George Stein, Jesse C. Cresswell, Rasa Hosseinzadeh, Yi Sui, Brendan Leigh Ross, Valentin Villecroze, Zhaoyan Liu, Anthony L. Caterini, J. Eric T. Taylor, and Gabriel Loaiza-Ganem. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

R. L. Stratonovich. Conditional Markov Processes. *Theory of Probability & Its Applications*, 5(2): 156–178, 1960.

Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*, volume 10. Cambridge University Press, 2019.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport, 2024. URL https://arxiv.org/abs/2302.00482.

Anh Tong, Thanh Nguyen-Tang, Toan Tran, and Jaesik Choi. Learning fractional white noises in neural stochastic differential equations. In *Advances in Neural Information Processing Systems*, volume 35, pp. 37660–37675. Curran Associates, Inc., 2022.

Eunbi BI Yoon, Keehun Park, Sungwoong Kim, and Sungbin Lim. Score-based generative models with Lévy processes. In *Advances in Neural Information Processing Systems*, volume 36, pp. 40694–40707. Curran Associates, Inc., 2023.

APPENDIX OF GENERATIVE FRACTIONAL INTERPOLANTS

A Quantitative results on imbalanced data **15**

B Quantitative comparison of fractional drivers **15**

C Mathematical details of the interpolation between diffusion and flow models **15**

C.1 Limiting behavior of fBM as $H \to 1$ . . . 16

C.2 Endpoint collapse of Type I fBM . . . 18

C.3 Type II fBM converges to a stochastic process for $H \to 1$ . . . 19

C.4 Intermediate results . . . 21

C.5 Convergence for general μ and g . . . 23

C.6 The canonical straight-line construction . . . 24

C.7 ODE limit of fractional variance preserving dynamics . . . 28

D An improved Markovian approximation of Type I fBM **30**

E Derivation of the KL divergence **34**

F Simplified augmented score matching **38**

F.1 KL-prior . . . 40

F.2 EDM . . . 41

G Flow-matching, OT-ODE and PF-ODE **44**

H Details on the covariance interpolation **46**

I Experimental setup **48**

J Additional experiments **49**

J.1 Toy examples . . . 49

J.2 Baseline on MNIST . . . 49

J.3 Additional experiments on CIFAR-10 with DiT-S/4 . . . 52

J.4 Additional experiments on CIFAR-10 with DiT-B/2 . . . 55

J.5 Additional experiments on CIFAR-10-LT with DiT-B/2 . . . 55

J.6 Additional experiments on celeba with DiT-B/2 . . . 56

J.7 Additional experiments for comparison against GFDM . . . 56

A Quantitative results on imbalanced data

Table 3: Quantitative results on CIFAR10-LT, where we train GFI over the Hurst range $(0.01, 0.995)$ and evaluate at $H_{\text{rough}} \in [0.01, 0.4]$, $H_{\text{smooth-1}} \in [0.5, 79]$ and $H_{\text{smooth-2}} \in (0.8, 0.99)$ with the smallest prior loss. Configurations with memory penalty greater than 1 are excluded from bolding.

		CIFAR-10-LT							
Schedule	Driving noise	H	K	FID* ↓	FD-DINOv2 ↓	Mem. penalty (DINOv2) ↓	Precision ↑	Recall ↑	Vendi Score ↑
FM	random straight-lines	-	-	15.29	195.42	1.00	0.88	0.40	19.61
VP	BM	-	-	18.02	280.96	1.00	0.70	**0.61**	17.32
Diffusion FM	BM	-	-	15.70	297.97	1.00	0.80	0.34	19.52
$\text{GFI}_{VP(K=1)}$	ε-MA-fBM	0.80	1	8.60	119.35	13.99	0.97	0.38	19.24
$\text{GFI}_{VP(K=2)}$	ε-MA-fBM	0.80	2	8.59	116.26	14.18	0.97	0.38	19.34
$\text{GFI}_{FM(K=1)}$	ε-MA-fBM	0.80	1	7.83	111.31	12.38	0.98	0.36	**19.65**
$\text{GFI}_{FM(K=2)}$	ε-MA-fBM	0.80	2	7.58	109.97	13.08	0.98	0.37	19.36
$\text{GFI}_{FM(K=3)}$	ε-MA-fBM	0.32	3	**11.51**	**156.52**	1.00	**0.92**	0.38	19.51
$\text{GFI}_{FM(K=4)}$	ε-MA-fBM	0.31	4	14.15	183.34	1.00	0.90	0.39	19.50

B Quantitative comparison of fractional drivers

Table 4: Best performance achieved by different drivers on MNIST and CIFAR-10 under fixed VP dynamics over $K \in \{2, 3, 4, 5\}$ and $H \in \{0.1, 0.2, \ldots, 0.9, 0.99\}$. The proposed ε-MA-fBM achieves better overall performance in terms of FD-DINOv2 and FID on both MNIST and CIFAR-10 than the MA-fBM used in Nobis et al. (2024). A more detailed comparison is provided in Tabs. 7 and 8, where the best MA-fBM configuration ranks only 19th on MNIST and 22nd on CIFAR-10 among all 80 tested configurations.

		CIFAR-10					MNIST				
Method	Driving noise	H	K	FD-DINOv2 ↓	FID ↓	Recall ↑	H	K	FD-DINOv2 ↓	FID ↓	Recall ↑
SBDM_{VP}	BM	0.5	-	698.68	12.71	0.55	0.5	-	10.82	1.13	0.73
GFDM_{VP}	MA-fBM	0.7	3	687.03	12.47	**0.57**	0.4	3	10.27	1.03	0.75
GFI_{VP} (ours)	ε-MA-fBM	0.8	3	**614.17**	**11.47**	0.54	0.9	4	**7.20**	**0.88**	**0.76**

C Mathematical details of the interpolation between diffusion and flow models

In this section, we provide the mathematical details of the interpolation of fBM between Brownian motion and random straight-line paths, and show how this interpolation induces an interpolation of an fBM-driven forward process between a continuous-time diffusion model and a flow model. We begin by recalling the definition of Type I fBM via its integral representation and its equivalent characterization through the covariance function.

Definition 3 (Type I Fractional Brownian Motion (Mandelbrot & Van Ness (1968))) *Define for $H \in (0, 1)$ the normalizing constant*

$$V_H = \frac{\Gamma(2-2H)\text{sinc}(H-1/2)}{2H}, \quad \text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \tag{27}$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space equipped with a complete and right continuous filtration $\{\mathcal{F}_t\}$ and Γ the Gamma function. For a two sided $\{\mathcal{F}_t\}$-Brownian motion B the centered Gaussian process $B_H^{(I)} = (B_H^{(I)}(t))_{t\in[0,T]}$ with

$$B_H^{(I)}(t) := \frac{V_H^{-1/2}}{\Gamma(H+\frac{1}{2})} \int_{-\infty}^{0} \left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right) \mathrm{d}B(s) + \frac{V_H^{-1/2}}{\Gamma(H+\frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} \,\mathrm{d}B(s) \tag{28}$$

uniquely characterized in law by its covariances

$$\text{Cov}\big(B_H^{(I)}(t), B_H^{(I)}(s)\big) = \tfrac{1}{2}\big(t^{2H} + s^{2H} - |t-s|^{2H}\big), \quad 0 < s \le t \le T. \tag{29}$$

is called Mandelbrot–Van Ness fractional Brownian motion (Type I fBM) with Hurst index $H \in (0, 1)$.

Type I fBM has stationary increments and includes the history on $(-\infty, 0)$, whereas the history of Type II fBM defined below begins at 0.

Definition 4 (Type II Fractional Brownian Motion (Lévy (1953))) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space equipped with a complete and right continuous filtration $\{\mathcal{F}_t\}$ and Γ the Gamma function. For a $\{\mathcal{F}_t\}$-Brownian motion B and $H \in (0,1)$, the centered Gaussian process $B_H^{(II)} = (B_H^{(II)}(t))_{t\in[0,T]}$ with*

$$B_H^{(II)}(t) := \frac{1}{\Gamma(H+\frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} \, \mathrm{d}B(s) \tag{30}$$

uniquely characterized in law by its covariances

$$\mathrm{Cov}\big(B_H^{(II)}(t), B_H^{(II)}(s)\big) = \frac{1}{\Gamma^2(H+\frac{1}{2})} \int_0^{\min\{t,s\}} ((t-u)(s-u))^{H-\frac{1}{2}} \quad 0 < s \le t \le T. \tag{31}$$

is called Riemann-Liouville fractional Brownian motion (Type II fBM) with Hurst index $H \in (0,1)$.

C.1 Limiting behavior of fBM as $H \to 1$

We now contrast the two types of fBM through their differing limiting behavior, recalling two well-known results and their proofs. Comparing Type I fBM in $L^2(\mathbb{P})$ accross different choices of the Hurst index $H \in (0,1)$ requires them to be defined on one common probability space.

A1 (Common-terminal coupling) *The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ carries $\{B_H^{(I)} : H \in (1/2, 1)\}$ and a random variable $\varepsilon \sim \mathcal{N}(0,1)$ such that, for every $H \in (1/2,1)$, $B_H^{(I)}$ is a Type I fractional Brownian motion of Hurst parameter H and*

$$B_H^{(I)}(1) = \varepsilon \qquad \textit{almost surely}. \tag{32}$$

For $H = 1$, we set $B_1^{(I)}(t) := t\varepsilon$.

Remark 1 (Existence of this coupling) *The coupling in A1 can always be constructed. Indeed, let*

$$R_H(s,t) := \frac{1}{2}\big(s^{2H} + t^{2H} - |t-s|^{2H}\big), \qquad c_H(t) := R_H(t,1),$$

and let β^H be a centered Gaussian bridge, independent of ε, with covariance

$$\mathrm{Cov}(\beta^H(s), \beta^H(t)) = R_H(s,t) - c_H(s)c_H(t).$$

The last kernel is positive semidefinite because it is the conditional covariance of a fractional Brownian motion given its terminal value. Then

$$B_H^{(I)}(t) := c_H(t)\varepsilon + \beta^H(t)$$

has covariance R_H and satisfies $B_H^{(I)}(1) = \varepsilon$ almost surely. Thus it is a Type I fractional Brownian motion with the required common terminal value.

We now show that Type I fBM converges towards a random straight-line, as $H \to 1$.

Proposition 7 *Assume A1. For every $t \in [0,1]$, Type I fBM collapses onto a random straight line with a Gaussian distributed slope:*

$$B_H^{(I)}(t) \xrightarrow[H\to 1]{L^2(\mathbb{P})} \varepsilon\, t, \qquad \varepsilon \sim \mathcal{N}(0,1). \tag{33}$$

Proof. Consider $B_H(t)$ for $t = 1$. Since Type 1 fBM is a centered Gaussian process with covariance

$$\mathrm{Cov}(B_H(t), B_H(s)) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right),$$

its variance at $t = 1$ is $\mathrm{Var}(B_H^{(I)}(t)) = \frac{1}{2}(1^{2H} + 1^{2H} - |1-1|^{2H}) = 1$, hence $B_H^{(I)}(1) \sim \mathcal{N}(0,1)$. We now show that, as $H \to 1$, $B_H(t)$ approaches the straight line connecting the origin to $B_H(1)$.

Since B_H is centered, its second moments coincide with its covariances. Therefore,

$$\begin{aligned}||B_H^{(I)}(t) - B_H^{(I)}(1)t||^2_{L^2(\mathbb{P})} &= \mathbb{E}[(B_H^{(I)}(t) - B_H^{(I)}(1)t)^2] \\ &= \mathbb{E}[(B_H^{(I)}(t))^2] - 2t\mathbb{E}[B_H^{(I)}(t)B_H^{(I)}(1)] + t^2\mathbb{E}[(B_H^{(I)}(1))^2] \\ = \mathrm{Cov}\big(B_H^{(I)}(t), B_H^{(I)}(t)\big) - 2t\mathrm{Cov}\big(B_H^{(I)}(t), B_H^{(I)}(1)\big) + t^2\mathrm{Cov}\big(B_H^{(I)}(1), B_H^{(I)}(1)\big) \\ &= t^{2H} - t(t^{2H} + 1 - |t-1|^{2H}) + t^2.\end{aligned}$$

Taking the limit as $H \to 1$ gives

$$\begin{aligned}\lim_{H\to 1} t^{2H} - t(t^{2H} + 1 - |t-1|^{2H}) + t^2 &= t^2 - t(t^2 + 1 - (t-1)^2) + t^2 \\ &= t^2 - t(t^2 + 1 - t^2 + 2t - 1) + t^2 = t^2 - 2t^2 + t^2 = 0.\end{aligned}$$

Thus

$$\Big\| B_H^{(I)}(t) - tB_H^{(I)}(1) \Big\|_{L^2(\mathbb{P})} \longrightarrow 0.$$

Since $B_H^{(I)}(1) = \varepsilon$ almost surely under the common-terminal coupling, this is exactly equation 33. □

In contrast to the limiting behavior observed in Eq. (33), Type II fBM converges to a stochastic integral and thus does not collapse to a random straight-line path, as shown below.

Proposition 8 *For every $t \in [0,1]$, Type II fBM converges to a Gaussian process*

$$B_H^{(II)}(t) \xrightarrow[H\to 1]{L^2(\mathbb{P})} G(t), \qquad G(t) \sim \mathcal{N}(0, \frac{2}{\pi}t^2). \tag{34}$$

Proof. Consider the stochastic process $G(t) = \frac{2}{\sqrt{\pi}} \int_0^t \sqrt{t-s} dB_s$. To motivate this choice, observe that

$$\begin{aligned}\lim_{H\to 1} \mathrm{Cov}\big(B_H^{(II)}(t), B_H^{(II)}(s)\big) &= \lim_{H\to 1} \frac{1}{\Gamma^2(H+\frac{1}{2})} \int_0^{\min\{t,s\}} \big((t-u)(s-u)\big)^{H-\frac{1}{2}} \,\mathrm{d}u. \\ &= \frac{4}{\pi} \int_0^{\min\{t,s\}} \sqrt{(t-u)(s-u)}\,\mathrm{d}u\end{aligned}$$

and recall that the Itô isometry respects the inner product. We used $\Gamma(3/2) = \sqrt{\pi}/2$ and dominated convergence. The domination is uniform for H in any interval $[H_0, 1]$ with $H_0 > 1/2$, because $0 \le t-u, s-u \le 1$ so the limit can be interchanged with the integral. Moreover, it follows that

$$G(t) \sim \mathcal{N}\big(0, \frac{4}{\pi}\int_0^t (t-s)ds\big)$$

and

$$\frac{4}{\pi}\int_0^t (t-s)ds = \frac{4}{\pi}(-\frac{1}{2}t^2 + t^2) = \frac{2}{\pi}t^2$$

so

$$G(t) \sim \mathcal{N}\big(0, \frac{2}{\pi}t^2\big).$$

It remains to show that $B_H^{(II)}(t) \xrightarrow[H\to 1]{L^2(\mathbb{P})} G(t)$. Since both processes are constructed from the same Brownian motion, the difference can be written as one Itô integral. Hence

$$
\begin{aligned}
||B_H^{(II)}(t) - G(t)||^2_{L^2(\mathbb{P})} &= \mathbb{E}[(B_H^{(II)}(t) - \frac{2}{\sqrt{\pi}}\int_0^t \sqrt{t-s}dB_s)^2] \\
&= \mathbb{E}[\big(\int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} - \frac{2}{\sqrt{\pi}}\sqrt{t-s}dB_s\big)^2] \\
&\overset{\text{Itô isometry}}{=} \mathbb{E}[\int_0^t \big(\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} - \frac{2}{\sqrt{\pi}}\sqrt{t-s}\big)^2 ds] \\
&= \int_0^t \big(\frac{(t-s)^{2H-1}}{\Gamma^2(H+\frac{1}{2})} - \frac{4(t-s)^H}{\sqrt{\pi}\Gamma(H+\frac{1}{2})} + \frac{4(t-s)}{\pi}\big) ds \\
&= -\frac{(t-s)^{2H}}{2H\Gamma^2(H+\frac{1}{2})}\Big|_0^t + \frac{2(t-s)^{H+1}}{(H+1)\sqrt{\pi}\Gamma(H+\frac{1}{2})}\Big|_0^t - \frac{2(t-s)^2}{\pi}\Big|_0^t \\
&= \frac{t^{2H}}{2H\Gamma^2(H+\frac{1}{2})} - \frac{4t^{H+1}}{(H+1)\sqrt{\pi}\Gamma(H+\frac{1}{2})} + \frac{2t^2}{\pi}
\end{aligned}
$$

Again, taking the limit as $H \to 1$ yields

$$
\begin{aligned}
&\lim_{H\to 1} \frac{t^{2H}}{2H\Gamma^2(H+\frac{1}{2})} - \frac{4t^{H+1}}{(H+1)\sqrt{\pi}\Gamma(H+\frac{1}{2})} + \frac{2t^2}{\pi} \\
&= \frac{2t^2}{\pi} - \frac{-4t^2}{\pi} + \frac{2t^2}{\pi} = 0.
\end{aligned}
$$

Therefore $\|B_H^{(II)}(t) - G(t)\|_{L^2(\mathbb{P})} \to 0$, which proves equation 41. □

This section collects the limiting results used in the main text. We first prove that Type I fractional Brownian motion collapses onto the random line joining its endpoints. We then transfer this result to linear equations by a deterministic stability estimate. Finally, we apply the estimate to the canonical coefficient that produces the straight-line marginals $(1-t)X_0 + tZ$.

Throughout, $B^H = (B_t^H)_{t\in[0,1]}$ denotes standard one-dimensional Type I fractional Brownian motion with Hurst index $H \in (0,1)$ and covariance

$$
R_H(s,t) = \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right).
$$

We write B_1^H for its terminal value at time 1. We use $C_0([0,T]) := \{b \in C([0,T]) : b(0) = 0\}$. All proofs were done on the unit interval. This can easily be extended to an interval $[0,S]$ by scaling $B_{Su}^H \overset{d}{=} S^H B_u^H$ and the corresponding change of variables in the coefficients. The terminal slope is then $\frac{B_S^H}{S} \sim \mathcal{N}(0, S^{2H-2})$. All path-space statements use the uniform norm. The proofs are written in one dimension; the corresponding results in $\mathbb{R}^d$ follow componentwise for independent coordinates.

C.2 Endpoint collapse of Type I fBM

Proposition 9 (Endpoint collapse of Type I fBM) *Let $Z \sim \mathcal{N}(0,1)$. As $H \uparrow 1$,*

$$
B^H \Longrightarrow (tZ)_{t\in[0,1]} \qquad \text{in } C([0,1]). \tag{35}
$$

Moreover, if

$$
D_H := \sup_{t\in[0,1]} |B_t^H - tB_1^H|, \tag{36}
$$

then, for every $p \in [1,\infty)$,

$$
D_H \xrightarrow{\mathbb{P}} 0, \qquad \mathbb{E}[D_H^p] \longrightarrow 0. \tag{37}
$$

In addition,

$$
(B^H, B_1^H) \Longrightarrow ((tZ)_{t\in[0,1]}, Z) \quad \text{in } C([0,1]) \times \mathbb{R}. \tag{38}
$$

Proof. For every $s, t \in [0,1]$,

$$R_H(s,t) \longrightarrow \frac{1}{2}\left(s^2 + t^2 - |t-s|^2\right) = st.$$

Thus the finite-dimensional distributions converge to those of $t \mapsto tZ$. To prove tightness, fix $H_0 \in (0,1)$ and restrict to $H \in [H_0, 1)$. If $N \sim \mathcal{N}(0,1)$ and $m_r := \mathbb{E}|N|^r$, stationarity of the increments gives

$$\mathbb{E}|B_t^H - B_s^H|^r = m_r |t-s|^{rH} \le m_r |t-s|^{rH_0}.$$

Choosing r such that $rH_0 > 1$, the Kolmogorov–Chentsov criterion yields tightness in $C([0,1])$. This proves equation 35. The continuous map $b \mapsto (b, b(1))$ then gives equation 38.

Next consider the continuous functional

$$F : C([0,1]) \to [0,\infty), \qquad F(b) := \sup_{t\in[0,1]} |b(t) - tb(1)|.$$

Since $F(tZ) = 0$, the continuous mapping theorem shows that $D_H = F(B^H) \to 0$ in probability.

It remains to upgrade this convergence to L^p. Set $Y_t^H := B_t^H - tB_1^H$, so that $D_H = \|Y^H\|_\infty$ and $Y_0^H = 0$. For $s, t \in [0,1]$,

$$\mathbb{E}|Y_t^H - Y_s^H|^2 \le 2\mathbb{E}|B_t^H - B_s^H|^2 + 2|t-s|^2\mathbb{E}|B_1^H|^2 \le 4|t-s|^{2H_0}.$$

Because the increment is centered Gaussian, for every $r \ge 2$ there is a constant $C_{inc}(r)$ such that

$$\mathbb{E}|Y_t^H - Y_s^H|^r \le C_{inc}(r)|t-s|^{rH_0}, \tag{39}$$

uniformly over $H \in [H_0, 1)$.

Choose $r > \max\{p, 2, 1/H_0\}$, and choose $a \in (1/r, H_0)$. The one-dimensional fractional Sobolev embedding gives

$$\|Y^H\|_\infty^r \le C_{emb}(a,r)\left(\int_0^1 |Y_t^H|^r\,dt + \int_0^1\int_0^1 \frac{|Y_t^H - Y_s^H|^r}{|t-s|^{1+ar}}\,ds\,dt\right).$$

Taking expectations and using equation 39 gives

$$\mathbb{E}\left[\|Y^H\|_\infty^r\right] \le C_{emb}(a,r)C_{inc}(r)\left(1 + \int_0^1\int_0^1 |t-s|^{rH_0-1-ar}\,ds\,dt\right).$$

The last integral is finite because

$$r(H_0 - a) > 0.$$

The bound is uniform in $H \in [H_0, 1)$. Consequently,

$$\sup_{H\in[H_0,1)} \mathbb{E}[D_H^r] = \sup_{H\in[H_0,1)} \mathbb{E}[\|Y^H\|_\infty^r] < \infty. \tag{40}$$

Let $M > 0$. Since $r/p > 1$, equation equation 40 implies

$$\sup_{H\in[H_0,1)} \mathbb{E}\left[D_H^p \mathbf{1}_{\{D_H^p > M\}}\right] \le M^{-\frac{r-p}{p}} \sup_{H\in[H_0,1)} \mathbb{E}[D_H^r] \xrightarrow{M\to\infty} 0.$$

Thus $(D_H^p)_{H\in[H_0,1)}$ is uniformly integrable. Since $D_H^p \to 0$ in probability, Vitali's theorem gives $\mathbb{E}[D_H^p] \to 0$. □

C.3 Type II fBM converges to a stochastic process for $H \to 1$

Proposition 10 *Let B be a standard Brownian motion and construct*

$$B_H^{(II)}(t) := \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-s)^{H-\frac{1}{2}}\,dB_s, \qquad G(t) := \frac{2}{\sqrt{\pi}}\int_0^t \sqrt{t-s}\,dB_s$$

using the same Brownian motion for every H. For every $t \in [0,1]$, Type II fBM converges to a Gaussian process

$$B_H^{(II)}(t) \xrightarrow[H\to 1]{L^2(\mathbb{P})} G(t), \qquad G(t) \sim \mathcal{N}(0, \frac{2}{\pi}t^2). \tag{41}$$

Consequently, the finite-dimensional distributions of $B_H^{(II)}$ converge to those of the centered Gaussian process G, whose covariance is

$$\mathrm{Cov}(G(s), G(t)) = \frac{4}{\pi} \int_0^{\min\{t,s\}} \sqrt{(s-u)(t-u)}\, du. \tag{42}$$

Moreover, G is almost surely not a straight line.

Proof. Consider the stochastic process $G(t) = \frac{2}{\sqrt{\pi}} \int_0^t \sqrt{t-s} dB_s$. To motivate this choice, observe that

$$\lim_{H\to 1} \mathrm{Cov}\big(B_H^{(II)}(t), B_H^{(II)}(s)\big) = \lim_{H\to 1} \frac{1}{\Gamma^2(H+\frac{1}{2})} \int_0^{\min\{t,s\}} \big((t-u)(s-u)\big)^{H-\frac{1}{2}}\, \mathrm{d}u.$$
$$= \frac{4}{\pi} \int_0^{\min\{t,s\}} \sqrt{(t-u)(s-u)}\, \mathrm{d}u$$

and recall that the Itô isometry respects the inner product. We used $\Gamma(3/2) = \sqrt{\pi}/2$ and dominated convergence. The domination is uniform for H in any interval $[H_0, 1]$ with $H_0 > 1/2$, because $0 \le t-u, s-u \le 1$ so the limit can be interchanged with the integral. Since $G(t)$ is an Itô integral with a deterministic integrand, it is centered Gaussian. By Itô isometry,

$$\mathrm{Var}(G(t)) = \frac{4}{\pi} \int_0^t (t-s)\, \mathrm{d}s = \frac{2t^2}{\pi}.$$

Hence $G(t) \sim \mathcal{N}(0, 2t^2/\pi)$. It remains to show that $B_H^{(II)}(t) \xrightarrow[H\to 1]{L^2(\mathbb{P})} G(t)$. Since both processes are constructed from the same Brownian motion, the difference can be written as one Itô integral. Hence

$$||B_H^{(II)}(t) - G(t)||^2_{L^2(\mathbb{P})} = \mathbb{E}[(B_H^{(II)}(t) - \frac{2}{\sqrt{\pi}} \int_0^t \sqrt{t-s} dB_s)^2]$$
$$= \mathbb{E}[\Big(\int_0^t \left[\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} - \frac{2}{\sqrt{\pi}}\sqrt{t-s}\right] dB_s\Big)^2]$$
$$\overset{\text{Itô isometry}}{=} \mathbb{E}[\int_0^t \big(\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} - \frac{2}{\sqrt{\pi}}\sqrt{t-s}\big)^2 ds]$$
$$= \int_0^t \big(\frac{(t-s)^{2H-1}}{\Gamma^2(H+\frac{1}{2})} - \frac{4(t-s)^H}{\sqrt{\pi}\Gamma(H+\frac{1}{2})} + \frac{4(t-s)}{\pi}\big) ds$$
$$= -\frac{(t-s)^{2H}}{2H\Gamma^2(H+\frac{1}{2})}\Big|_0^t + \frac{4(t-s)^{H+1}}{(H+1)\sqrt{\pi}\Gamma(H+\frac{1}{2})}\Big|_0^t - \frac{2(t-s)^2}{\pi}\Big|_0^t$$
$$= \frac{t^{2H}}{2H\Gamma^2(H+\frac{1}{2})} - \frac{4t^{H+1}}{(H+1)\sqrt{\pi}\Gamma(H+\frac{1}{2})} + \frac{2t^2}{\pi}$$

Again, taking the limit as $H \to 1$ yields

$$\lim_{H\to 1} \frac{t^{2H}}{2H\Gamma^2(H+\frac{1}{2})} - \frac{4t^{H+1}}{(H+1)\sqrt{\pi}\Gamma(H+\frac{1}{2})} + \frac{2t^2}{\pi}$$
$$= \frac{2t^2}{\pi} - \frac{4t^2}{\pi} + \frac{2t^2}{\pi} = 0.$$

Therefore $\|B_H^{(II)}(t) - G(t)\|_{L^2(\mathbb{P})} \to 0$, which proves equation 41. Applying this argument to finitely many times gives finite-dimensional convergence. Finally, for any fixed $t \in (0,1)$, Itô's isometry gives

$$\mathrm{Var}(G(t) - tG(1)) = \frac{4}{\pi} \int_0^1 \left(\sqrt{(t-u)_+} - t\sqrt{1-u}\right)^2 du \ge \frac{2}{\pi} t^2 (1-t)^2 > 0,$$

where $(x)_+ := \max\{x, 0\}$. Thus the centered Gaussian variable $G(t) - tG(1)$ is non-degenerate, so $\mathbb{P}(G(t) = tG(1)) = 0$. Since $G(0) = 0$, this excludes straight-line paths almost surely. □

C.4 INTERMEDIATE RESULTS

Denote $BV([0,T])$ as the set of all functions with bounded variation and let $TV_{[0,T]}(f)$ be its value for $f \in BV([0,T])$.

Lemma 1 (Continuity of the integration-by-parts map) *Let $T \in (0,1]$ and let $q \in C([0,T]) \cap BV([0,T])$. For $b \in C_0([0,T])$, define*

$$\mathcal{I}_q(b)(t) := q(t)b(t) - \int_0^t b(s)\, dq(s), \qquad t \in [0,T]. \tag{43}$$

Then $\mathcal{I}_q : C_0([0,T]) \to C([0,T])$ is linear and

$$\|\mathcal{I}_q(b) - \mathcal{I}_q(\widetilde{b})\|_\infty \le \big(\|q\|_\infty + \mathrm{TV}_{[0,T]}(q)\big)\|b - \widetilde{b}\|_\infty. \tag{44}$$

Proof. For every $t \in [0,T]$,

$$\left| \int_0^t (b(s) - \widetilde{b}(s))\, dq(s) \right| \le \|b - \widetilde{b}\|_\infty \,\mathrm{TV}_{[0,T]}(q).$$

Combining this with

$$|q(t)(b(t) - \widetilde{b}(t))| \le \|q\|_\infty \|b - \widetilde{b}\|_\infty$$

and taking the supremum over t proves the claim. □

Theorem 1 (Stability under endpoint collapse) *Fix $T \in (0,1]$. Let $\mu \in L^1([0,T])$ and define*

$$\Phi(t) := \exp\left(\int_0^t \mu(r)\, dr \right).$$

Let $q_H \in C([0,T]) \cap BV([0,T])$ and let $q \in L^1([0,T])$ satisfy

$$M_T := \sup_{H \in [H_0, 1)} \big(\|q_H\|_\infty + \mathrm{TV}_{[0,T]}(q_H)\big) < \infty \tag{45}$$

for some $H_0 \in (0,1)$, and

$$\varepsilon_H := \int_0^T |q_H(s) - q(s)|\, ds \longrightarrow 0. \tag{46}$$

Let X_0 have an arbitrary Borel probability distribution and be independent of B^H. Define

$$X_t^H := \Phi(t)\left(X_0 + \mathcal{I}_{q_H}(B^H)(t)\right), \tag{47}$$

$$\widehat{X}_t^H := \Phi(t)\left(X_0 + B_1^H \int_0^t q(s)\, ds\right). \tag{48}$$

Writing $g_H := \Phi q_H$, the process X^H is the integration-by-parts solution of

$$dX_t^H = \mu(t) X_t^H\, dt + g_H(t)\, dB_t^H, \qquad X_0^H = X_0.$$

Then, with

$$D_{H,T} := \sup_{t \in [0,T]} |B_t^H - tB_1^H|,$$

we have the pathwise estimate

$$\|X^H - \widehat{X}^H\|_{C([0,T])} \le \|\Phi\|_\infty \left(M_T D_{H,T} + \varepsilon_H |B_1^H|\right). \tag{49}$$

Consequently, for every finite $p \ge 1$,

$$\mathbb{E}\|X^H - \widehat{X}^H\|_{C([0,T])}^p \longrightarrow 0, \qquad \|X^H - \widehat{X}^H\|_{C([0,T])} \xrightarrow{\mathbb{P}} 0. \tag{50}$$

If $Z \sim \mathcal{N}(0,1)$ is independent of X_0 and

$$X_t^\star := \Phi(t)\left(X_0 + Z \int_0^t q(s)\, ds\right),$$

then

$$(X^H, B_1^H) \Longrightarrow (X^\star, Z) \qquad \text{in } C([0,T]) \times \mathbb{R}. \tag{51}$$

Proof. Set

$$L_H(t) := tB_1^H, \qquad \mathcal{R}_H(t) := B_t^H - L_H(t).$$

Observe that

$$X_t^H - \widehat{X}_t^H = \Phi(t)\left(\mathcal{I}_{q_H}(B^H)(t) - B_1^H \int_0^t q(s)ds\right). \tag{52}$$

By linearity of $\mathcal{I}_{q_H}$ and integration by parts,

$$\mathcal{I}_{q_H}(B^H)(t) = \mathcal{I}_{q_H}(\mathcal{R}_H)(t) + \mathcal{I}_{q_H}(L_H)(t) = \mathcal{I}_{q_H}(\mathcal{R}_H)(t) + B_1^H \int_0^t q_H(s)\, ds.$$

Lemma 1 therefore gives

$$\sup_{t\in[0,T]} |\mathcal{I}_{q_H}(\mathcal{R}_H)(t)| \le M_T D_{H,T}.$$

Moreover,

$$\sup_{t\in[0,T]} \left| B_1^H \int_0^t (q_H(s) - q(s))\, ds \right| \le \varepsilon_H |B_1^H|.$$

Applying these bounds to 52 proves 49.

Set

$$Y_H := \|X^H - \widehat{X}^H\|_{C([0,T])}.$$

Taking $L^p(\mathbb{P})$-norms in equation 49 and applying Minkowski's inequality gives

$$\|Y_H\|_{L^p(\mathbb{P})} \le \|\Phi\|_{\infty,[0,T]} \left(M_T \|D_{H,T}\|_{L^p(\mathbb{P})} + \varepsilon_H \|B_1^H\|_{L^p(\mathbb{P})}\right).$$

Since $[0,T] \subseteq [0,1]$, we have

$$D_{H,T} = \sup_{t\in[0,T]} |B_t^H - tB_1^H| \le \sup_{t\in[0,1]} |B_t^H - tB_1^H| = D_H.$$

Consequently, Proposition 9 yields

$$\|D_{H,T}\|_{L^p(\mathbb{P})} \le \|D_H\|_{L^p(\mathbb{P})} = \left(\mathbb{E}[D_H^p]\right)^{1/p} \longrightarrow 0.$$

Moreover, $B_1^H \sim \mathcal{N}(0,1)$ for every H. Hence, writing

$$m_p := \mathbb{E}|N|^p, \qquad N \sim \mathcal{N}(0,1),$$

we have

$$\|B_1^H\|_{L^p(\mathbb{P})} = m_p^{1/p},$$

which is finite and independent of H. Since $\varepsilon_H \to 0$, it follows that

$$\varepsilon_H \|B_1^H\|_{L^p(\mathbb{P})} = \varepsilon_H m_p^{1/p} \longrightarrow 0.$$

The constants $\|\Phi\|_{\infty,[0,T]}$ and M_T are finite and independent of H. Therefore,

$$\|Y_H\|_{L^p(\mathbb{P})} \longrightarrow 0.$$

Equivalently,

$$\mathbb{E}\left[\|X^H - \widehat{X}^H\|^p_{C([0,T])}\right] \longrightarrow 0,$$

which proves equation 50. Convergence in probability follows immediately.

It remains to identify the joint weak limit. By assumption, X_0 is independent of B^H, and hence it is independent of B_1^H. Moreover,

$$B_1^H \sim \mathcal{N}(0,1)$$

for every H. Since $Z \sim \mathcal{N}(0,1)$ is independent of X_0, we therefore have

$$(X_0, B_1^H) \overset{d}{=} (X_0, Z).$$

Define the deterministic map

$$\Psi : \mathbb{R}^2 \longrightarrow C([0,T]) \times \mathbb{R}$$

by

$$\Psi(x,z) := \left(\left[\Phi(t) \left(x + z \int_0^t q(s)\, ds \right) \right]_{t\in[0,T]}, z \right).$$

Because

$$(\widehat{X}^H, B_1^H) = \Psi(X_0, B_1^H)$$

and

$$(X^\star, Z) = \Psi(X_0, Z),$$

the preceding identity in distribution implies

$$(\widehat{X}^H, B_1^H) \overset{d}{=} (X^\star, Z)$$

for every H. In particular,

$$(\widehat{X}^H, B_1^H) \Longrightarrow (X^\star, Z) \qquad \text{in } C([0,T]) \times \mathbb{R}.$$

On the other hand, equation 50 gives

$$\|X^H - \widehat{X}^H\|_{C([0,T])} \xrightarrow{\mathbb{P}} 0.$$

Consequently,

$$(X^H - \widehat{X}^H, 0) \xrightarrow{\mathbb{P}} (0,0) \qquad \text{in } C([0,T]) \times \mathbb{R}.$$

Since

$$(X^H, B_1^H) = (\widehat{X}^H, B_1^H) + (X^H - \widehat{X}^H, 0),$$

Slutsky's theorem yields

$$(X^H, B_1^H) \Longrightarrow (X^\star, Z) \qquad \text{in } C([0,T]) \times \mathbb{R},$$

which proves equation 51. □

C.5 Convergence for general μ and g

Corollary 1 (Linear equations with fixed coefficients) *Fix $T \in (0,1]$. Let $\mu \in L^1([0,T])$, let*

$$\Phi(t) := \exp\left(\int_0^t \mu(r)\, dr \right), \qquad q(t) := \frac{g(t)}{\Phi(t)},$$

and assume $q \in C([0,T]) \cap BV([0,T])$. Let X_0 have an arbitrary Borel probability distribution and be independent of B^H. Let $Z \sim \mathcal{N}(0,1)$ be independent of X_0, and define

$$X_t^H := \Phi(t) \left(X_0 + \mathcal{I}_q(B^H)(t) \right)$$

and

$$\widehat{X}_t^H := \Phi(t) \left(X_0 + B_1^H \int_0^t q(s)\, ds \right).$$

Then, for every finite $p \geq 1$,

$$\mathbb{E}\|X^H - \widehat{X}^H\|^p_{C([0,T])} \longrightarrow 0,$$

and

$$X^H \Longrightarrow X^\star \qquad \text{in } C([0,T]), \tag{53}$$

where

$$X_t^\star = \Phi(t) \left(X_0 + Z \int_0^t q(s)\, ds \right) \tag{54}$$

is the unique absolutely continuous solution of

$$dX_t^\star = \mu(t) X_t^\star\, dt + g(t) Z\, dt, \qquad X_0^\star = X_0.$$

Proof. Apply Theorem 1 with $q_H = q$. Then $\varepsilon_H = 0$ and

$$\|X^H - \widehat{X}^H\|_\infty \leq \|\Phi\|_\infty \big(\|q\|_\infty + \mathrm{TV}_{[0,T]}(q) \big) D_{H,T}.$$

The ODE identity follows by differentiating equation 54. For uniqueness, the difference W of two solutions with the same Z and initial value satisfies $W' = \mu W$ almost everywhere and $W(0) = 0$. Hence $(W/\Phi)' = 0$ almost everywhere, so $W = 0$. □

C.6 THE CANONICAL STRAIGHT-LINE CONSTRUCTION

We now specialize to the singular drift

$$\mu(t) = -\frac{1}{1-t}, \qquad 0 \le t < 1,$$

whose fundamental solution is $\Phi(t) = 1 - t$. The coefficients below are chosen so that the one-time marginal of the stochastic convolution has variance t^2 for every $H \in (1/2, 1)$.

Set

$$\alpha_H := H(2H-1), \qquad \mathcal{B}_H := \mathrm{B}(2-H, 2H-1), \qquad c_H := (\alpha_H \mathcal{B}_H)^{-1/2}. \tag{55}$$

Proposition 11 (Canonical coefficient and prescribed marginals) *For $H \in (1/2, 1)$, define*

$$q_H(t) := c_H t^{1-H}(1-t)^{-H-1}, \qquad g_H(t) := (1-t)q_H(t) = c_H t^{1-H}(1-t)^{-H}, \tag{56}$$

for $t \in (0, 1)$, and set $q_H(0) = g_H(0) = 0$. Then

$$q_H(t) \int_0^t q_H(u)(t-u)^{2H-2}\, du = \frac{t}{\alpha_H (1-t)^3}. \tag{57}$$

Moreover, $\psi_H(t) := (1-t)^{-H-1}$ is the unique positive function $\psi \in C([0,1))$ satisfying

$$\psi(t) \int_0^1 \psi(tv) w_H(v)\, dv = (1-t)^{-3}, \qquad w_H(v) := \frac{v^{1-H}(1-v)^{2H-2}}{\mathcal{B}_H}. \tag{58}$$

In particular, $q_H(t) = c_H t^{1-H}\psi_H(t)$ and $1 \le \psi_H(t) \le (1-t)^{-3}$. If

$$S_t^H := (1-t) \int_0^t q_H(u)\, dB_u^H, \tag{59}$$

then

$$\mathrm{Var}(S_t^H) = t^2, \qquad 0 \le t < 1. \tag{60}$$

Consequently, for X_0 independent of B^H, the process

$$X_t^H := (1-t)X_0 + S_t^H \tag{61}$$

satisfies

$$X_t^H = X_0 - \int_0^t \frac{X_s^H}{1-s}\, ds + \int_0^t g_H(s)\, dB_s^H \tag{62}$$

and has the prescribed marginals

$$X_t^H \overset{d}{=} (1-t)X_0 + tZ, \qquad Z \sim \mathcal{N}(0,1), \quad Z \perp X_0. \tag{63}$$

Proof. The beta-integral identity

$$\frac{1}{\mathcal{B}_H} \int_0^1 \frac{v^{1-H}(1-v)^{2H-2}}{(1-tv)^{H+1}}\, dv = (1-t)^{H-2} \tag{64}$$

follows, for example, from the substitution $y = (1-t)v/(1-tv)$. It also verifies equation 58 for ψ_H. For uniqueness, let $\widetilde{\psi}$ be another positive continuous solution and set $r = \widetilde{\psi}/\psi_H$. The equation at zero gives $r(0) = 1$, and division of the two equations gives

$$r(t) \int_0^1 r(tv)\, \nu_t(dv) = 1, \qquad \nu_t(dv) := \frac{\psi_H(tv) w_H(v)\, dv}{\int_0^1 \psi_H(tv) w_H(v)\, dv}.$$

If $r \not\equiv 1$, choose $T < 1$ and $0 < a < \max_{[0,T]} |\log r|$, and let τ be the first time $|\log r(\tau)| = a$. Then $\tau > 0$ and $e^{-a} < r(\tau v) < e^a$ for every $v \in (0,1)$. Since ν_τ is a probability measure on $(0,1)$, the same strict bounds hold for its integral. The preceding identity implies $|\log r(\tau)| < a$, a contradiction. Thus $\widetilde{\psi} = \psi_H$.

Changing variables $u = tv$ and using equation 64 gives

$$\int_0^t q_H(u)(t-u)^{2H-2}\,du = c_H t^H \mathcal{B}_H (1-t)^{H-2}.$$

Multiplication by $q_H(t)$ and the identity $c_H^2 \mathcal{B}_H = 1/\alpha_H$ prove equation 57.

On each $[0,T]$ with $T<1$, the integrals against B^H are interpreted via $\mathcal{I}_q$. For continuous integrands of bounded variation, these agree with the Gaussian Wiener integrals. For bounded deterministic integrands on $[0,t]$ and $H > 1/2$,

$$\operatorname{Cov}\left(\int f\,dB^H, \int h\,dB^H\right) = \alpha_H \iint f(u)h(v)|u-v|^{2H-2}\,du\,dv.$$

Define

$$Q_H(t) := \alpha_H \int_0^t \int_0^t q_H(u)q_H(v)|u-v|^{2H-2}\,du\,dv. \tag{65}$$

By symmetry, Tonelli's theorem, and equation 57,

$$Q_H'(t) = 2\alpha_H q_H(t) \int_0^t q_H(u)(t-u)^{2H-2}\,du = \frac{2t}{(1-t)^3}.$$

Since $Q_H(0) = 0$,

$$Q_H(t) = \frac{t^2}{(1-t)^2}.$$

Multiplying by $(1-t)^2$ proves equation 60.

To verify equation 62, set $Y_t^H := \int_0^t q_H(u)\,dB_u^H$. After expressing both sides by integration by parts, Fubini's theorem for Lebesgue measure and the finite signed measure dq_H gives

$$\int_0^t Y_s^H\,ds = \int_0^t (t-u)q_H(u)\,dB_u^H.$$

Therefore

$$-\int_0^t Y_s^H\,ds + \int_0^t (1-u)q_H(u)\,dB_u^H = (1-t)Y_t^H = S_t^H.$$

The deterministic initial-condition term satisfies

$$X_0 - \int_0^t X_0\,ds = (1-t)X_0,$$

which proves equation 62. Finally, S_t^H is centered Gaussian, independent of X_0, with variance t^2, proving equation 63. □

Theorem 2 (Convergence of the canonical process) *Let X_0 have an arbitrary Borel probability distribution and be independent of B^H. Let X^H be given by equation 61. Fix $T \in (0,1)$ and define*

$$\widehat{X}_t^H := (1-t)X_0 + tB_1^H, \qquad t \in [0,T]. \tag{66}$$

The same formula, extended to $t \in [0,1)$, is the unique solution of

$$d\widehat{X}_t^H = \frac{B_1^H - \widehat{X}_t^H}{1-t}\,dt, \qquad \widehat{X}_0^H = X_0. \tag{67}$$

Then, for every finite $p \geq 1$,

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^H - \widehat{X}_t^H|^p\right] \longrightarrow 0. \tag{68}$$

In particular,

$$\sup_{t\in[0,T]} |X_t^H - \widehat{X}_t^H| \xrightarrow{\mathbb{P}} 0. \tag{69}$$

Moreover,

$$\big((X_t^H)_{t\in[0,T]}, B_1^H\big) \Longrightarrow \big(((1-t)X_0 + tZ)_{t\in[0,T]}, Z\big) \tag{70}$$

in $C([0,T]) \times \mathbb{R}$*, where* $Z \sim \mathcal{N}(0,1)$ *is independent of* X_0*.*

The limiting path

$$X_t^\star = (1-t)X_0 + tZ$$

is the unique pathwise solution on $[0,1)$ *of*

$$dX_t^\star = \frac{Z - X_t^\star}{1-t}\,dt, \qquad X_0^\star = X_0. \tag{71}$$

Proof. On $[0,T]$, the functions q_H in equation 56 are increasing. Hence

$$\|q_H\|_{\infty,[0,T]} + \mathrm{TV}_{[0,T]}(q_H) = 2q_H(T).$$

Since $c_H \to 1$, the right-hand side is bounded uniformly for H close to one. Define

$$q_1(t) := (1-t)^{-2}.$$

For every fixed $s \in (0,T]$, we have

$$c_H \longrightarrow 1, \qquad s^{1-H} \longrightarrow 1, \qquad (1-s)^{-H-1} \longrightarrow (1-s)^{-2}.$$

Hence

$$q_H(s) = c_H s^{1-H}(1-s)^{-H-1} \longrightarrow (1-s)^{-2} = q_1(s).$$

Notice, that the convergence fails at $s = 0$, since $q_H(0) = 0$ whereas $q_1(0) = 1$.

Moreover, q_H is increasing on $[0,T]$, and therefore

$$\sup_{s\in[0,T]} q_H(s) = q_H(T).$$

Since $q_H(T) \to q_1(T) < \infty$, there exist $H_0 < 1$ and a constant $C_T < \infty$ such that

$$\sup_{H\in[H_0,1)} \sup_{s\in[0,T]} q_H(s) \le C_T.$$

Also,

$$\sup_{s\in[0,T]} q_1(s) = (1-T)^{-2}.$$

It follows that

$$|q_H(s) - q_1(s)| \le C_T + (1-T)^{-2}, \qquad s \in [0,T],$$

and the right-hand side is integrable on $[0,T]$. The dominated convergence theorem therefore yields

$$\varepsilon_H := \int_0^T |q_H(s) - q_1(s)|\,ds \longrightarrow 0.$$

Apply Theorem 1 with $\Phi(t) = 1 - t$ and $q = q_1$. Since

$$(1-t)\int_0^t q_1(s)\,ds = t,$$

the endpoint-driven path equation 48 is exactly $\widehat{X}^H$. The theorem gives equation 68, equation 69, and equation 70.

Finally,

$$\frac{d}{dt}X_t^\star = Z - X_0 = \frac{Z - X_t^\star}{1-t},$$

which proves equation 71. For uniqueness, the difference W of two solutions with the same Z and initial value satisfies $W' = -W/(1-t)$ and $W(0) = 0$. Thus $(W/(1-t))' = 0$ and $W = 0$ on $[0,1)$. The same argument applies to $\widehat{X}^H$ with Z replaced by B_1^H. □

Equivalently, for every $\delta > 0$,

$$\mathbb{P}\left(\sup_{t\in[0,T]} |X_t^H - (1-t)X_0 - tB_1^H| > \delta\right) \longrightarrow 0. \tag{72}$$

Under a common-terminal coupling satisfying $B_1^H = Z$ for every H, we also have

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^H - X_t^\star|^p\right] \longrightarrow 0.$$

C.6.1 WHY THE CANONICAL CONVERGENCE DOES NOT EXTEND TO $T = 1$

The stability theorem permits $T = 1$ when its assumptions hold on $[0, 1]$. For the canonical coefficients, however, the restriction $T < 1$ is essential; it is not merely a limitation of the estimate.

Proposition 12 (Failure of uniform convergence at the terminal time) *Let X^H and $\widehat{X}^H$ be the canonical processes, and let $N \sim \mathcal{N}(0, 1)$ and $m_p := \mathbb{E}|N|^p$. For every $H \in (1/2, 1)$ and every finite $p \geq 1$,*

$$\mathbb{E}\left[\sup_{0\leq t<1} |X_t^H - \widehat{X}_t^H|^p\right] \geq 2^{p/2} m_p.$$

Furthermore, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{0\leq t<1} |X_t^H - \widehat{X}_t^H| > \varepsilon\right) \geq \mathbb{P}(\sqrt{2}|N| > \varepsilon) > 0.$$

Thus uniform convergence in probability and in L^p fails on the full interval. For fixed $H < 1$, X^H has no continuous modification extending to $t = 1$, so the canonical weak convergence cannot be formulated in $C([0, 1])$ either.

Proof. Fix $H \in (1/2, 1)$ and set $m_H(t) := \mathrm{Cov}(S_t^H, B_1^H)$. The deterministic fBM covariance formula gives

$$m_H(t) = (1 - t)H \int_0^t q_H(u)\big(u^{2H-1} + (1 - u)^{2H-1}\big)\, du.$$

Since $q_H(u) \leq c_H(1 - u)^{-H-1}$ and each power in parentheses is at most one,

$$0 \leq m_H(t) \leq 2H(1 - t) \int_0^t q_H(u)\, du \leq 2c_H\big((1 - t)^{1-H} - (1 - t)\big) \longrightarrow 0 \quad (t \uparrow 1).$$

The initial condition cancels in $X_t^H - \widehat{X}_t^H = S_t^H - tB_1^H$. This difference is centered Gaussian, and the variance identity gives

$$\mathbb{E}|X_t^H - \widehat{X}_t^H|^2 = 2t^2 - 2t\, m_H(t) \longrightarrow 2 \quad (t \uparrow 1).$$

Its pth absolute moment therefore tends to $2^{p/2}m_p$, and its tail probability at ε tends to $\mathbb{P}(\sqrt{2}|N| > \varepsilon)$. Bounding the supremum below by the value at any fixed $t < 1$ proves the two asserted inequalities.

To exclude a continuous extension, fix $s < 1$ and set $a = (1 + s)/2$. For $t > a$, splitting the covariance integral

$$\mathrm{Cov}(S_s^H, S_t^H) = \alpha_H(1 - s)(1 - t) \int_0^s \int_0^t q_H(u)q_H(v)|u - v|^{2H-2}\, dv\, du$$

at $v = a$ gives

$$0 \leq \mathrm{Cov}(S_s^H, S_t^H) \leq C_{H,s}(1 - t)\left(1 + \int_0^t q_H(v)\, dv\right) \longrightarrow 0.$$

Indeed, on $v \leq a$ the coefficients are bounded and the kernel is integrable, while on $v > a$ the distance $v - u$ is bounded below by $a - s > 0$. Now choose $s_n \uparrow 1$ and $s_n < t_n < 1$ such that $\mathrm{Cov}(S_{s_n}^H, S_{t_n}^H) \to 0$. Then

$$\mathrm{Var}(S_{t_n}^H - S_{s_n}^H) = t_n^2 + s_n^2 - 2\,\mathrm{Cov}(S_{s_n}^H, S_{t_n}^H) \longrightarrow 2.$$

These centered Gaussian differences do not tend to zero in probability. Hence S^H has no limit in probability as $t \uparrow 1$, and therefore no continuous modification on $[0, 1]$. The same holds for X^H, since $(1 - t)X_0 \to 0$ almost surely. □

The limiting path $X_t^\star = (1 - t)X_0 + tZ$ itself does extend continuously to $X_1^\star = Z$. This does not give convergence of X^H on the full interval.

C.7 ODE LIMIT OF FRACTIONAL VARIANCE PRESERVING DYNAMICS

Let $0 \leq \beta_{\min} \leq \beta_{\max} < \infty$ and set

$$\beta(t) = \beta_{\min} + t(\beta_{\max} - \beta_{\min}), \qquad \mu(t) = -\frac{1}{2}\beta(t), \qquad g(t) = \sqrt{\beta(t)}.$$

Then

$$\Phi(t) = \exp\left(-\frac{1}{2}\beta_{\min} t - \frac{1}{4}(\beta_{\max} - \beta_{\min})t^2\right).$$

The function $q = \Phi^{-1}\sqrt{\beta}$ is continuous and of bounded variation on $[0, T]$ for every $T \in (0, 1]$. Corollary 1 gives the limiting random ODE

$$dX_t^\star = -\frac{1}{2}\beta(t)X_t^\star\, dt + \sqrt{\beta(t)}Z\, dt, \qquad X_0^\star = X_0,$$

with solution

$$X_t^\star = \Phi(t)\left(X_0 + Z\int_0^t \Phi(s)^{-1}\sqrt{\beta(s)}\, ds\right).$$

Definition 5 (Wiener integral) *Fix $H \in (1/2, 1)$ and $T \in (0, 1]$. Let $\mathcal{E}_T$ be the deterministic step functions on $[0, T]$, and define the integrand Hilbert space $\mathcal{H}_H([0, T])$ as the completion of $\mathcal{E}_T$ under the norm $\|f\|_H^2 := \langle f, f\rangle_H$, where*

$$\langle f, h\rangle_H := \alpha_H \int_0^T \int_0^T f(u)h(v)|u - v|^{2H-2}\, du\, dv, \qquad \alpha_H := H(2H - 1),$$

For $f = \sum_{i=0}^{n-1} a_i \mathbf{1}_{(t_i, t_{i+1}]} \in \mathcal{E}_T$, where $0 = t_0 < \cdots < t_n = T$, set

$$I_2^H(f) := \sum_{i=0}^{n-1} a_i (B^H_{t_{i+1}} - B^H_{t_i}).$$

Because $\mathbb{E}[I_2^H(f) I_2^H(h)] = \langle f, h\rangle_H$, this map extends uniquely to an isometry $I_2^H : \mathcal{H}_H([0, T]) \to L^2(\mathbb{P})$. We write $\int_0^T f(s)\, dB_s^H := I_2^H(f)$. In particular, $I_2^H(\mathbf{1}_{[0,t]}) = B_t^H$; indicator values at single endpoints are immaterial.

Proposition 13 (Agreement with pathwise Riemann–Stieltjes integration) *If $q \in C([0, T]) \cap BV([0, T])$ is deterministic, then for each $t \in [0, T]$,*

$$I_2^H(q\mathbf{1}_{[0,t]}) = \lim_{|\pi|\to 0} \sum_{i=0}^{n-1} q(t_i)(B^H_{t_{i+1}} - B^H_{t_i}) = q(t)B_t^H - \int_0^t B_s^H\, dq(s) = \mathcal{I}_q(B^H)(t) \quad \textit{almost surely}, \tag{73}$$

where $\pi : 0 = t_0 < \cdots < t_n = t$. Thus the pathwise process $\mathcal{I}_q(B^H)$ is a continuous version of the Wiener-integral process. In particular, this applies to the canonical q_H on every $[0, T]$ with $T < 1$.

Proof. The case $t = 0$ is immediate, so fix $t > 0$. Let $q_\pi = \sum_i q(t_i)\mathbf{1}_{(t_i, t_{i+1}]}$ on $[0, t]$, with $q_\pi(0) := q(0)$. Uniform continuity of q gives $\|q_\pi - q\|_{\infty,[0,t]} \to 0$. Because $\|\mathbf{1}_{[0,t]}\|_H^2 = t^{2H}$,

$$\|(q_\pi - q)\mathbf{1}_{[0,t]}\|_H^2 \le t^{2H} \|q_\pi - q\|^2_{\infty,[0,t]} \longrightarrow 0.$$

The defining isometry therefore shows that the sums in equation 73 converge in $L^2(\mathbb{P})$ to $I_2^H(q\mathbf{1}_{[0,t]})$.

On the other hand, discrete integration by parts and $B_0^H = 0$ give

$$\sum_{i=0}^{n-1} q(t_i)(B^H_{t_{i+1}} - B^H_{t_i}) = q(t)B_t^H - \sum_{i=1}^{n} B^H_{t_i}\big(q(t_i) - q(t_{i-1})\big).$$

Almost surely, B^H is continuous. Since q has bounded variation, the final sum converges pathwise to the Riemann–Stieltjes integral $\int_0^t B_s^H\, dq(s)$. Thus the same sums converge almost surely to $\mathcal{I}_q(B^H)(t)$. Uniqueness of limits in probability proves equation 73. The right-hand side is continuous in t, so it provides the asserted continuous version. □

D AN IMPROVED MARKOVIAN APPROXIMATION OF TYPE I FBM

Since the time reversal of an fBM-driven SDE with $H > 1/2$ does not admit the structure required to train a score-based diffusion model (Darses & Saussereau, 2007), Nobis et al. (2024) instead use the Markovian approximation of fBM (MA-fBM) (Harms & Stefanovits, 2019; Daems et al., 2024) defined below.

Definition 6 (Markov approximation of fBM (MA-fBM) (Daems et al., 2024)) *Define for $\gamma > 0$ the family of Ornstein–Uhlenbeck (OU) processes*

$$Y_\gamma(t) = Y_\gamma(0)e^{-\gamma t} + \int_0^t e^{-\gamma(t-s)}\,\mathrm{d}B(s), \tag{74}$$

with dynamics

$$\mathrm{d}Y_\gamma(t) = -\gamma Y_\gamma(t)\,\mathrm{d}t + \mathrm{d}B(t), \tag{75}$$

where all processes are driven by the same two-sided Brownian motion B. The initial condition is chosen according to the type of fBM to be approximated, where

$$Y_\gamma(0) = \begin{cases} \int_{-\infty}^0 e^{\gamma s}\,\mathrm{d}B(s), & \textit{(for Type I fBM)} \\ 0, & \textit{(for Type II fBM).} \end{cases} \tag{76}$$

Fix $K \in \mathbb{N}$, $r > 1$, set $n = \frac{K+1}{2}$, and let $\gamma_k = r^{k-n}$, $k = 1, \dots, K$, be a geometrically spaced grid of mean-reversion speeds and write $Y_k = Y_{\gamma_k}$ for all $1 \le k \le K$. Given a Hurst index $H \in (0,1)$ and a terminal time $T > 0$, we call

$$\hat{B}_H(t) := \sum_{k=1}^K \omega_k (Y_k(t) - Y_k(0)), \qquad t \in [0,T], \tag{77}$$

Markov-approximate fractional Brownian motion (MA-fBM), where the approximation coefficients $\boldsymbol{\omega} = (\omega_1, \dots, \omega_K) \in \mathbb{R}^K$ are are chosen to minimize the $L^2(\mathbb{P})$-error

$$\mathcal{E}(\boldsymbol{\omega}) := \int_0^T \mathbb{E}\left[\left(B_H(t) - \hat{B}_H(t)\right)^2\right]\mathrm{d}t \tag{78}$$

with unique minimizer obtained by solving the linear system $\boldsymbol{A}\boldsymbol{\omega} = \boldsymbol{b}$ for the matrix $\boldsymbol{A} \in \mathbb{R}^{K,K}$ depending on $\gamma_1, \dots \gamma_K$ and some vector $\boldsymbol{b} \in \mathbb{R}^K$ depending on $\gamma_1, \dots \gamma_K$ and H as given in Daems et al. (2024, Proposition 5).

The MA-fBM defined above is not well suited for approximating Type I fBM near the boundary $H \to 1$. As shown in Sec. 4, the approximation error deteriorates as H approaches 1. This is also visible in the sample paths in Sec. 4, where the fBM and MA-fBM paths are not aligned. Our goal is to adjust MA-fBM such that the limit of Type I fBM for $H \to 1$ from Prop. 7 is preserved by our approximation, i.e. for a given time horizon $T > 0$ the approximation $\tilde{B}_H(t)$ should satisfy

$$\lim_{H\to 1} \tilde{B}_H(t) = \varepsilon \frac{t}{T}, \quad \varepsilon \overset{d}{=} B_H^{(I)}(T). \tag{79}$$

The idea is to explicitly add a straight line to the the approximation:

$$\hat{B}_H^\varepsilon(t) = \sum_{k=1}^K \omega_k \left(Y_k(t) - Y_k(0)\right) + \omega_{K+1}\varepsilon\frac{t}{T}\,. \tag{80}$$

Similar to Daems et al. (2024), this permits an approximation error in quadratic form:

$$\mathcal{E}(\boldsymbol{\omega}) = \frac{\int_0^T \mathbb{E}\left[\left(\hat{B}_H(t) - B_H(t)\right)^2\right]\mathrm{d}t}{\int_0^T \mathbb{E}\left[B_H(t)^2\right]\mathrm{d}t} = \frac{\int_0^T \mathbb{E}\left[\hat{B}_H(t)^2\right]\mathrm{d}t}{\int_0^T \mathbb{E}\left[B_H(t)^2\right]\mathrm{d}t} - 2\frac{\int_0^T \mathbb{E}\left[\hat{B}_H(t)B_H(t)\right]\mathrm{d}t}{\int_0^T \mathbb{E}\left[B_H(t)^2\right]\mathrm{d}t} + 1 \tag{81}$$

$$= \sum_{i,j}\omega_i\omega_j\frac{\boldsymbol{A}_{i,j}}{c} - 2\sum_k \omega_k\frac{\boldsymbol{b}_k}{c} + 1 = \boldsymbol{\omega}^T\frac{\boldsymbol{A}}{c}\boldsymbol{\omega} - 2\frac{\boldsymbol{b}^\top}{c}\boldsymbol{\omega} + 1\,, \tag{82}$$

where $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_K, \omega_{K+1}]$. This yields a closed-form solution for the optimal $\boldsymbol{\omega}^*$:

$$\boldsymbol{A}\boldsymbol{\omega}^* = \boldsymbol{b}\,, \tag{83}$$

which results in an approximation error

$$\mathcal{E}(\boldsymbol{\omega}^*) = 1 - \frac{\boldsymbol{b}}{c}^\top \boldsymbol{\omega}^*\,. \tag{84}$$

For clarity, $\boldsymbol{A}$ and $\boldsymbol{b}$ have the following structure:

$$\boldsymbol{A} = \left(\begin{array}{c|c} \boldsymbol{A}_{i,j} & \boldsymbol{A}_{k,\varepsilon}^\top \\ \hline \boldsymbol{A}_{k,\varepsilon} & \boldsymbol{A}_{\varepsilon,\varepsilon} \end{array}\right), \boldsymbol{b} = \left(\begin{array}{c} \boldsymbol{b}_k \\ \hline \boldsymbol{b}_\varepsilon \end{array}\right). \tag{85}$$

$$\boldsymbol{A}_{i,j} = \int_0^T \frac{2 - e^{-\gamma_i t} - e^{-\gamma_j t}}{\gamma_i + \gamma_j}\,\mathrm{d}t = \frac{2T + \frac{e^{-\gamma_i T}-1}{\gamma_i} + \frac{e^{-\gamma_j T}-1}{\gamma_j}}{\gamma_i + \gamma_j}\,, \tag{86}$$

$$\boldsymbol{A}_{\varepsilon,\varepsilon} = \int_0^T \frac{t^2}{T^2}\mathbb{E}\left[\varepsilon^2\right]\mathrm{d}t = \int_0^T \frac{t^2}{T^2}\mathbb{E}\left[B_H^{(I)}(T)^2\right]\mathrm{d}t = \frac{T^{2H+1}}{3} \tag{87}$$

$$\boldsymbol{A}_{k,\varepsilon} = \int_0^T \frac{t}{T}\mathbb{E}\left[B_H^{(I)}(T)\left(Y_k(t) - Y_k(0)\right)\right]\mathrm{d}t \tag{88}$$

$$= \frac{V_H^{-1/2}}{\Gamma(H+1/2)}\int_0^T \frac{t}{T}\Bigg(\left(e^{-\gamma_k t} - 1\right)\int_{-\infty}^0 \left((T-s)^{H-1/2} - (-s)^{H-1/2}\right)e^{\gamma_k s}\,\mathrm{d}s + \int_0^t (T-s)^{H-1/2} e^{-\gamma_k(t-s)}\,\mathrm{d}s\Bigg)\mathrm{d}t \tag{89}$$

$$= V_H^{-1/2}\int_0^T \frac{t}{T}\frac{1 - e^{-\gamma_k t} - e^{\gamma_k T}Q(H+1/2, \gamma_k T) + e^{\gamma_k(T-t)}Q(H+1/2, \gamma_k(T-t))}{\gamma_k^{H+1/2}}\,\mathrm{d}t \tag{90}$$

$$= \frac{V_H^{-1/2}}{2\gamma_k^{H+5/2}T}\Bigg(-4 + \gamma_k T(\gamma_k T - 2) + (\gamma_k T)^{H+1/2}\frac{2\gamma_k T + 2H + 3}{\Gamma(H+5/2)} + 2e^{-\gamma_k T}(\gamma_k T + 1) - (\gamma_k^2 T^2 - 2)e^{\gamma_k T}Q(H+1/2, \gamma_k T)\Bigg) \tag{91}$$

$$\boldsymbol{b}_k = V_H^{-1/2}\int_0^T \frac{2 - e^{-\gamma_k t} - Q\left(H+1/2, \gamma_k t\right)e^{\gamma_k t}}{\gamma_k^{H+1/2}}\,\mathrm{d}t \tag{92}$$

$$= V_H^{-1/2}\left(\frac{2T}{\gamma_k^{H+1/2}} - \frac{T^{H+1/2}}{\gamma_k \Gamma(H+3/2)} + \frac{e^{-\gamma_k T} - Q\left(H+1/2, \gamma_k T\right)e^{\gamma_k T}}{\gamma_k^{H+3/2}}\right), \tag{93}$$

$$\boldsymbol{b}_\varepsilon = \int_0^T \frac{t}{T}\mathbb{E}\left[B_H(t)^{(I)}B_H^{(I)}(T)\right]\mathrm{d}t = \frac{1}{2}\int_0^T \frac{t}{T}(t^{2H} + T^{2H} - (T-t)^{2H})\,\mathrm{d}t \tag{94}$$

$$= \frac{2H^2 + 5H + 1}{4(2H+1)(H+1)}T^{2H+1}\,, \tag{95}$$

$$c = \int_0^T t^{2H}\,\mathrm{d}t = \frac{T^{2H+1}}{2H+1}\,. \tag{96}$$

We summarize the calcualation in the following Propostion.

Proposition 14 (Optimal approximation coefficients) *The approximation coefficients that minimize the L^2 error*

$$\mathcal{E}^\varepsilon(\boldsymbol{\omega}) := \int_0^T \mathbb{E}\left[\left(B_H(t) - \hat{B}_H^\varepsilon(t)\right)^2\right]\mathrm{d}t \tag{97}$$

are given by the closed-form expression $\boldsymbol{A}^\varepsilon\boldsymbol{\omega} = \boldsymbol{b}^\varepsilon$, with $\boldsymbol{A} = [A_{i,j}]_{1\le i\le K, 1\le j\le K}$ and $= [b_i]_{1\le K}$ as defined in Eq. (85).

To define a SDE driven by ε-MA-fBM we introduce a new Y_ε process:

$$Y_\varepsilon(0) = \frac{V_H^{-1/2}}{\Gamma(H+1/2)} \int_{-\infty}^{0} \left((T-t)^{H-1/2} - (-t)^{H-1/2} \right) \mathrm{d}W(t)\,, \tag{98}$$

$$\mathrm{d}Y_\varepsilon(t) = G_H(t)\,\mathrm{d}W(t)\,, \quad G_H(t) = \frac{V_H^{-1/2}(T-t)^{H-1/2}}{\Gamma(H+1/2)}\,, \tag{99}$$

for which the solution at time T is:

$$Y_\varepsilon(T) \overset{d}{=} B_H(T) \overset{d}{=} \varepsilon\,. \tag{100}$$

We now consider the system of SDEs

$$\mathrm{d}\hat{B}_H^{(I)}(t) = \sum_k \omega_k\,\mathrm{d}Y_k(t) + \omega_\varepsilon \frac{\varepsilon}{T}\,\mathrm{d}t \tag{101}$$

$$\mathrm{d}Y_k(t) = -\gamma_k Y_k(t)\,\mathrm{d}t + \mathrm{d}W(t), \quad k = 1, \ldots, K \tag{102}$$

$$\mathrm{d}Y_\varepsilon(t) = G_H(t)\,\mathrm{d}W(t)\,. \tag{103}$$

Define $Y = (Y_1, \ldots, Y_K, Y_\varepsilon)|Y_1(0) = y_1(0), \ldots, Y_K(0) = y_K(0), Y_\varepsilon(0) = y_\varepsilon(0))$ with dynamics

$$\mathrm{d}Y(t) = FY(t)\,\mathrm{d}t + G(t)\,\mathrm{d}W(t), \quad Y(0) = y(0) \in \mathbb{R}^{K+1}, \tag{104}$$

where

$$F = \begin{pmatrix} -\mathrm{diag}(\gamma_1, \ldots, \gamma_K) & 0 \\ 0 & 0 \end{pmatrix}, \quad G(t) = (1 \quad \ldots \quad 1 \quad G_H(t))^T\,. \tag{105}$$

Steering the process Y towards a fixed value ε in it's last entry Y_ε yields the partially pinned process $Y_{|0,1} = Y|(Y_\varepsilon(0) = y_\varepsilon(0),), Y_\varepsilon(1) = y_\varepsilon(1)$ with dynamics

$$\mathrm{d}Y_{|0,1}(t) = FY_{|0,1}(t)\,\mathrm{d}t + G(t)G(t)^T \nabla_{y_t} \log p^\varepsilon_{1|t}(y_\varepsilon(1)|Y_{|0,1}(t))\,\mathrm{d}t + G(t)\,\mathrm{d}W(t), \tag{106}$$

where $p^\varepsilon_{1|t}$ is the density of $Y_\varepsilon(1)|Y(t)$. Since

$$Y_\varepsilon(1) = Y_\varepsilon(t) + \int_t^1 G_H(s)\,\mathrm{d}W(t) \tag{107}$$

we see that $Y_\varepsilon(1)|Y(t) \sim \mathcal{N}(Y_\varepsilon(t), \sigma^2_{1|t})$ does not depend on the values of $Y_1(t), \ldots, Y_K(t)$ and

$$\sigma^2_{1|t} = \int_t^T G_H(s)^2\,\mathrm{d}s. \tag{108}$$

Hence

$$\nabla_{y_\varepsilon(t)} \log p^\varepsilon(y_\varepsilon(1)|y_\varepsilon(t)) = \frac{y_\varepsilon(1) - y_\varepsilon(t)}{\sigma^2_{1|t}} \tag{109}$$

and

$$\nabla_{y_k(t)} \log p^\varepsilon(y_\varepsilon(1)|y_\varepsilon(t)) = 0, \quad 1 \le k \le K \tag{110}$$

such that

$$G(t)G(t)^T \nabla_{y(t)} \log p^\varepsilon(y_\varepsilon(1)|y(t)) = \begin{pmatrix} \mathbf{1}_{K,K} & G_H(t)\mathbf{1}_K \\ G_H(t)\mathbf{1}_K & G_H^2(t) \end{pmatrix} \nabla_{y(t)} \log p^\varepsilon(y_\varepsilon(1)|y(t)) \tag{111}$$

$$= \begin{pmatrix} G_H(t)\mathbf{1}_K \\ G_H^2(t) \end{pmatrix} \frac{y_\varepsilon(1) - y_\varepsilon(t)}{\sigma^2_{1|t}}. \tag{112}$$

In total steering towards $Y_\varepsilon(1) = \varepsilon$ yields by the above derivations and Eq. (106) the dynamics

$$\mathrm{d}Y_{|0,1}(t) = FY_{|0,1}(t)\,\mathrm{d}t + u_\varepsilon(Y_\varepsilon(t), t)\,\mathrm{d}t + G(t)\,\mathrm{d}W(t), \quad Y_{|0,1}(0) = y(0) \tag{113}$$

$$u_\varepsilon(y, t) = (G_H(t), \ldots, G_H(t), G_H^2(t))^T \frac{\varepsilon - y_\varepsilon(t)}{\sigma^2_{1|t}}. \tag{114}$$

We get for the conditioned SDEs an extra control term $u_\varepsilon\left(Y_\varepsilon(t), t\right)$

$$\mathrm{d}\hat{B}_H^{(I)}(t) = \sum_k \omega_k \,\mathrm{d}Y_k(t) + \omega_\varepsilon \frac{\varepsilon}{T}\,\mathrm{d}t \tag{115}$$

$$\mathrm{d}Y_k(t) = -\gamma_k Y_k(t)\,\mathrm{d}t + u_\varepsilon\left(Y_\varepsilon(t), t\right)\mathrm{d}t + \mathrm{d}W(t), \quad k = 1, \ldots, K \tag{116}$$

$$\mathrm{d}Y_\varepsilon(t) = G_H(t) u_\varepsilon\left(Y_\varepsilon(t), t\right)\mathrm{d}t + G_H(t)\,\mathrm{d}W(t)\,. \tag{117}$$

As usual (Daems et al., 2025):

$$u_\varepsilon\left(Y_\varepsilon(t), t\right) = G_H(t)\partial_{Y_\varepsilon} \log p\left(Y_\varepsilon(T) = \varepsilon | Y_\varepsilon(t)\right) \tag{118}$$

where the probability density p is with respect to the unconditioned SDE. Note, this is only a function of Y_ε, not the other Y_k's. We have

$$Y_\varepsilon(T) = Y_\varepsilon(t) + \int_t^T G_H(t)\,\mathrm{d}W(t) \tag{119}$$

$$\mathbb{E}\left[Y_\varepsilon(T) | Y_\varepsilon(t)\right] = Y_\varepsilon(t) \tag{120}$$

$$v_t = \mathrm{Var}(Y_\varepsilon(T) | Y_\varepsilon(t)) \tag{121}$$

$$= \int_t^T G_H(s)^2\,\mathrm{d}s \tag{122}$$

$$= \frac{V_H^{-1}(T-t)^{2H}}{2H\Gamma(H+1/2)^2}\,. \tag{123}$$

Hence

$$\log p\left(Y_\varepsilon(T) = \varepsilon | Y_\varepsilon(t)\right) = -\frac{(\varepsilon - Y_\varepsilon(t))^2}{2v_t} + \text{const} \tag{124}$$

and

$$u_\varepsilon\left(Y_\varepsilon(t), t\right) = G_H(t)\frac{\varepsilon - Y_\varepsilon(t)}{v_t}\,. \tag{125}$$

$Y_\varepsilon(0)$ can be sampled together with the usual $Y_1(0), \ldots, Y_K(0)$ from their joint Gaussian distributions using

$$\mathrm{Var}(\varepsilon) = T^{2H} \tag{126}$$

$$\mathrm{Var}(Y_\varepsilon(0)) = \mathrm{Var}(\varepsilon) - v_0 \tag{127}$$

$$= T^{2H} - v_0 \tag{128}$$

$$\mathrm{Cov}(Y_\varepsilon(0), Y_k(0)) = \frac{V_H^{-1/2}}{\Gamma(H+1/2)}\mathbb{E}\left[\int_{-\infty}^0 e^{\gamma_k s}\,\mathrm{d}W(s)\left(\int_{-\infty}^0 (T-s)^{H-1/2}\,\mathrm{d}W(s) - \int_{-\infty}^0 (-s)^{H-1/2}\,\mathrm{d}W(s)\right)\right] \tag{129}$$

$$= \frac{V_H^{-1/2}}{\Gamma(H+1/2)}\int_{-\infty}^0 e^{\gamma_k s}\left((T-s)^{H-1/2} - (-s)^{H-1/2}\right)\mathrm{d}s \tag{130}$$

$$= V_H^{-1/2}\frac{e^{\gamma_k T}Q(H+1/2, \gamma_k T) - 1}{\gamma_k^{H+1/2}}\,. \tag{131}$$

This ultimately leads to an adapted augmented system SDEs driven by ε-MA-fBM, where we add ε to the state, which now becomes the augmented state $Z(t) = [X(t), Y_1(t), \ldots, Y_K(t), Y_\varepsilon(t), \varepsilon]$ with dynamics

$$\mathrm{d}Z(t) = F(t, H)Z(t)\,\mathrm{d}t + G(t, H)\,\mathrm{d}W(t) \tag{132}$$

with

$$F(t,H) = \begin{pmatrix} -\theta_t & -\sigma_t\omega_1\gamma_1 & \cdots & -\sigma_t\omega_K\gamma_K & -\sigma_t\bar{\omega}G_H(t)/v_t & \theta_t\mu_t + \sigma_t\left(\omega_\varepsilon/T + \bar{\omega}G_H(t)/v_t\right) \\ 0 & -\gamma_1 & & & -G_H(t)/v_t & G_H(t)/v_t \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & & & -\gamma_K & -G_H(t)/v_t & G_H(t)/v_t \\ 0 & 0 & \cdots & 0 & -G_H^2(t)/v_t & G_H^2(t)/v_t \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}, \tag{133}$$

$$G(t,H) = \begin{pmatrix} \sigma_t\bar{\omega} \\ 1 \\ \vdots \\ 1 \\ G_H(t) \\ 0 \end{pmatrix}. \tag{134}$$

Now, we can simply use the ODE formulations to numerically approximate the covariance:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = F(t,H)m \tag{135}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = F(t,H)P + PF^\top(t,H) + G(t,H)\bar{G}^\top(t,H) \tag{136}$$

where $m(0) = [x_0, 0, \ldots, 0]^\top$, and the extra terms in $P(0)$, are given by

$$\mathrm{Var}(\varepsilon) = T^{2H}, \tag{137}$$

$$\mathrm{Cov}(\varepsilon, X(0)) = 0, \tag{138}$$

$$\mathrm{Cov}(\varepsilon, Y_\varepsilon(0)) = \mathrm{Var}(Y_\varepsilon(0)), \tag{139}$$

$$\mathrm{Cov}(\varepsilon, Y_k(0)) = \mathrm{Cov}(Y_\varepsilon(0), Y_k(0)). \tag{140}$$

E Derivation of the KL divergence

The proof follows the arguments of Blessing et al. (2025, Appendix A.4–A.5), with their degenerate diffusion coefficient $\eta(t) = (0, \sigma(t))^\mathsf{T}$ replaced with an arbitrary $\eta \in C([0,T], \mathbb{R}^{D\times d})$ that has rank d for every $t \in [0,T]$.

Lemma 2 *Fix a data dimension $d \in \mathbb{N}$ and an augmented state dimension $D \in \mathbb{N}$ with $D > d$. Let B be a d-dimensional Brownian motion, and let $f \in C(\mathbb{R}^D \times [0,T], \mathbb{R}^D)$ and $\eta \in C([0,T], \mathbb{R}^{D\times d})$ such that $\eta(t)$ has rank d for every $t \in [0,T]$. Denote by $\mathbb{P}_Z$ the path measure associated with the process Z solving*

$$\mathrm{d}Z(t) = f(Z(t),t)\,\mathrm{d}t + \eta(t)\,\mathrm{d}B(t), \qquad Z(0) \sim p_Z(\cdot, 0). \tag{141}$$

For any time-dependent function φ, we denote its time reversal by $\bar{\varphi}(\cdot,t) := \varphi(\cdot, T-t)$. For a control $u \in C(\mathbb{R}^D \times [0,T], \mathbb{R}^d)$ let $\mathbb{P}^u$ denote the path measure associated with the parameterized reverse-time process R_u solving

$$\mathrm{d}R_u(t) = \big(\bar{f} + \bar{\eta}\bar{u}\big)(R_u(t),t)\,\mathrm{d}t + \bar{\eta}(t)\,\mathrm{d}\bar{B}(t), \qquad R_u(0) \sim \pi. \tag{142}$$

With

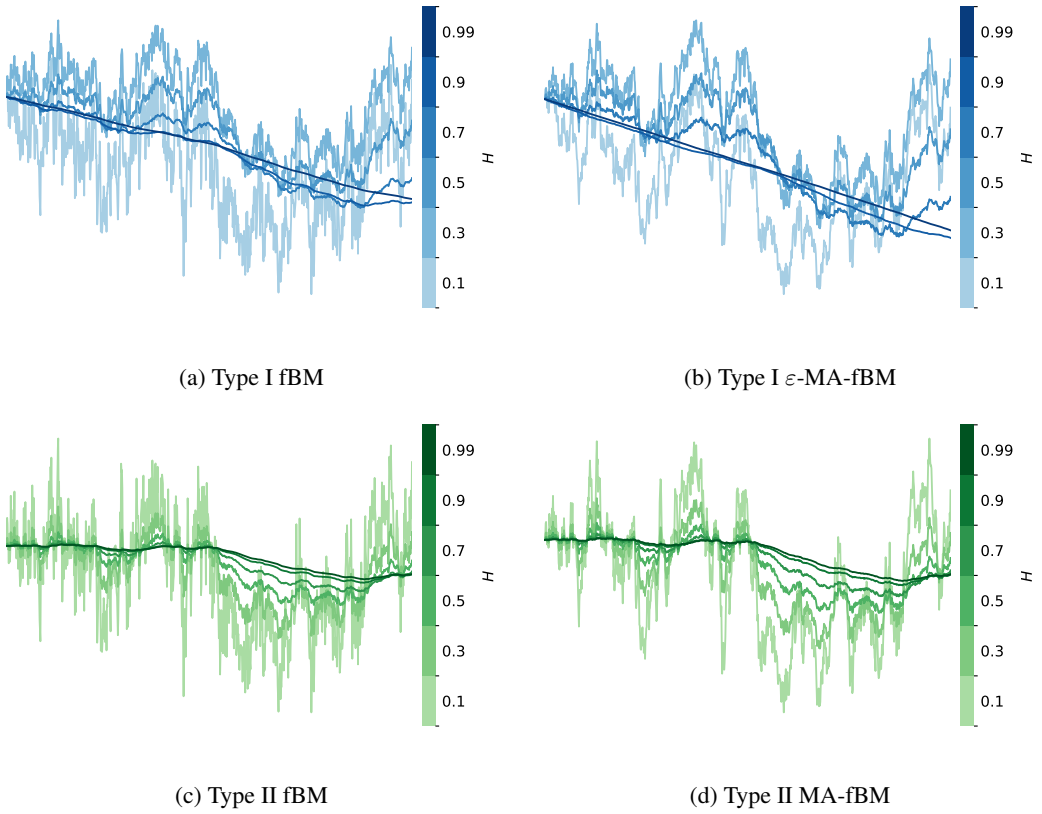
$$u^\star(z,t) = \eta^\mathsf{T}(t)\nabla_z \log p_Z(z,t), \tag{143}$$

the process

$$\mathrm{d}R^\star(t) = \big(-\bar{f} + \bar{\eta}\bar{u}^\star\big)(R^\star(t),t)\,\mathrm{d}t + \bar{\eta}(t)\,\mathrm{d}\bar{B}(t), \qquad R^\star(0) \sim p_Z(\cdot, T), \tag{144}$$

is the time-reversal of Z and denoting its path measure by $\mathbb{P}^{u^\star}$ yields ()

$$D_{\mathrm{KL}}\left(\mathbb{P}^{u^\star} \| \mathbb{P}^u\right) = D_{\mathrm{KL}}\left(p_Z(\cdot,T) \| \pi\right) + \frac{1}{2}\int_0^T \mathbb{E}_{Z(t)}\left[\|u^\star(Z(t),t) - u(Z(t),t)\|_2^2\right]\mathrm{d}t. \tag{145}$$



(a) Type I fBM (b) Type I ε-MA-fBM

(c) Type II fBM (d) Type II MA-fBM

Figure 3: Exact fBM vs. its Markov approximation ($K = 5$) across H. All four panels share the same seed for randomness.

Proof. Since (Haussmann & Pardoux, 1986) holds for degenerate η, the time-reversal $R^\star$ of Z solves

$$\mathrm{d}R^\star(t) = \left[-f(R^\star(t), T-t) + \eta(T-t)\eta^\mathsf{T}(T-t)\nabla_z \log p_Z(R^\star(t), T-t)\right]\mathrm{d}t + \eta(T-t)\,\mathrm{d}\bar{B}(t) \tag{146}$$

and hence Eq. (144) holds true. Now for the second claim, let $\mathbb{P}_z^{u^\star}$ and $\mathbb{P}_z^u$ denote the conditional path measures conditioned on their initial value z. Disintegrating the path measures with respect to the evaluation map $R \mapsto R(0)$ and applying the additive property of relative entropy (**?**, Theorem 2.4) gives

$$\begin{aligned} D_{\mathrm{KL}}\left(\mathbb{P}^{u^\star} \| \mathbb{P}^u\right) &= D_{\mathrm{KL}}\left(p_Z(\cdot, T) \| \pi\right) \\ &\quad + \mathbb{E}_{z\sim p_Z(\cdot,T)}\left[D_{\mathrm{KL}}\left(\mathbb{P}_z^{u^\star} \| \mathbb{P}_z^u\right)\right]. \end{aligned} \tag{147}$$

It remains to evaluate the conditional path-space KL divergence. For fixed initial value z, the two processes have the same diffusion coefficient $\bar{\eta}$, while their drift difference is

$$\bar{\eta}(t)\left[u^\star(\cdot, T-t) - u(\cdot, t)\right]. \tag{148}$$

Since $\eta(t) \in \mathbb{R}^{D\times d}$ has rank d, it has full column rank. Hence its Moore–Penrose pseudoinverse is given by

$$\eta^+(t) = \left(\eta^\mathsf{T}(t)\eta(t)\right)^{-1}\eta^\mathsf{T}(t), \tag{149}$$

and satisfies

$$\eta^+(t)\eta(t) = I_d. \tag{150}$$

The same therefore holds for $\bar{\eta}(t)$.

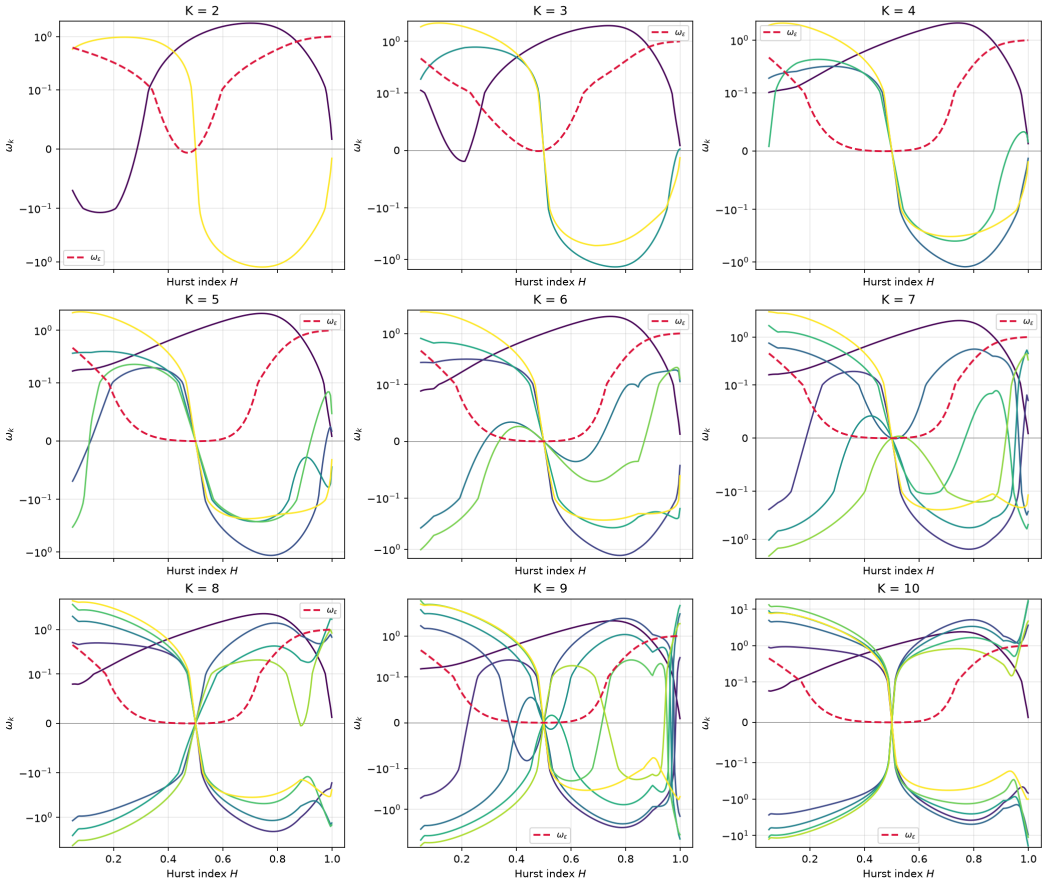


Figure 4: Evolution of the optimal approximation weights ω over the Hurst index H for an increasing number of auxiliary processes K. Lighter colour indicate a higher index of the corresponding weight. At $H = 0.5$, ω_0 equals 1 and $\omega_i, i! = 0$ is 0. We therefore recover Brownian motion in our approximation.

Applying Girsanov's theorem to the d-dimensional driving Brownian motion (Karatzas & Shreve, 2004, Theorem 3.5.1) and using Eq. (150) gives

$$D_{\mathrm{KL}}\left(\mathbb{P}_z^{u^\star}\|\mathbb{P}_z^{u}\right) = \frac{1}{2}\mathbb{E}_{\mathbb{P}_z^{u^\star}}\left[\int_0^T \left\|\bar{\eta}^+(t)\bar{\eta}(t)\left(u^\star(R^\star(t),T-t)-u(R^\star(t),t)\right)\right\|_2^2 \mathrm{d}t\right]$$
$$= \frac{1}{2}\mathbb{E}_{\mathbb{P}_z^{u^\star}}\left[\int_0^T \left\|u^\star(R^\star(t),T-t)-u(R^\star(t),t)\right\|_2^2 \mathrm{d}t\right]. \tag{151}$$

Equivalently, this is the full-column-rank analogue of the degenerate Girsanov calculation in Blessing et al. (2025, Theorem A.4); the particular block structure of their diffusion coefficient is not required here because Eq. (150) follows directly from the full-column-rank assumption.

Substituting the preceding identity into Eq. (147) and integrating over the initial distribution of $R^\star$ yields

$$D_{\mathrm{KL}}\left(\mathbb{P}^{u^\star}\|\mathbb{P}^{u}\right) = D_{\mathrm{KL}}\left(p_Z(\cdot,T)\|\pi\right)$$
$$+\frac{1}{2}\int_0^T \mathbb{E}_{R^\star(t)}\left[\left\|u^\star(R^\star(t),T-t)-u(R^\star(t),t)\right\|_2^2\right]\mathrm{d}t. \tag{152}$$

Finally, by the change of variables $s = T-t$, we obtain

$$D_{\mathrm{KL}}\left(\mathbb{P}^{u^\star}\|\mathbb{P}^{u}\right) = D_{\mathrm{KL}}\left(p_Z(\cdot,T)\|\pi\right)$$
$$+\frac{1}{2}\int_0^T \mathbb{E}_{Z(s)}\left[\left\|u^\star(Z(s),s)-u(Z(s),T-s)\right\|_2^2\right]\mathrm{d}t, \tag{153}$$

which proves the claim. □

Corollary 2 *Recall that $Z_u = (X_u, Y_u)$ is the parameterized process with dynamics Eq. (141) with marginal density $p_{Z_u}(\cdot,t)$, $p_{X_u}(\cdot,t)$ is the the density of $X_u(t)$ and $p_X(x,0)$ denotes the unknown data distribution.*

$$\mathbb{E}_{X(0)}\left[\log p_{X_u}(X(0),T)\right] \geq -\mathcal{H}(p_X(\cdot,0)) - D_{KL}(p_{Z_T}\|\pi) - \frac{1}{2}\int_0^T \mathbb{E}_{Z(t)}\left[\|\eta(t)^\mathsf{T}\nabla_z \log p(Z(t),t) - u(Z(t),T-t)\|_2^2\right]\mathrm{d}t. \tag{154}$$

Proof. Recall that for any measurable map $\phi : \mathbb{C}([0,T],\mathbb{R}^D) \to \mathbb{C}([0,T],\mathbb{R}^d)$ and path measures $\mathbb{P}$ and $\mathbb{Q}$ with associated processes taking values in $\mathbb{R}^D$ we have by the data processing inequality

$$D_{KL}(\Phi_\#\mathbb{P}\|\Phi_\#\mathbb{Q}) \leq D_{KL}(\mathbb{P}\|\mathbb{Q}) \tag{155}$$

for the respective pushforward under ϕ. Therefore by defining projections on the first coordinate we have

$$D_{KL}(\mathbb{P}_{\overline{X}(T)}\|\mathbb{P}_{X_u(T)}) \leq D_{KL}(\mathbb{P}_{\overline{X}}\|\mathbb{P}_{X_u}) \leq D_{KL}(\mathbb{P}_{\overline{Z}}\|\mathbb{P}_{Z_u}). \tag{156}$$

By definition of the reverse process $\overline{X}(T) = X(0) \sim p_X(\cdot,0)$ we thus find

$$D_{KL}\left(p_X(\cdot,0)\|p_{X_u}(\cdot,T)\right) = D_{KL}\left(\mathbb{P}_{\overline{X}(T)}\|\mathbb{P}_{X_u(T)}\right) \leq D_{KL}(\mathbb{P}_{\overline{Z}}\|\mathbb{P}_{Z_u}). \tag{157}$$

On the other hand we find

$$D_{KL}\left(p_X(\cdot,0)\|p_{X_u}(\cdot,T)\right) = \mathbb{E}_{X(0)}\left[\log\frac{p_X(X(0),0)}{p_{X_u}(X(0),T)}\right] \tag{158}$$
$$= \mathbb{E}_{X(0)}\left[\log p_X(X(0),0)\right] - \mathbb{E}_{X(0)}\left[\log p_{X_u}(X(0),T)\right] \tag{159}$$
$$= -\mathcal{H}(p_X(\cdot,0)) - \mathbb{E}_{X(0)}\left[\log p_{X_u}(X(0),T)\right] \tag{160}$$

and hence

$$\mathbb{E}_{X(0)}\left[\log p_{X_u}(X(0),T)\right] \geq -\mathcal{H}(p_X(\cdot,0)) - D_{KL}(\mathbb{P}_{\overline{Z}}\|\mathbb{P}_{Z_u}) \tag{161}$$

yields the result. □

The marginal mean $m(t) = \mathbb{E}[Z_t|X_0]$ and the marginal covariance $\Sigma(t) = \mathbb{V}[Z_t|X_0]$ of the process $Z = (X, Y_1, \ldots, Y_K, Y_\varepsilon, \varepsilon)$ as defined in **??** can be approximated numerically by solving the system of ODEs (Särkkä & Solin, 2019, Eq. (6.2))

$$\frac{\mathrm{d}m}{\mathrm{d}t} = F(t)m \tag{162}$$

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = F(t)P + PF^\top(t) + G(t)G^\top(t) \tag{163}$$

where $m(0) = [x_0, 0, \ldots, 0]^\top$, and the entries of $\Sigma(0)$ are

$$\mathrm{Cov}(X(0), Y_k(0)) = 0 \tag{164}$$

$$\mathrm{Cov}(X(0), Y_\varepsilon(0)) = 0 \tag{165}$$

$$\mathrm{Cov}(X(0), \varepsilon) = 0 \tag{166}$$

$$\mathrm{Cov}(\varepsilon, X(0)) = 0\,, \tag{167}$$

$$\mathrm{Cov}(Y_i(0), Y_j(0)) = \mathbb{E}\left[\int_{-\infty}^0 e^{\gamma_i s}\,\mathrm{d}B(t) \int_{-\infty}^0 e^{\gamma_j s}\,\mathrm{d}B(t)\right] = \int_{-\infty}^0 e^{(\gamma_i+\gamma_j)s}\,\mathrm{d}t = \frac{1}{\gamma_i + \gamma_j} \tag{168}$$

$$\mathrm{Var}(\varepsilon) = \mathrm{Var}(B_H(T)) = T^{2H}\,, \tag{169}$$

$$\mathrm{Cov}(\varepsilon, Y_\varepsilon(0)) = Cov(Y_\varepsilon(T), Y_\varepsilon(0)) = Cov(Y_\varepsilon(0) + \int_0^T G_H(t)\,\mathrm{d}B(t)), Y_\varepsilon(0)) \tag{170}$$

$$= \mathrm{Var}(Y_\varepsilon(0))\,, \tag{171}$$

$$\mathrm{Cov}(\varepsilon, Y_k(0)) = \mathrm{Cov}(Y_\varepsilon(0), Y_k(0))\,. \tag{172}$$

and

$$\mathrm{Cov}(Y_\varepsilon(0), Y_k(0)) = \frac{V_H^{-1/2}}{\Gamma(H+1/2)}\mathbb{E}\left[\int_{-\infty}^0 e^{\gamma_k s}\,\mathrm{d}W(s)\left(\int_{-\infty}^0 (T-s)^{H-1/2}\,\mathrm{d}W(s) - \int_{-\infty}^0 (-s)^{H-1/2}\,\mathrm{d}W(s)\right)\right] \tag{173}$$

$$= \frac{V_H^{-1/2}}{\Gamma(H+1/2)}\int_{-\infty}^0 e^{\gamma_k s}\left((T-s)^{H-1/2} - (-s)^{H-1/2}\right)\mathrm{d}s \tag{174}$$

$$= V_H^{-1/2}\frac{e^{\gamma_k T}Q(H+1/2, \gamma_k T) - 1}{\gamma_k^{H+1/2}}\,. \tag{175}$$

F SIMPLIFIED AUGMENTED SCORE MATCHING

We write

$$U(t) := X(t) - \sum_{k=0}^{K+1} \alpha_t^k Y_k(t) \tag{176}$$

for the process of score-model input and with a slight abuse of notation we write $u = u(x, y)$ where $x \in \mathbb{R}^d$, $y \in \mathbb{R}^{d(K+2)}$ and $u(x, y) = x - \sum_{k=0}^{K+2} \alpha_t^k y_k$ for $z = (x, y)$ with $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^{d(K+2)}$. We first note

$$U(t, x_0) := (X(t) - \sum_{k=0}^{K+3} \alpha_t^k Y^k(t))|(X(0) = x_0) \sim \mathcal{N}(c(t)x_0, \zeta_t^2 I_d) \tag{177}$$

where $c(t) = \exp(\int_0^t \mu(s)\,\mathrm{d}s)$ is the mean scale, ζ_t^2 is the conditional variance of $X(t)|(Y(t), X_0 = x_0)$ and $x_0 \in \mathbb{R}^d$ a fixed data point. Hence, we have in distribution

$$U(t, x_0) \overset{d}{=} c(t)x_0 + \zeta_t \xi, \quad \xi \sim \mathcal{N}(0, I_d). \tag{178}$$

Define the augmented score-model $S_\theta : \mathbb{R}^{d(K+3)} \times [0, 1] \to \mathbb{R}^{d(K+3)}$ via

$$(x, y, t) = (z, t) \mapsto A(t)s_\theta(u, t) + (0, \nabla_y \log q_t(y)), \tag{179}$$

where s_θ is a time-dependent neural network, q_t the known density of Y_t and $A(t) \in \mathbb{R}^{d(K+3),d}$ with

$$A(t) = \begin{pmatrix} \alpha(t) & 0_{K+3} & \cdots & 0_{K+3} \\ 0_{K+3} & \alpha(t) & \cdots & 0_{K+3} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{K+3} & 0_{K+3} & \cdots & \alpha(t) \end{pmatrix} \in \mathbb{R}^{d(K+3)\times d}, \quad \alpha(t) = (\alpha_t^0, -\alpha_t^1, \ldots, -\alpha_t^{K+2})^\mathsf{T} \in \mathbb{R}^{K+3} \tag{180}$$

with $\alpha_t^0 \equiv 1$ and

$$\alpha_t^k = \sum_{k=1}^{K+2} \mathbb{E}[X(t)Y_k(t)] \left[(\Sigma_t^y)^{-1}\right]_{l,k}, \quad 1 \le k \le K+2. \tag{181}$$

Define

$$s^\star(z,t) = \mathbb{E}_{X_0|Z_t}\left[\nabla_x \log p_{0t}(X_t|Y_t, X_0)\right] \tag{182}$$

such that we have by

$$\nabla_z \log p_t^a(z) - S_\theta(z,t) = A(t)\left(s^\star(z,t) - s_\theta(u,t)\right), \tag{183}$$

since the $(0, \nabla_y q_t(z))$ on the RHS of the augmented score-model parameterization in Eq. (179) cancels by . Setting

$$\Lambda(t) = G(t,H)^\mathsf{T} A(t) = \lambda(t) I_d \tag{184}$$

where $\lambda : [0,T] \to \mathbb{R}$ is a scalar function. Hence, we have

$$\mathbb{E}_{Z_t}\left[\|G(t,H)^\mathsf{T}\left(\nabla_z \log p_t^a(Z_t) - S_\theta(Z_t,t)\right)\|^2\right] \tag{185}$$

$$= \mathbb{E}_{Z_t}\left[\lambda^2(t)\|s^\star(Z_t,t) - s_\theta(U_t,t)\|^2\right] \tag{186}$$

$$= \mathbb{E}_{Z_t}\left[\lambda^2(t)\|\mathbb{E}_{X_0|Z_t}\left[\nabla_x \log p_{0t}(X_t|Y_t,X_0)\right] - s_\theta(U_t,t)\|^2\right] \tag{187}$$

$$= \mathbb{E}_{Z_t}\left[\lambda^2(t)\|\mathbb{E}_{X_0|Z_t}\left[\frac{c(t)X_0 - U(t)}{\zeta_t^2}\right] - s_\theta(U_t,t)\|^2\right] \tag{188}$$

$$= \mathbb{E}_{Z_t}\left[\lambda^2(t)\|\mathbb{E}\left[\frac{c(t)X_0 - U(t)}{\zeta_t^2}|\sigma(Z_t)\right] - s_\theta(U_t,t)\|^2\right] \tag{189}$$

$$= \mathbb{E}_{U_t}\left[\lambda^2(t)\|\mathbb{E}\left[\frac{c(t)X_0 - U(t)}{\zeta_t^2}|\sigma(U_t)\right] - s_\theta(U_t,t)\|^2\right] \tag{190}$$

$$= \lambda(t)\mathbb{E}_{U_t}\left[\|\mathbb{E}\left[\frac{c(t)X_0 - U(t)}{\zeta_t^2} - s_\theta(U_t,t)|\sigma(U_t)\right]\|^2\right] \tag{191}$$

$$\overset{\text{Pythagoras in L2}}{=} \lambda^2(t)\mathbb{E}_{U_t}\left[\left\|\frac{c(t)X_0 - U(t)}{\zeta_t^2} - s_\theta(U(t),t)\right\|^2\right] \tag{192}$$

$$- \lambda^2(t)\mathbb{E}_{U_t}\left[\|\frac{c(t)X_0 - U(t)}{\zeta_t^2} - s_\theta(U_t,t) - \mathbb{E}\left[\frac{c(t)X_0 - U(t)}{\zeta_t^2} - s_\theta(U_t,t)|\sigma(U_t)\right]\|^2\right] \tag{193}$$

$$= \lambda^2(t)\mathbb{E}_{U_t}\left[\left\|\frac{c(t)X_0 - U(t)}{\zeta_t^2} - s_\theta(U(t),t)\right\|^2\right] - \underbrace{\lambda^2(t)\mathbb{E}_{U_t}\left[\|\frac{c(t)X_0 - U(t)}{\zeta_t^2} - \mathbb{E}\left[\frac{c(t)X_0 - U(t)}{\zeta_t^2}|\sigma(U_t)\right]\|^2\right]}_{:=C(t) \text{ does not depend on } \theta} \tag{194}$$

$$\overset{(178)\text{does this not only hold conditioned on } x_0?}{=} \frac{\lambda^2(t)}{\zeta_t^2}\mathbb{E}_{U_t}\left[\left\|\xi - \underbrace{\zeta_t s_\theta(c(t)X_0 + \zeta_t\xi, t)}_{:=\xi_\theta(U_t,t)}\right\|^2\right] - C(t) \tag{195}$$

Now define

$$x_\theta(u,t) := \frac{u - \zeta_t \xi_\theta(u,t)}{c(t)} \tag{196}$$

such that for a fixed ξ and x_0 we have with $U(t) = c(t)x_0 + \zeta_t\xi$

$$\xi - \xi_\theta(U(t),t) = \xi - \frac{U(t) - c(t)\frac{U(t)-\zeta_t\xi_\theta(u,t)}{c(t)}}{\zeta_t} = \frac{U(t) - c(t)x_\theta(u,t)}{\zeta_t} \tag{197}$$

$$= \xi - \frac{c(t)X_0 + \zeta_t\xi - c(t)x_\theta(u,t)}{\zeta_t} = \frac{c(t)}{\zeta(t)}(x_\theta(u,t) - x_0) \tag{198}$$

Hence, we have in total with $\pi = \mathcal{N}(0_{d(K+1)}, \Sigma_T)$

$$D_{KL}(\mathbb{P}\|\mathbb{P}_\theta) = D_{KL}(p_T^a\|\pi) + \frac{1}{2}\int_0^T \mathbb{E}\left[\|G(t,H)^\mathsf{T}\left(\nabla_z \log p_t^a(Z_t) - S_\theta(Z_t,t)\right)\|\right] \mathrm{d}t \tag{199}$$

$$= D_{KL}(p_T^a\|\pi) + \frac{1}{2}\int_0^T \mathbb{E}\left[\frac{\lambda^2(t)}{\zeta_t^2}\|\xi - \xi_\theta(U(t),t)\|\right] \mathrm{d}t \tag{200}$$

$$= D_{KL}(p_T^a\|\pi) + \frac{1}{2}\int_0^T \mathbb{E}\left[\frac{\lambda^2(t)c^2(t)}{\zeta_t^4}\|\left(x_\theta(U(t),t) - x_0\right)\|\right] \mathrm{d}t \tag{201}$$

where we call $Eq.$ (200) "noise prediction" and $Eq.$ (201) "signal prediction". We introduce an additional weighting function $w : [0,T] \to (0,\infty)$ and define our loss via

$$\mathcal{L}(\theta) = D_{KL}(p_T^a\|\pi) + \frac{1}{2}\int_0^T \mathbb{E}\left[\frac{\lambda^2(t)w(t)}{\zeta_t^2}\|\xi - \xi_\theta(U(t),t)\|\right] \mathrm{d}t \quad \textit{(noise prediction)} \tag{202}$$

$$= D_{KL}(p_T^a\|\pi) + \frac{1}{2}\int_0^T \mathbb{E}\left[\frac{\lambda^2(t)c^2(t)w(t)}{\zeta_t^4}\|x_\theta(U(t),t) - x_0\|\right] \mathrm{d}t \quad \textit{(signal prediction)} \tag{203}$$

F.1 KL-PRIOR

First note

$$D_{KL}(p_T^a\|\pi) = \int_{-\infty}^{\infty} p_T^a(z)\log\left(\frac{p^a(z)}{\pi(z)}\right)\mathrm{d}z \tag{204}$$

$$= \int_{-\infty}^{\infty} q_T(y)\int_{-\infty}^{\infty} p(x|y)\log\left(\frac{p(x|y)q_T(y)}{\pi(x|y)q_T(y)}\right)\mathrm{d}x\,\mathrm{d}y \tag{205}$$

$$= \int_{-\infty}^{\infty} q_T(y)\int_{-\infty}^{\infty} p(x|y)\log\left(\frac{p(x|y)}{\pi(x|y)}\right)\mathrm{d}x\,\mathrm{d}y \tag{206}$$

$$= \int_{-\infty}^{\infty} q_T(y) D_{KL}(p_{X_T|Y_T}\|\pi_{X|Y})\,\mathrm{d}y \tag{207}$$

$$= \mathbb{E}_{Y\sim q_T(y)}\left[D_{KL}(p_{X_T|Y_T}\|\pi_{X|Y})\right] \tag{208}$$

and

$$p_{X_T|Y_T}(x|y) = \int_{-\infty}^{\infty} p_{X_T,X_0|Y_T}(x,x_0|y)\,\mathrm{d}x_0 \tag{209}$$

$$= \int_{-\infty}^{\infty} p_{X_T,X_0|Y_T}(x,|y,x_0)p_{X_0|Y_T}(x_0|y)\,\mathrm{d}x_0 \tag{210}$$

$$= \int_{-\infty}^{\infty} p_{X_T|Y_T,X_0}(x,|y,x_0)p_{X_0}(x_0)\,\mathrm{d}x_0 \tag{211}$$

$$= \mathbb{E}_{X_0\sim\mathcal{D}}[p_{X_T|Y_T,X_0}(x,|y,X_0)] \tag{212}$$

$$= \mathbb{E}_{X_0\sim\mathcal{D}}[\mathcal{N}(c(t)X_0,\zeta_t^2 I_d)] \tag{213}$$

where we use $X_0 \perp Y_T$. Therefore

$$D_{KL}(p_T^a\|\pi) = \mathbb{E}_{Y\sim q_T(y)}\left[D_{KL}(p_{X_T|Y_T}\|\pi_{X|Y})\right] \tag{214}$$

$$= \mathbb{E}_{Y\sim q_T(y)}\left[D_{KL}(\mathbb{E}_{X_0\sim\mathcal{D}}[\mathcal{N}(c(t)X_0,\zeta_t^2 I_d)]\|\mathcal{N}(0,\zeta_t^2 I_d)\right] \tag{215}$$

$$\leq \mathbb{E}_{Y\sim q_T(y)}\left[\mathbb{E}_{X_0\sim\mathcal{D}}[D_{KL}(\mathcal{N}(c(t)X_0,\zeta_t^2 I_d)\|\mathcal{N}(0,\zeta_t^2 I_d)]\right] \tag{216}$$

and since

$$D_{KL}(\mathcal{N}(c(t)x_0, \zeta_t^2 I_d)\|\mathcal{N}(0, \zeta_t^2 I_d)] = \frac{1}{2}\left[\mathrm{tr}\left(\frac{\zeta_T^1}{\zeta_T^1}\right) - d + (c(t)x_0^\mathsf{T})\frac{1}{\zeta_T^2}c(t)x_0 + \ln\frac{\det\zeta_T^2}{\det\zeta_T^2}\right] \tag{217}$$

$$= \frac{1}{2}\left[1 - d + \frac{c^2(t)}{\zeta_T^2}x_0^\mathsf{T}x_0\right] \tag{218}$$

$$= \frac{1-d}{2} + \frac{c^2(t)}{2\zeta_T^2}x_0^\mathsf{T}x_0 \tag{219}$$

$$\leq \frac{c^2(t)}{2\zeta_T^2}x_0^\mathsf{T}x_0 \tag{220}$$

we have in total for each dimension

$$D_{KL}(p_T^a\|\pi) \leq \frac{c^2(t)x_{0,j}^2}{2\zeta_T^2} =: \mathcal{L}_{prior,j}(H), \quad 1 \leq j \leq d. \tag{221}$$

and we optimize (for noise prediction)

$$\mathcal{L}^{eff}(\theta) = \frac{2}{(T_{ode} - \epsilon_{train})}\mathcal{L}(\theta) \tag{222}$$

$$\leq \frac{2}{(T_{ode} - \epsilon_{train})}\left\{\mathcal{L}_{prior}(H) + \frac{1}{2}\int_0^T \mathbb{E}\left[\frac{\lambda^2(t)w(t)}{\zeta_t^2}\|\xi - \xi_\theta(U(t), t)\|\right]\mathrm{d}t\right\} \tag{223}$$

$$= \frac{c^2(t)\mathbb{E}[X_0^\mathsf{T}X_0]}{(T_{ode} - \epsilon_{train})\zeta_T^2} + \frac{1}{(T_{ods} - \epsilon_{train})}\int_0^T \mathbb{E}\left[\frac{\lambda^2(t)w(t)}{\zeta_t^2}\|\xi - \xi_\theta(U(t), t)\|\right]\mathrm{d}t \tag{224}$$

$$\approx \frac{c^2(t)\mathbb{E}[X_0^\mathsf{T}X_0]}{(T_{ode} - \epsilon_{train})\zeta_T^2} + \frac{1}{(T_{ods} - \epsilon_{train})}\int_{\epsilon_{train}}^{T_{ode}} \mathbb{E}\left[\frac{\lambda^2(t)w(t)}{\zeta_t^2}\|\xi - \xi_\theta(U(t), t)\|\right]\mathrm{d}t \tag{225}$$

$$\approx \frac{c^2(t)\mathbb{E}[X_0^\mathsf{T}X_0]}{(T_{ode} - \epsilon_{train})\zeta_T^2} + \mathbb{E}_{t\sim\mathcal{U}[\epsilon_{train}, T_{ode}]}\left[\mathbb{E}\left[\frac{\lambda^2(t)w(t)}{\zeta_t^2}\|\xi - \xi_\theta(U(t), t)\|\right]\right]\mathrm{d}t \tag{226}$$

F.2 EDM

Following Karras et al. (2022) we define a fixed interpolation between noise and signal prediction. Towards that goal, we first define

$$D_\theta(u, t) := c_{skip}(t)u + c_{out}(t)F_\theta(c_{in}(t)u, t) \tag{227}$$

such that for $w : [0, T] \to (0, \infty)$ in the signal prediction setting we have

$$\mathbb{E}\left[\frac{\lambda^2(t)c^2(t)w(t)}{\zeta_t^4}\|x_\theta(U(t), t) - X(0)\|\right] \tag{228}$$

$$= \mathbb{E}\left[\frac{\lambda^2(t)c^2(t)w(t)}{\zeta_t^4}\|D_\theta(U(t), t) - X(0)\|\right] \tag{229}$$

$$= \mathbb{E}\left[\frac{\lambda^2(t)c^2(t)w(t)}{\zeta_t^4}\|c_{skip}(t)U(t) + c_{out}(t)F_\theta(c_{in}(t)U(t), t) - X(0)\|\right] \tag{230}$$

$$= \mathbb{E}\left[\frac{\lambda^2(t)c^2(t)w(t)c_{out}^2(t)}{\zeta_t^4}\|F_\theta(c_{in}(t)U(t), t) - \underbrace{\frac{1}{c_{out}(t)}\left[X(0) - c_{skip}(t)U(t)\right]}_{F_{target}(X(0), U(t))}\|\right] \tag{231}$$

Following Karras et al. (2022) we aim for the training inputs to have unit variance

$$\mathbb{V}[c_{in}(t)U(t)] = 1 \tag{232}$$
$$c_{in}(t)^2\mathbb{V}[c(t)X(0) + \zeta_t\xi] = 1 \tag{233}$$
$$c_{in}(t)^2(c(t)^2\sigma_{data}^2 + \zeta_t^2) = 1 \tag{234}$$
$$c_{in}(t) = \frac{1}{\sqrt{c(t)^2\sigma_{data}^2 + \zeta_t^2}} \tag{235}$$

for the training target to have unit variance

$$\mathbb{V}\left[\frac{1}{c_{out}(t)}[X(0) - c_{skip}(t)U(t)]\right] = 1 \tag{236}$$
$$\frac{1}{c_{out}^2(t)}\mathbb{V}\left[X(0) - c_{skip}(t)(c(t)X(0) + \zeta_t\xi)\right] = 1 \tag{237}$$
$$\frac{1}{c_{out}^2(t)}\mathbb{V}\left[(1 - c_{skip}(t)c(t))X(0) - c_{skip}(t)\zeta_t\xi)\right] = 1 \tag{238}$$
$$\frac{1}{c_{out}^2(t)}(1 - c_{skip}(t)c(t))^2\sigma_{data}^2 + c_{skip}(t)^2\zeta_t^2 = 1 \tag{239}$$
$$c_{out}^2(t) = (1 - c_{skip}(t)c(t))^2\sigma_{data}^2 + c_{skip}(t)^2\zeta_t^2 \tag{240}$$
$$P_t(x) = (1 - xc(t))^2\sigma_{data}^2 + x^2\zeta_t^2 \tag{241}$$
$$\tag{242}$$

with $x = c_{skip}(t)$ and $P(x) = c_{out}^2(t)$. To reduce error amplification we select $c_{skip}(t)$ to minimize $c_{out}(t)$ and calculate the derivative

$$0 = \frac{\partial P_t}{\partial x}(x) \tag{243}$$
$$0 = -c(t)2(1 - xc(t))\sigma_{data}^2 + 2x\zeta_t^2 \tag{244}$$
$$0 = -\sigma_{data}^2c(t) + xc(t)^2\sigma_{data}^2 + x\zeta_t^2 \tag{245}$$
$$x(c(t)^2\sigma_{data}^2 + \zeta_t^2) = \sigma_{data}^2c(t) \tag{246}$$
$$0 = \sigma_{data}^2c(t) + xc(t)^2\sigma_{data}^2 + x\zeta_t^2 \tag{247}$$
$$x(c(t)^2\sigma_{data}^2 + \zeta_t^2) = \sigma_{data}^2c(t) \tag{248}$$
$$c_{skip}(t) = \frac{\sigma_{data}^2c(t)}{c(t)^2\sigma_{data}^2 + \zeta_t^2} \tag{249}$$

we calculate

$$c_{out}^2(t) = (1 - c_{skip}(t)c(t))^2\sigma_{data}^2 + c_{skip}(t)^2\zeta_t^2 \tag{250}$$
$$= (1 - [\frac{\sigma_{data}^2c(t)}{c(t)^2\sigma_{data}^2 + \zeta_t^2}]c(t))^2\sigma_{data}^2 + [\frac{\sigma_{data}^2c(t)}{c(t)^2\sigma_{data}^2 + \zeta_t^2}]^2\zeta_t^2 \tag{251}$$
$$= (\frac{c(t)^2\sigma_{data}^2 + \zeta_t^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2} - [\frac{\sigma_{data}^2c(t)^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2}])^2\sigma_{data}^2 + [\frac{\sigma_{data}^2c(t)}{c(t)^2\sigma_{data}^2 + \zeta_t^2}]^2\zeta_t^2 \tag{252}$$
$$= (\frac{\zeta_t^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2})^2\sigma_{data}^2 + [\frac{\sigma_{data}^2c(t)}{c(t)^2\sigma_{data}^2 + \zeta_t^2}]^2\zeta_t^2 \tag{253}$$
$$= (\frac{\zeta_t^4\sigma_{data}^2}{[c(t)^2\sigma_{data}^2 + \zeta_t^2]^2}) + [\frac{\sigma_{data}^4c^2(t)\zeta_t^2}{[c(t)^2\sigma_{data}^2 + \zeta_t^2]^2}] \tag{254}$$
$$= \frac{\sigma_{data}^2\zeta_t^2(\zeta_t^2 + c(t)^2\sigma_{data}^2)}{[c(t)^2\sigma_{data}^2 + \zeta_t^2]^2} \tag{255}$$
$$= \frac{\sigma_{data}^2\zeta_t^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2} \tag{256}$$

such that

$$c_{out}(t) = \frac{\sigma_{data}\zeta_t}{\sqrt{c(t)^2\sigma_{data}^2 + \zeta_t^2}}. \tag{257}$$

In total we can write

$$\begin{aligned}
F_{target}(X(0), U(t)) &= \frac{1}{c_{out}(t)} \left[X(0) - c_{skip}(t)U(t)\right] && (258)\\
&= \frac{1}{\frac{\sigma_{data}\zeta_t}{\sqrt{c(t)^2\sigma_{data}^2+\zeta_t^2}}} \left[(1 - c_{skip}(t)c(t))X(0) - c_{skip}(t)\zeta_t\xi\right] && (259)\\
&= \frac{\sqrt{c(t)^2\sigma_{data}^2 + \zeta_t^2}}{\sigma_{data}\zeta_t} \left[\left(\frac{c(t)^2\sigma_{data}^2 + \zeta_t^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2} - \frac{\sigma_{data}^2 c(t)^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2}\right) X(0) - \frac{\sigma_{data}^2 c(t)}{c(t)^2\sigma_{data}^2 + \zeta_t^2}\zeta_t\xi\right] && (260)\\
&= \frac{\sqrt{c(t)^2\sigma_{data}^2 + \zeta_t^2}}{\sigma_{data}\zeta_t} \left[\frac{\zeta_t^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2} X(0) - \frac{\sigma_{data}^2 c(t)\zeta_t}{c(t)^2\sigma_{data}^2 + \zeta_t^2}\xi\right] && (261)\\
&\frac{\zeta_t}{\sqrt{c(t)^2\sigma_{data}^2 + \zeta_t^2}\sigma_{data}\zeta_t} \left[\zeta_t X(0) - \sigma_{data}^2 c(t)\xi\right] && (262)\\
&= \frac{\zeta_t X(0) - \sigma_{data}^2 c(t)\xi}{\sqrt{c(t)^2\sigma_{data}^2 + \zeta_t^2}\sigma_{data}} && (263)\\
& && (264)
\end{aligned}$$

Loss weighting. When using vpred we have

$$\mathcal{L}(\theta) = \mathcal{L}_{prior}(H) + \mathbb{E}\left[\underbrace{\frac{\lambda^2(t)c^2(t)w(t)c_{out}^2(t)}{\zeta_t^4}}_{=w_{eff}(t)} \|F_\theta(c_{in}(t)U(t), t) - F_{target}(X(0), U(t))\|^2\right] \tag{265}$$

with

$$\begin{aligned}
w_{eff}(t) &= w(t)\frac{\lambda^2(t)c^2(t)}{\zeta_t^4}c_{out}^2 = w(t)\frac{\lambda^2(t)c^2(t)}{\zeta_t^4}\left[\frac{\sigma_{data}\zeta_t}{\sqrt{c(t)^2\sigma_{data}^2 + \zeta_t^2}}\right]^2 && (266)\\
&= w(t)\frac{\lambda^2(t)c^2(t)}{\zeta_t^2}\frac{\sigma_{data}^2}{c(t)^2\sigma_{data}^2 + \zeta_t^2} && (267)
\end{aligned}$$

where $w(t)$ is the weight that deviates from the theory of minimizing the kl between path measures. We calculate

$$\begin{aligned}
D &= c(t)(1 - pred_lambda) + pred_lambda\zeta_t && (268)\\
&= c(t)(\frac{c(t)\sigma_{data}^2}{c(t)\sigma_{data}^2 + \zeta_t}) + \frac{\zeta_t}{c(t)\sigma_{data}^2 + \zeta_t}\zeta_t && (269)
\end{aligned}$$

so

$$\tilde{w}_{eff}(t) = \frac{c^2(t)[\sqrt{(1-lam)^2 + lam^2\sigma_{data}^2}]^2}{D^2} \tag{270}$$

$$= \frac{c^2(t)\{(\frac{c(t)\sigma_{data}^2}{\sigma_{data}^2+\zeta_t})^2 + [\frac{\zeta_t}{c(t)\sigma_{data}^2+\zeta_t}]^2\sigma_{data}^2\}}{[(\frac{c^2(t)\sigma_{data}^2}{c(t)\sigma_{data}^2+\zeta_t}) + \frac{\zeta_t}{c(t)\sigma_{data}^2+\zeta_t}\zeta_t]^2} \tag{271}$$

$$= \frac{c^2(t)\frac{c(t)^2\sigma_{data}^4}{[c(t)\sigma_{data}^2+\zeta_t]^2} + \frac{c(t)^2\zeta_t^2\sigma_{data}^2}{[c(t)\sigma_{data}^2+\zeta_t]^2}}{\frac{[c^2(t)\sigma_{data}^2+\zeta_t^2]^2}{[c(t)\sigma_{data}^2+\zeta_t]^2}} \tag{272}$$

$$= \frac{\sigma_{data}^2 c^2(t)[c(t)^2\sigma_{data}^2 + \zeta_t^2]}{[c^2(t)\sigma_{data}^2 + \zeta_t^2]^2} \tag{273}$$

$$= \frac{\sigma_{data}^2 c^2(t)}{c^2(t)\sigma_{data}^2 + \zeta_t^2}. \tag{274}$$

sigma. The loss weighting for "sigma" is simply the above calculated

$$w_{eff}(t) = \frac{\sigma_{data}^2 c^2(t)}{c^2(t)\sigma_{data}^2 + \zeta_t^2} \Rightarrow w(t) = \frac{\zeta_t^2}{\lambda^2(t)} \tag{275}$$

kl_sigma. We only need to multiply the derived "sigma" weight with the weight calculated in line 198 of training.py $w = \lambda^2(t)$ so that

$$w_{eff}(t) = \lambda^2(t)\frac{\sigma_{data}^2 c^2(t)}{c^2(t)\sigma_{data}^2 + \zeta_t^2} \Rightarrow w(t) = \zeta_t^2 \tag{276}$$

kl. We only need to multiply the derived "sigma" weight with the weight calculated in line 200 of training.py $w = \lambda^2(t)/cond_var$ so that

$$w_{eff}(t) = \frac{\lambda^2(t)}{\zeta_t^2}\frac{\sigma_{data}^2 c^2(t)}{c^2(t)\sigma_{data}^2 + \zeta_t^2} \Rightarrow w(t) \equiv 1 \tag{277}$$

Hence we minimize in this setting the exact kl-divergence between path measures.

uniform. We simply choose

$$w_{eff}(t) \equiv 1 \Rightarrow w(t) = \frac{\zeta_t^2(c^2(t)\sigma_{data}^2 + \zeta_t^2)}{\lambda^2(t)c^2(t)\sigma_{data}^2} \tag{278}$$

kl_uniform. We simply choose

$$w_{eff}(t) = \lambda^2(t) \Rightarrow w(t) = \frac{\zeta_t^2(c^2(t)\sigma_{data}^2 + \zeta_t^2)}{c^2(t)\sigma_{data}^2} \tag{279}$$

G FLOW-MATCHING, OT-ODE AND PF-ODE

Define for a data distribution $X_0 \sim p_0$ and $\varepsilon \sim p_1 = \mathcal{N}(0, I_d)$ the path

$$X(t) = (1-t)X(0) + t\varepsilon. \tag{280}$$

with

$$\partial_t X(t) = \varepsilon - X(0) = \begin{cases} = \varepsilon - \frac{X(t)-t\varepsilon}{(1-t)} = \frac{\varepsilon - X(t)}{(1-t)}, & (p_0 \to p_1) \\ = \frac{X(t)-(1-t)X(0)}{t} - X(0) = \frac{X(t)-X(0)}{t}, & (p_1 \to p_0) \end{cases} \tag{281}$$

so that X_t solves the random OT-ODEs

$$\mathrm{d}X(t) = \begin{cases} = \frac{\varepsilon - X(t)}{(1-t)}\,\mathrm{d}t, & (p_0 \to p_1), \\ = \frac{X(t)-X(0)}{t}\,\mathrm{d}t & (p_1 \to p_0). \end{cases} \tag{282}$$

- For any fixed given pair of data and noise (x_0, ε) we can construct the straight-line interpolation $x_t = (1-t)x_0 + t\varepsilon$.
- We can sample trajectories of the conditional OT-ODE by first sampling $(x_0, \varepsilon) \sim p_0 \otimes p_1$ and second follow the conditional trajectory of the OT-ODE.
- Optimal transport (OT) or entroptic optimal transport (EOT) with quadratic costs on the other hand samples $(x_0, \varepsilon) \sim \pi^\star_{0,1}$ from an *optimal* coupling, that is the solution to

$$\pi^\star_{0,1} = \arg\min_{\pi} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|x-y\|^2 \,\mathrm{d}\pi(x,y) + \sigma D_{KL}(\pi \| p_0 \otimes p_1) \right\} \tag{283}$$

  which is called OT-problem for $\sigma = 0$ and EOT-problem for ($\sigma > 0$).
- We can transform p_0 to via p_1 via OT or EOT if we *first* sample $(x_0, \varepsilon) \sim \pi^\star_{0,1}$ and *second* follow the conditional OT-ODE or the "EOT-SDE", which is the the scaled Brownian bridge that solves

$$\mathrm{d}X_t = \frac{\varepsilon - X(t)}{(1-t)}\,\mathrm{d}t + \sqrt{\sigma}\,\mathrm{d}B(t), X(0) = x_0. \tag{284}$$

Note that both drifts of the OT-ODE depend on the end points of the trajectory, so we can not use the OT-ODE directly to build a generative model $p_1 \to p_0$, since we would already need to know where we want to go. To get rid of this dependence, we define the vector field

$$v_t(x) = \frac{x - \mathbb{E}[X(0)|X_t = x]}{t} \tag{285}$$

and aim to approximate the path of the ODE

$$\mathrm{d}\hat{X}(t) = v_t(\hat{X}(t))\,\mathrm{d}t. \tag{286}$$

One can show that X and $\hat{X}$ share share the same marginals, but X and $\hat{X}$ *do not* have the same path distribution, i.e. their corresponding path measures differ. Denote with p_t the density of X_t. Since $X(0)|(X(0) = x_0) \sim \mathcal{N}((1-t)x_0, t^2 I_d)$ is Gaussian we have

$$\begin{aligned}
\nabla_x \log p_t(x) &= \mathbb{E}[\nabla_x \log p_t(x|X_0)|X(t) = x] && (287)\\
&= \mathbb{E}[-\frac{x - (1-t)X(0)}{t^2}|X(t) = x] && (288)\\
&= -\frac{x - (1-t)\mathbb{E}[X(0)|X(t) = x]}{t^2} && (289)
\end{aligned}$$

and hence in total

$$\mathbb{E}[X(0)|X(t) = x] = \frac{x + t^2 \nabla_x \log p_t(x)}{1-t} \tag{290}$$

such that

$$\begin{aligned}
v_t(x) &= \frac{x - \mathbb{E}[X(0)|X_t = x]}{t} && (291)\\
&= \frac{x - \frac{x + t^2\nabla_x \log p_t(x)}{1-t}}{t} && (292)\\
&= \frac{\frac{(1-t)x}{(1-t)} - \frac{x + t^2 \nabla_x p_t(x)}{1-t}}{t} && (293)\\
&= \frac{(1-t)x - x - t^2 \nabla_x \log p_t(x)}{(1-t)t} && (294)\\
&= -\frac{(x + t\nabla_x \log p_t(x)}{(1-t)} && (295)
\end{aligned}$$

and in total we have the flow matching ODE that is used to go from noise at $t = 1$ to data at $t = 0$ by simulating

$$\mathrm{d}\hat{X}(t) = \left[-\frac{\hat{X}(t)}{1-t} - \frac{t}{1-t}\nabla_x \log p_t(\hat{X}_t) \right] \mathrm{d}t \tag{296}$$

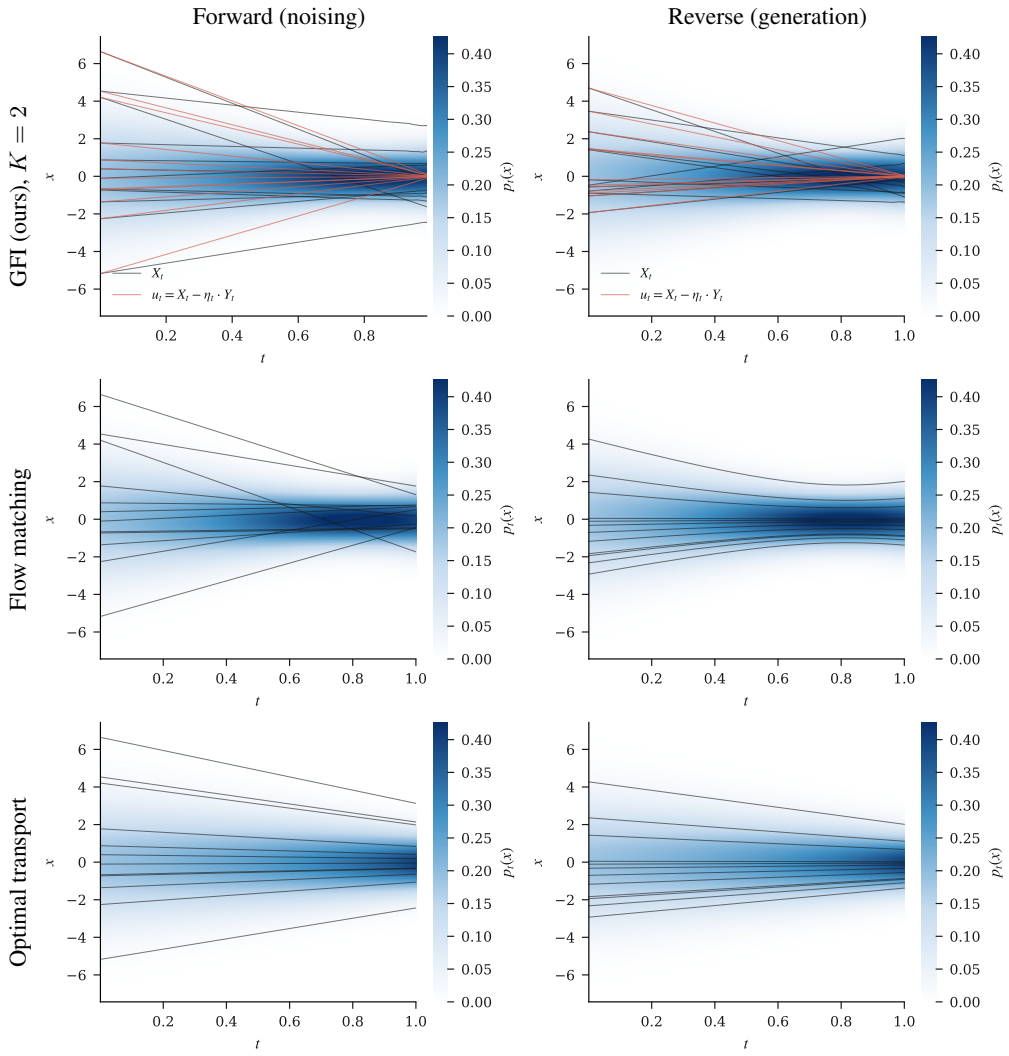


Figure 5: Three transports between $\mathcal{N}(0, 2.124^2)$ and the prior $\mathcal{N}(0, 1)$, over the exact marginal $p_t(x)$. Red curves are the neural network input $u_t = X_t - \alpha_t^\top Y_t$ for $K > 0$. **Left:** forward, data $\to$ prior. **Right:** reverse, using the closed-form score. All three methods move mass along straight chords in the forward process $x_a = (1-a)x_0 + ax_1$ and differ only in the coupling: independent for GFNs and flow matching, so their chords cross. For optimal transport the chords do not cross.

which is the PF-ODE corresponding to the the forward SDE

$$\mathrm{d}X(t) = \mu(t)X_t\,\mathrm{d}t + g(t)\,\mathrm{d}B(t), \quad \mu(t) = -\frac{1}{1-t}, \quad g(t) = \sqrt{\frac{2t}{1-t}}. \tag{297}$$

The PF-ODE has by construction no crossings *at the same time*. So if $X_t^1 = X_t^2$, we already know that both trajectories X^1 and X^2 coincide. Our framework on the other hand *has* crossings, but the lines are more straight.

H Details on the covariance interpolation

The covariance matrix Σ_t lacks a closed-form expression and must be computed by solving an ODE over the domain $t \in [0, T]$ for a fixed Hurst exponent H. Our goal is to make H learnable without repeatedly solving the ODE during training.

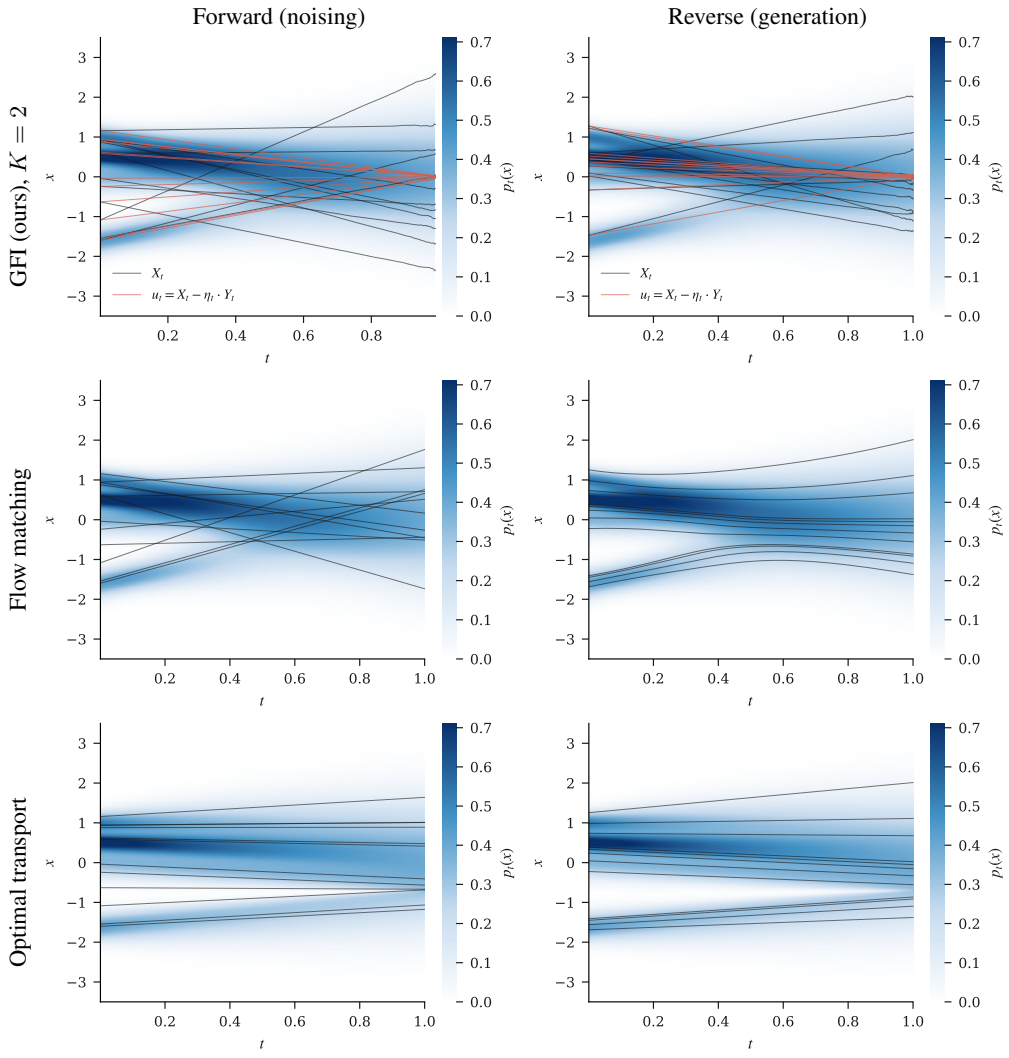


Figure 6: Three transports between a four-component Gaussian mixture and the prior $\mathcal{N}(0, 1)$ in one dimension, over the exact marginal $p_t(x)$. Red curves are the neural network input $u_t = X_t - \alpha_t^\top Y_t$ for $K > 0$. **Left:** forward, data $\to$ prior. **Right:** reverse, using the closed-form score. All three methods both move mass along straight chords $x_a = (1-a)x_0 + ax_1$ and differ only in the coupling: independent for GFNs and flow matching, so their chords cross. For optimal transport the chords do not cross..

The key observation is that the Hurst exponent enters the ODE parametrically rather than through differential operators. We can therefore sample the function $f : [0, T] \times [0, 1] \to \mathbb{R}^{K+3\times K+3}$ that maps time t and Hurst exponent H to the covariance matrix Σ_t for a fixed H by solving the ODE for a (t, H) pair.

Since the vector field in is a composition of continuous functions, it itself is continuous and the ODE solution is differentiable. This informs the hypothesis that the Chebyshev interpolation should converge quickly, meaning a small number of nodes suffices for high accuracy. We precompute the ODE solutions on a tensor grid of Chebyshev nodes over $H \in [0, 1]$ and $t \in [0, T]$, then construct a bivariate Chebyshev interpolant for each component function of f.

During training, we treat H as a learnable parameter and evaluate Σ_t using the Chebychev series. Since the series is differentiable, gradients with respect to H are obtained through standard back-propagation. The complete procedure is illustrated in 7.

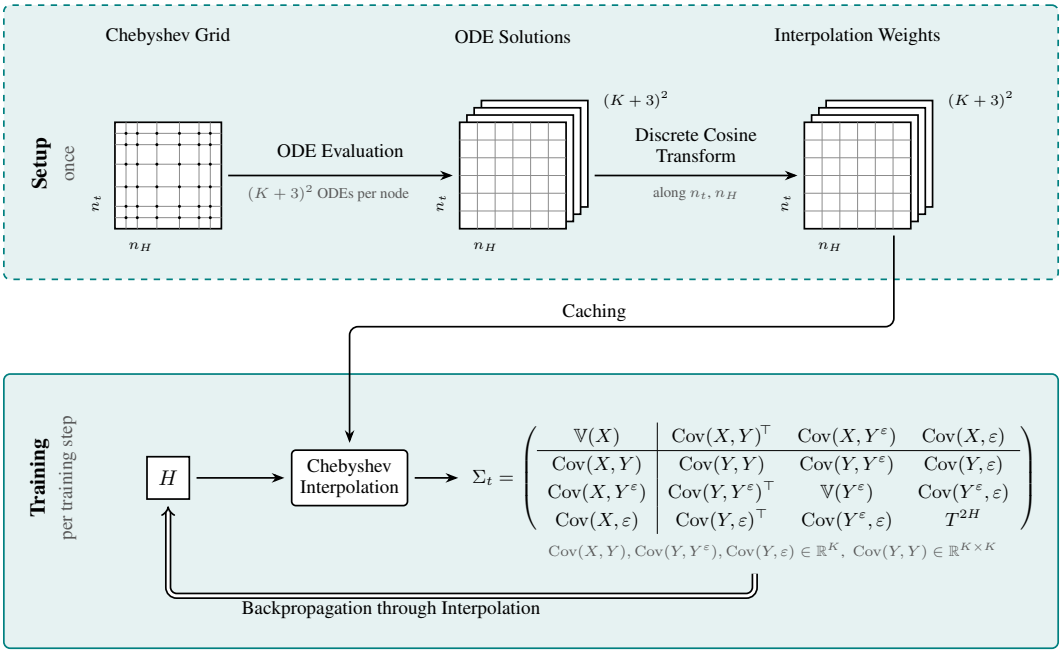


Figure 7: Computation of and backpropagation through the covariance matrix Σ_t with learnable Hurst parameter H at time t using Chebyshev interpolation for efficient evaluation of the covariance Σ_t. The setup stage is performed once and carries no gradients; only the interpolation is traversed during training. All $(K+3)^2$ entries of $\Sigma_t \in \mathbb{R}^{(K+3)\times(K+3)}$ are interpolated jointly from the cached coefficients.

The covariance matrix Σ_t lacks a closed-form expression and must be computed by solving an ODE over the domain $t \in [0, T]$ for a fixed Hurst exponent H. Our goal is to make H learnable without repeatedly solving the ODE during training.

The key observation is that the Hurst exponent enters the ODE parametrically rather than through differential operators. We can therefore sample the function $f : [0, T] \times [0, 1] \to \mathbb{R}^{K+3\times K+3}$ that maps time t and Hurst exponent H to the covariance matrix Σ_t for a fixed H by solving the ODE for a (t, H) pair.

Since the vector field in is a composition of continuous functions, it itself is continuous and the ODE solution is differentiable. This informs the hypothesis that the Chebyshev interpolation should converge quickly, meaning a small number of nodes suffices for high accuracy. We precompute the ODE solutions on a tensor grid of Chebyshev nodes over $H \in [0, 1]$ and $t \in [0, T]$, then construct a bivariate Chebyshev interpolant for each component function of f.

During training, we treat H as a learnable parameter and evaluate Σ_t using the Chebychev series. Since the series is differentiable, gradients with respect to H are obtained through standard back-propagation. The complete procedure is illustrated in 7.

I Experimental setup

For all reported experiments we use models from the family of Diffusion Transformers introduced by Peebles & Xie (2023). While initially only introduced for latent diffusion and lacking the inductive bias of convolution-based models such as U-Nets (Ronneberger et al. (2015)), our experiments have shown that the increased compute efficiency of DiTs makes up for its slower convergence speed on pixel datasets. We therefore select DiTs of different scales for both pixel and latent datasets. We deviate from the canonical implementation of DiTs by using QK-normalization (Dehghani et al. (2023)) in the attention blocks, which allows us to train the models in BFLOAT16 precision. All reported results are trained in BFLOAT16 precision for the DiT backbone. During sampling, we use

FLOAT32 evaluations of the score network and FLOAT64 accumulators to prevent accumulation of floating point errors in the integration.

For sampling, a common method is the utilization of Exponential Moving Averages of model weights to improve visual fidelity and image quality metrics. As shown by Karras et al. (2024) the classical formulation of EMA is heavily influenced by its random initialization during training. We instead utilize the Power Function EMA (Karras et al. (2024)) with an EMA length $\sigma_{\text{rel}} = 0.1$ for all reported experiments. To increase reproducability, all training and sampling experiments reported here share the same initial seed.

For the evaluation of image quality such as Fréchet Distance, for both metrics based on Inception-V3 and DINO-v2 model embeddings, we follow the protocol of ADM Dhariwal & Nichol (2021) and sample 50 000 samples from a trained model and compute statistics against the full training dataset. For the computation of the Improved Precision and Recall we follow the protocol introduced by Kynkäänniemi et al. (2019) and subsample the generated samples or the real samples to equal length and compute the metrics on these sets. This results in 50000 samples for both real and fake samples on MNIST, CIFAR-10 and Imagenet and 28000 samples on CelebA. In contrast to the original implementation we compute Improved Precision and Recall in the embedding space of a Inception-V3 model. In order to evaluate diversity of the generated samples, we compute the Vendi score (Friedman & Dieng (2023)) in the embedding space of Inception-V3 for the whole set of generated samples.

J ADDITIONAL EXPERIMENTS

J.1 TOY EXAMPLES

Figs. 8 and 9 visualize the variability of reverse sample paths and neural network inputs u in dependence of the Hurst exponent H and the number of auxilliary processes K.

J.2 BASELINE ON MNIST

Tab. 5 shows all evaluated baselines with varying target and weighting formulations for the MNIST dataset.

Table 5: Prediction target and loss weighting on MNIST ($K = 0$, $H = 0.5$, DiT-S/4, 100k steps, 50k samples) for VP. FID, improved precision/recall and Vendi use Inception features; FD-DINOv2 uses DINOv2. Best value per column in bold; ϵ-prediction was only run for three weightings.

		VP				
Target	Weighting	FID ↓	FD-DINO ↓	Prec. ↑	Rec. ↑	Vendi
ϵ-pred.	σ	1.42	17.04	0.772	**0.736**	4.73
	KL	3.30	43.61	0.691	0.693	4.69
	KL-σ	1.53	16.24	0.802	0.714	4.73
v-pred.	uniform	1.99	25.05	0.781	0.688	4.74
	σ	1.72	15.64	0.779	0.710	4.74
	KL	2.72	28.73	0.716	0.685	4.65
	KL-σ	1.57	14.75	0.803	0.707	4.74
	KL-uniform	5.31	85.55	0.614	0.614	4.83
x_0-pred.	uniform	3.62	43.28	0.801	0.611	4.62
	σ	**1.13**	**10.82**	0.814	0.734	4.71
	KL	2.75	31.64	0.731	0.680	4.65
	KL-σ	1.38	13.09	**0.836**	0.707	4.68
	KL-uniform	8.69	137.77	0.727	0.412	4.64

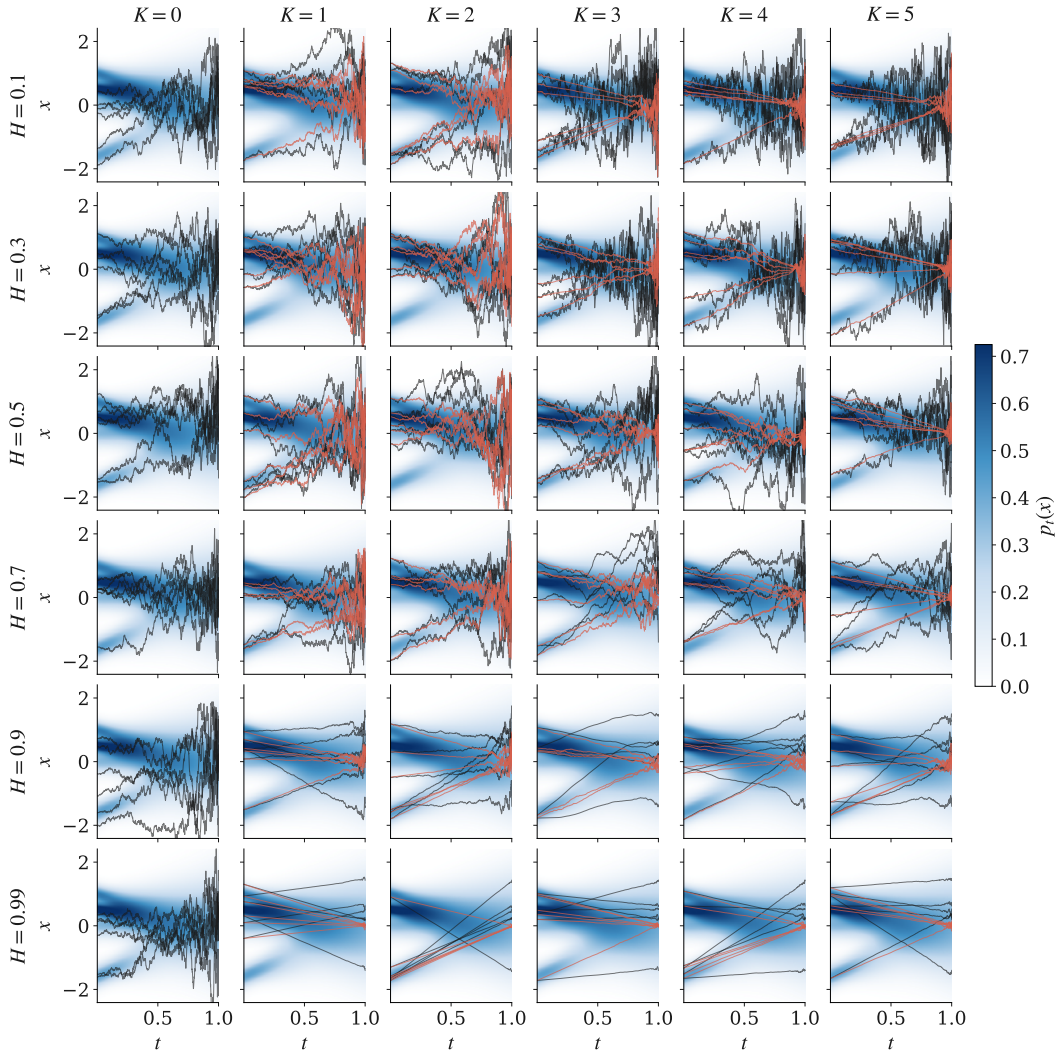


Figure 8: **Markov approximation order K and Hurst index H shape the transport.** Fractional flow matching (Type I) on a standardised four-component Gaussian mixture in 1D, over $H \in \{0.1, 0.3, 0.5, 0.7, 0.9, 0.99\}$ (rows) and $K \in \{0, \dots, 5\}$ OU modes (columns), with $\gamma \in [0.1, 20]$. Shading is the marginal $p_t(x)$ of the forward process. Dark curves are six reverse trajectories X_t integrated with the analytic mixture score by Euler–Maruyama over 1000 steps from $t = 10^{-3}$ to $T_{\text{ode}} = 1$. All cells share one prior draw. Red curves are the neural network input $u_t = X_t - \alpha_t^\top Y_t$ for $K > 0$. It is absent at $K = 0$, where standard non-augmented score matching is shown. Increasing H smooths the paths and increasing K smooths the neural network input u_t.

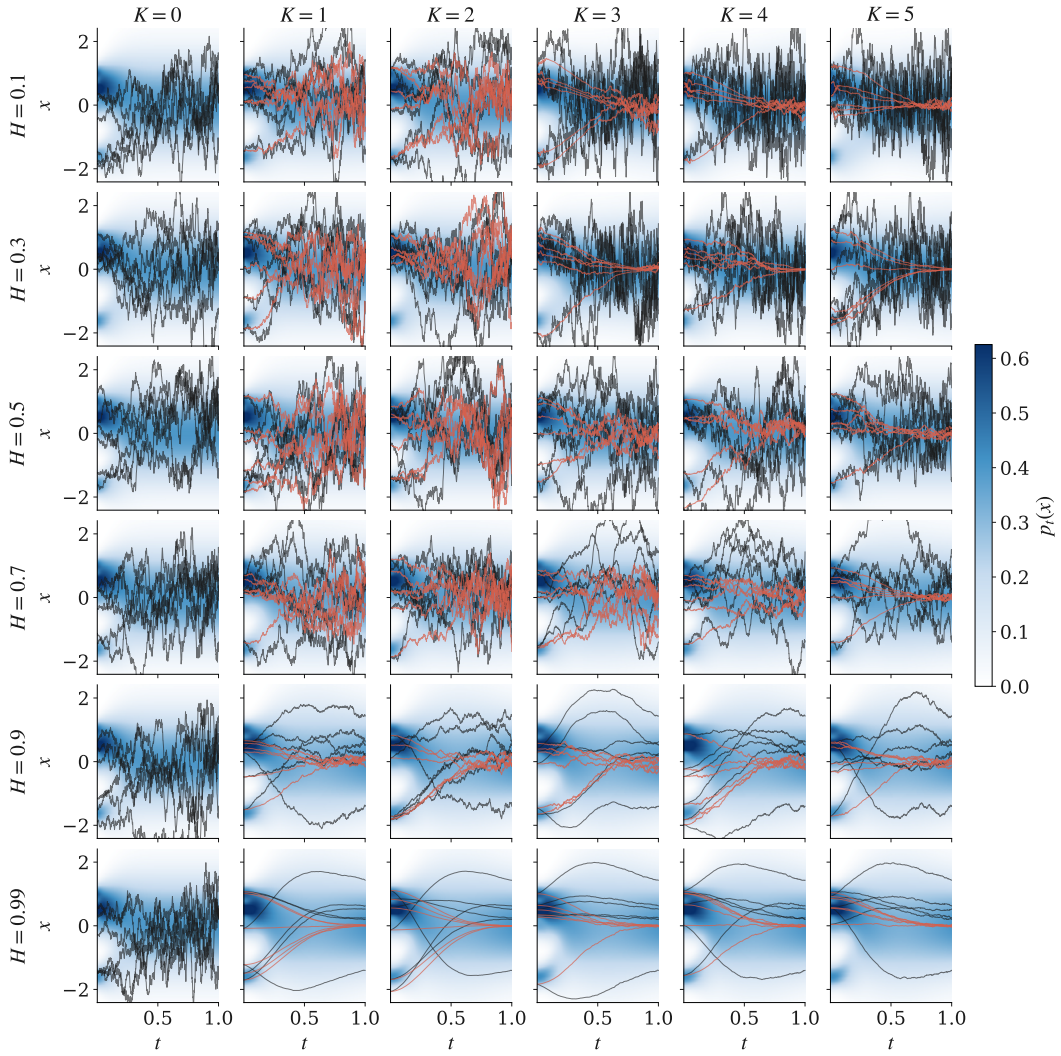


Figure 9: **Markov approximation order K and Hurst index H shape the transport.** FVP Type I on a standardised four-component Gaussian mixture in 1D, over $H \in \{0.1, 0.3, 0.5, 0.7, 0.9, 0.99\}$ (rows) and $K \in \{0, \ldots, 5\}$ OU modes (columns), with $\gamma \in [0.1, 20]$. Shading is the marginal $p_t(x)$ of the forward process. Dark curves are six reverse trajectories X_t integrated with the analytic mixture score by Euler–Maruyama over 1000 steps from $t = 10^{-3}$ to $T_{\text{ode}} = 1$. All cells share one prior draw. Red curves are the neural network input $u_t = X_t - \alpha_t^\top Y_t$ for $K > 0$. It is absent at $K = 0$, where standard non-augmented score matching is shown. Increasing H smooths the paths and increasing K smooths the neural network input u_t.

J.3 Additional experiments on CIFAR-10 with DiT-S/4

Tab. 6 shows all evaluated baselines with varying target and weighting formulations for the CIFAR10 dataset.

Table 6: Prediction target and loss weighting on CIFAR-10 ($K = 0$, $H = 0.5$, DiT-S/4, 400k steps, 50k samples) for VE, VP and diffusion flow matching (DFM). FID, improved precision/recall and Vendi use Inception features; FD-DINOv2 uses DINOv2. Best value per column in bold; ϵ-prediction was only run for three weightings. The two VE KL-uniform runs and the DFM x_0-pred. KL run diverged (†) and are excluded from bolding.

		VE					VP					DFM				
Target	Weighting	FID ↓	FD-DINO ↓	Prec. ↑	Rec. ↑	Vendi	FID ↓	FD-DINO ↓	Prec. ↑	Rec. ↑	Vendi	FID ↓	FD-DINO ↓	Prec. ↑	Rec. ↑	Vendi
ϵ-pred.	σ	20.16	884.81	0.570	0.538	15.36	**12.71**	**698.68**	0.604	**0.547**	16.92	**12.25**	**668.91**	0.603	**0.561**	16.54
	KL	21.36	897.25	0.574	**0.541**	15.35	26.82	1023.09	0.581	0.491	15.66	32.79	1099.77	0.582	0.425	14.74
	KL-σ	25.96	1008.85	0.535	0.461	15.94	20.43	792.03	0.608	0.455	16.38	17.68	747.70	0.612	0.484	16.32
v-pred.	uniform	21.09	886.72	**0.598**	0.507	15.70	24.25	935.89	0.602	0.452	16.31	18.56	828.63	0.606	0.502	16.13
	σ	21.34	889.68	0.578	0.515	14.95	13.47	701.40	0.602	0.535	16.74	12.74	686.00	0.607	0.545	16.37
	KL	**19.02**	854.69	0.575	0.533	15.23	27.30	1008.64	0.582	0.462	15.28	34.12	1082.84	0.585	0.426	14.79
	KL-σ	24.06	955.43	0.574	0.469	15.89	22.76	818.47	0.609	0.439	16.39	20.47	780.92	**0.614**	0.454	16.20
	KL-uniform	144.26†	1499.97†	0.262†	0.014†	15.77	36.51	1078.37	0.575	0.354	16.51	33.58	1095.83	0.566	0.374	16.04
x_0-pred.	uniform	28.76	884.14	0.533	0.435	15.42	27.46	943.06	0.568	0.399	15.52	33.86	919.22	0.519	0.366	15.96
	σ	19.30	**843.59**	0.583	0.529	15.07	14.33	717.04	**0.617**	0.528	16.09	19.61	821.72	0.611	0.500	15.30
	KL	19.95	857.42	0.586	0.519	15.17	26.74	1018.77	0.584	0.468	14.86	150.23†	1510.50†	0.307†	0.025†	12.69
	KL-σ	25.54	895.95	0.557	0.449	15.36	21.69	771.68	0.606	0.433	15.56	22.02	754.75	0.584	0.440	15.47
	KL-uniform	197.91†	2179.85†	0.138†	0.009†	11.51	37.30	1046.39	0.465	0.363	15.56	77.77	1265.40	0.371	0.213	14.48

Table 7: Type I vs Type II FVP on MNIST, by Hurst index H (rows) and number of OU components K (column groups). All use $\mathrm{loss_weight} = \mathrm{kl_sigma}$. Bold marks the better of the two kinds at each (H, K). A box marks the overall best value of each metric block.

	$K = 2$		$K = 3$		$K = 4$		$K = 5$	
H	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
				FD-DINOv2 ↓				
0.1	**10.81**	13.04	**12.27**	13.61	14.90	**13.50**	16.96	**14.52**
0.2	**10.85**	12.53	**15.29**	18.85	22.05	**16.50**	35.29	**14.85**
0.3	**11.36**	12.74	**22.62**	188.37	**45.26**	159.33	155.49	**21.38**
0.4	**11.16**	12.19	**7.46**	10.27	**8.00**	12.46	**7.43**	202.16
0.5	**11.12**	12.75	**9.73**	11.01	**9.89**	11.45	**8.95**	16.29
0.6	**10.68**	12.99	**10.70**	12.47	**10.16**	11.58	**7.76**	14.88
0.7	**10.25**	12.80	**10.64**	11.80	**10.36**	11.95	**7.65**	15.48
0.8	**9.63**	12.69	**8.44**	11.87	**9.60**	12.50	**7.36**	16.84
0.9	**8.74**	13.05	**7.21**	12.06	**7.20**	13.40	**8.03**	20.21
0.99	258.38	**14.53**	52.56	**12.16**	48.04	**17.45**	41.36	**36.23**
				FID ↓				
0.1	**1.24**	1.26	1.30	**1.27**	1.41	**1.24**	1.55	**1.38**
0.2	1.26	**1.24**	**1.37**	1.62	1.80	**1.42**	3.09	**1.40**
0.3	1.29	**1.25**	**2.13**	31.17	**7.39**	22.74	26.19	**1.87**
0.4	1.29	**1.22**	**0.97**	1.03	**1.05**	1.40	**1.01**	33.12
0.5	**1.28**	1.29	1.17	**1.09**	**1.13**	1.27	**1.12**	1.61
0.6	1.28	**1.24**	**1.23**	1.35	**1.18**	1.26	**0.99**	1.43
0.7	**1.20**	1.28	**1.23**	1.24	**1.27**	1.29	**0.95**	1.47
0.8	**1.11**	1.21	**1.08**	1.16	**1.18**	1.28	**0.92**	1.61
0.9	**0.95**	1.24	**0.91**	1.18	**0.88**	1.31	**0.90**	1.85
0.99	41.79	**1.36**	7.59	**1.19**	6.73	**1.61**	5.53	**3.56**

Table 8: Type I vs Type II FVP on CIFAR-10, by Hurst index H (rows) and number of OU components K (column groups). All use $\mathrm{loss_weight} = \mathrm{kl_sigma}$. Bold marks the better of the two kinds at each (H, K). A box marks the overall best value of each metric block.

	$K=2$		$K=3$		$K=4$		$K=5$	
H	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
				FD-DINOv2 ↓				
0.1	**701.85**	713.60	801.36	**723.91**	858.46	**788.91**	903.57	**727.74**
0.2	**704.15**	717.74	**861.26**	878.07	907.43	**846.21**	991.23	**761.64**
0.3	**705.08**	705.70	**858.19**	1115.52	**799.86**	1257.39	895.96	**885.86**
0.4	**706.79**	708.81	**636.26**	736.23	624.24	–	**658.77**	1380.57
0.5	706.45	**705.52**	**646.16**	704.15	**648.02**	718.58	**632.18**	834.79
0.6	**702.29**	710.33	**675.77**	694.79	**653.71**	712.11	**620.54**	787.84
0.7	**683.95**	709.15	**678.04**	687.03	**661.47**	715.72	**624.19**	786.22
0.8	**655.65**	712.13	**614.17**	690.26	**636.51**	721.03	**640.33**	818.86
0.9	**653.98**	714.54	**660.99**	708.36	**663.75**	751.74	**686.30**	893.20
0.99	1139.72	**764.88**	913.36	**710.09**	917.91	**829.39**	**919.36**	1013.80
				FID ↓				
0.1	15.50	**14.03**	15.68	**13.06**	17.80	**15.21**	19.87	**14.45**
0.2	15.44	**14.07**	**17.71**	19.45	22.28	**17.44**	30.20	**15.31**
0.3	15.55	**13.21**	**20.29**	43.54	**22.21**	57.09	30.02	**20.21**
0.4	15.65	**13.06**	**12.31**	13.58	12.10	–	**13.49**	91.54
0.5	15.40	**13.26**	12.82	**12.65**	**13.03**	13.73	**12.69**	17.96
0.6	15.47	**13.63**	14.12	**12.54**	13.42	**13.35**	**11.99**	15.45
0.7	14.74	**13.53**	13.73	**12.16**	**13.35**	13.53	**11.77**	15.21
0.8	**13.21**	13.49	**11.47**	12.47	**12.21**	13.48	**11.74**	16.58
0.9	**11.80**	13.35	**12.41**	13.02	**12.00**	14.59	**12.38**	20.43
0.99	47.83	**14.93**	26.11	**13.03**	26.16	**17.62**	**25.46**	29.60

Table 9: Flow matching (FM_TypeI) on CIFAR-10, by Hurst index H (rows) and number of OU components K (columns). There is no Type II FlowMatching class in the codebase, so FM is reported on its own rather than as a Type I/II comparison. All use $\mathrm{loss_weight} = \mathrm{kl_sigma}$. Bold marks the best K at each H. A box marks the overall best value of each metric block.

H	$K=2$	$K=3$	$K=4$	$K=5$
		FD-DINOv2 ↓		
FM (K=0)		–		
0.1	**735.50**	831.00	852.40	844.30
0.2	**730.30**	821.80	744.80	856.30
0.3	733.00	710.50	**664.10**	766.60
0.4	738.20	**657.00**	664.50	720.20
0.5	740.30	679.30	676.80	**643.60**
0.6	727.90	705.00	**671.50**	688.90
0.7	706.70	699.00	**663.00**	689.40
0.8	709.70	**666.10**	693.50	758.70
0.9	728.80	**707.50**	717.00	747.70
0.99	1120.00	974.50	967.10	**947.50**
		FID ↓		
FM (K=0)		–		
0.1	17.84	**17.09**	17.71	17.19
0.2	17.21	15.96	**13.75**	17.69
0.3	16.85	13.34	**12.83**	15.09
0.4	17.31	**13.65**	14.53	14.34
0.5	17.60	15.83	15.61	**14.39**
0.6	16.63	15.91	14.98	**12.84**
0.7	15.16	14.83	13.98	**12.46**
0.8	14.03	**12.34**	13.83	14.70
0.9	13.37	**12.71**	13.09	13.82
0.99	42.44	27.79	26.29	**24.05**

J.4 Additional experiments on CIFAR-10 with DiT-B/2

Table 10: Quantitative results on CIFAR10, where we train GFI over the Hurst range $(0.01, 0.995)$ and evaluate at the $H \in (0.01, 0.4)$ in the rough regime with the smallest prior loss, $H \in (0.5, 0.79)$ in the smooth regime with the smallest prior loss, $H \in (0.8, 0.99)$ with the smallest prior loss and at the H with the smallest path-KL bound; we report the H with the lowest FD-DINOv2.

		CIFAR-10									
Schedule	Driving noise	H	K	Target	weighting	FID* ↓	FID ↓	FD-DINOv2 ↓	Precision ↑	Recall ↑	Vendi Score ↑
FM	random straight-lines	-	-	$\xi - x_0$	uniform	8.63	10.55	513.87	0.64	0.54	16.92
FM	random straight-lines	-	-	$\xi - x_0$	variance	6.87	8.49	399.17	0.65	0.55	17.11
VP	BM	-	-	noise	variance	6.15	7.14	394.78	0.66	0.55	17.60
VP	BM	-	-	vpred	variance	6.46	9.22	408.53	0.67	0.52	17.69
Diffusion FM	BM	-	-	noise	variance	6.37	8.30	374.24	0.67	0.53	17.50
Diffusion FM	BM	-	-	vpred	variance	5.80	7.50	365.16	0.67	0.55	17.42
$\text{GFI}_{VP(K=1)}$	ε-MA-fBM	0.80	1	noise	KL-σ	6.27	8.43	403.77	0.67	0.53	17.00
$\text{GFI}_{VP(K=2)}$	ε-MA-fBM	0.80	2	noise	KL-σ	6.38	10.18	412.83	0.67	0.50	16.95
$\text{GFI}_{VP(K=3)}$	ε-MA-fBM	$0.03^\dagger$	3	noise	KL-σ	5.19	$5.19^\dagger$	$381.54^\dagger$	$0.65^\dagger$	$0.58^\dagger$	$17.42^\dagger$
$\text{GFI}_{VP(K=4)}$	ε-MA-fBM	0.73	4	noise	KL-σ	5.26	8.03	329.17	0.68	0.52	17.40
$\text{GFI}_{VP(K=5)}$	ε-MA-fBM	0.66	5	noise	KL-σ	5.89	8.75	351.73	0.65	0.53	17.77
$\text{GFI}_{FM(K=1)}$	ε-MA-fBM	0.80	1	noise	KL-σ	6.43	7.69	415.30	0.65	0.55	17.19
$\text{GFI}_{FM(K=2)}$	ε-MA-fBM	0.80	2	noise	KL-σ	6.72	7.35	426.18	0.65	0.55	17.23
$\text{GFI}_{FM(K=3)}$	ε-MA-fBM	0.50	3	noise	KL-σ	5.69	8.47	375.80	0.67	0.53	17.35
$\text{GFI}_{FM(K=4)}$	ε-MA-fBM	0.50	4	noise	KL-σ	5.83	6.78	369.07	0.65	0.56	17.41
$\text{GFI}_{FM(K=5)}$	ε-MA-fBM	0.92	5	noise	KL-σ	5.64	6.96	359.61	0.65	0.57	17.28
$\text{GFI}_{VP(K=1)}$	ε-MA-fBM	0.80	1	vpred	KL-σ	6.82	8.94	408.35	0.67	0.52	17.11
$\text{GFI}_{VP(K=2)}$	ε-MA-fBM	0.80	2	vpred	KL-σ	7.06	8.71	419.85	0.66	0.52	17.01
$\text{GFI}_{VP(K=3)}$	ε-MA-fBM	$0.03^\dagger$	3	vpred	KL-σ	5.49	$5.49^\dagger$	$369.80^\dagger$	$0.66^\dagger$	$0.57^\dagger$	$17.54^\dagger$
$\text{GFI}_{VP(K=4)}$	ε-MA-fBM	0.73	4	vpred	KL-σ	5.32	8.36	316.65	0.69	0.51	17.44
$\text{GFI}_{VP(K=5)}$	ε-MA-fBM	0.66	5	vpred	KL-σ	5.62	7.36	342.03	0.65	0.55	17.56
$\text{GFI}_{VP(K=6)}$	ε-MA-fBM	0.61	6	vpred	KL-σ	6.51	7.90	398.52	0.62	0.58	17.66
$\text{GFI}_{VP(K=7)}$	ε-MA-fBM	0.02	7	vpred	KL-σ	6.51	6.51	426.72	0.61	0.60	17.26
$\text{GFI}_{VP(K=8)}$	ε-MA-fBM	0.02	8	vpred	KL-σ	7.33	7.35	426.36	0.61	0.58	17.29
$\text{GFI}_{VP(K=9)}$	ε-MA-fBM	0.03	9	vpred	KL-σ	7.78	7.78	432.89	0.61	0.59	17.20
$\text{GFI}_{VP(K=10)}$	ε-MA-fBM	0.01	10	vpred	KL-σ	9.08	9.08	569.76	0.55	0.61	17.49
$\text{GFI}_{FM(K=1)}$	ε-MA-fBM	0.80	1	vpred	KL-σ	7.13	9.81	418.70	0.67	0.50	17.28
$\text{GFI}_{FM(K=2)}$	ε-MA-fBM	0.80	2	vpred	KL-σ	7.17	8.63	428.43	0.66	0.52	17.24
$\text{GFI}_{FM(K=3)}$	ε-MA-fBM	0.02	3	vpred	KL-σ	5.93	8.66	365.45	0.68	0.51	17.25
$\text{GFI}_{FM(K=4)}$	ε-MA-fBM	0.92	4	vpred	KL-σ	6.34	11.07	365.41	0.69	0.48	17.27
$\text{GFI}_{FM(K=5)}$	ε-MA-fBM	0.92	5	vpred	KL-σ	5.72	8.28	333.34	0.66	0.54	17.62
$\text{GFI}_{FM(K=6)}$	ε-MA-fBM	0.92	6	vpred	KL-σ	5.19	5.19	374.33	0.65	0.59	17.07
$\text{GFI}_{FM(K=7)}$	ε-MA-fBM	0.92	7	vpred	KL-σ	5.24	5.24	409.84	0.64	0.59	16.81
$\text{GFI}_{FM(K=8)}$	ε-MA-fBM	0.93	8	vpred	KL-σ	4.77	4.77	380.81	0.65	0.59	17.19
$\text{GFI}_{FM(K=9)}$	ε-MA-fBM	0.94	9	vpred	KL-σ	5.35	5.35	395.35	0.64	0.60	17.06
$\text{GFI}_{FM(K=10)}$	ε-MA-fBM	0.95	10	vpred	KL-σ	5.24	5.24	405.25	0.64	0.59	16.89

J.5 Additional experiments on CIFAR-10-LT with DiT-B/2

Table 11: Quantitative results on CIFAR10_LT (DiT-B/2, 400k steps, 50k samples). GFI is trained over the Hurst range $(0.01, 0.995)$ and evaluated at the H with the smallest prior loss in each of the regimes $(0.01, 0.4)$, $(0.5, 0.79)$ and $(0.8, 0.99)$; we report the regime whose H attains the lowest FD-DINOv2. All metrics are taken at the checkpoint with the lowest FD-DINOv2 (selected on a 5k-sample trajectory at $H = 0.5$); FID* is the FID at the checkpoint with the lowest FID. MiFID/FD-DINOv2 is the memorization penalty of MiFID at the same checkpoint: the reciprocal of the mean cosine distance between each sample and its nearest training image in DINOv2 feature space when that mean is below 0.1, and 1 otherwise. Precision, recall and Vendi score use Inception features. A dash marks rows whose best-FID checkpoint has not been evaluated with 50k samples.

		CIFAR-10_LT										
Schedule	Driving noise	H	K	Target	weighting	FID* ↓	FID ↓	FD-DINOv2 ↓	MiFID/FD-DINOv2 ↓	Precision ↑	Recall ↑	Vendi Score ↑
FM	random straight-lines	-	-	$\xi - x_0$	uniform	15.29	15.29	195.42	1.00	0.88	0.40	19.61
FM	random straight-lines	-	-	$\xi - x_0$	variance	17.50	34.95	482.56	1.00	0.67	0.30	18.45
VP	BM	-	-	noise	variance	18.02	33.28	280.96	1.00	0.70	0.61	17.32
VP	BM	-	-	vpred	variance	16.20	23.84	302.28	1.00	0.82	0.36	19.05
Diffusion FM	BM	-	-	noise	variance	15.70	23.33	297.97	1.00	0.80	0.34	19.52
Diffusion FM	BM	-	-	vpred	variance	15.39	27.77	338.73	1.00	0.78	0.35	19.26
$\text{GFI}_{VP(K=1)}$	ε-MA-fBM	0.80	1	vpred	KL-σ	8.60	8.60	119.35	13.99	0.97	0.38	19.24
$\text{GFI}_{VP(K=2)}$	ε-MA-fBM	0.80	2	vpred	KL-σ	8.59	8.59	116.26	14.18	0.97	0.38	19.34
$\text{GFI}^\dagger_{VP(K=3)}$	ε-MA-fBM	0.03	3	vpred	KL-σ	–	19.58	252.53	1.00	0.84	0.36	19.59
$\text{GFI}_{VP(K=4)}$	ε-MA-fBM	0.80	4	vpred	KL-σ	–	18.35	233.66	1.00	0.86	0.36	19.60
$\text{GFI}_{VP(K=5)}$	ε-MA-fBM	0.66	5	vpred	KL-σ	–	34.37	423.23	1.00	0.71	0.30	19.62
$\text{GFI}_{FM(K=1)}$	ε-MA-fBM	0.80	1	vpred	KL-σ	7.83	7.83	111.31	12.38	0.98	0.36	19.65
$\text{GFI}_{FM(K=2)}$	ε-MA-fBM	0.80	2	vpred	KL-σ	7.58	7.58	109.97	13.08	0.98	0.37	19.36
$\text{GFI}_{FM(K=3)}$	ε-MA-fBM	0.32	3	vpred	KL-σ	11.51	11.51	156.52	1.00	0.92	0.38	19.51
$\text{GFI}_{FM(K=4)}$	ε-MA-fBM	0.31	4	vpred	KL-σ	14.15	14.15	183.34	1.00	0.90	0.39	19.50
$\text{GFI}_{FM(K=5)}$	ε-MA-fBM	0.21	5	vpred	KL-σ	–	32.47	422.01	1.00	0.72	0.30	19.23

J.6 ADDITIONAL EXPERIMENTS ON CELEBA WITH DIT-B/2

Table 12: Quantitative results on CelebA-HQ (latent space, DiT-B/2, 400k training steps, 50k samples), where we train GFI over the Hurst range $(0.01, 0.995)$. Every row reports the checkpoint with the lowest in-training FD-DINOv2; FID* is the FID of the checkpoint with the lowest in-training FID. GFI is evaluated at the $H \in (0.01, 0.4)$ in the rough regime with the smallest prior loss, at the $H \in (0.5, 0.79)$ and $H \in (0.8, 0.99)$ in the smooth regimes with the smallest prior loss, and at the H with the smallest path-KL bound; we report the H with the lowest FD-DINOv2.

						CelebA-HQ					
Schedule	Driving noise	H	K	Target	weighting	FID* ↓	FID ↓	FD-DINOv2 ↓	Precision ↑	Recall ↑	Vendi Score ↑
FM	random straight-lines	-	-	$\xi - x_0$	uniform	10.03	10.03	251.31	0.74	0.32	7.39
FM	random straight-lines	-	-	$\xi - x_0$	variance	8.09	9.61	214.38	0.75	0.34	7.48
Diffusion FM	BM	-	-	$\xi - x_0$	uniform	9.84	9.84	218.27	0.78	0.31	7.19
Diffusion FM	BM	-	-	$\xi - x_0$	KL-σ	14.00	14.00	260.62	0.76	0.22	6.83
Diffusion FM	BM	-	-	$\xi - x_0$	variance	8.17	9.88	190.62	0.78	0.33	7.26
Diffusion FM	BM	-	-	noise	variance	8.05	11.64	198.13	0.81	0.31	6.93
Diffusion FM	BM	-	-	vpred	variance	8.25	9.93	192.37	0.79	0.33	7.16
VP	BM	-	-	$\xi - x_0$	uniform	11.85	11.85	250.77	0.76	0.27	6.98
VP	BM	-	-	$\xi - x_0$	KL-σ	13.44	13.44	265.34	0.75	0.23	6.91
VP	BM	-	-	$\xi - x_0$	variance	8.42	10.23	197.44	0.79	0.30	7.12
VP	BM	-	-	noise	variance	8.02	10.87	196.43	0.80	0.31	6.99
VP	BM	-	-	vpred	variance	8.98	9.75	198.14	0.79	0.32	7.14
GFI$_{FM(K=1)}$	ε-MA-fBM	0.38	1	$\xi - x_0$	KL-σ	10.34	10.34	234.40	0.76	0.30	7.16
GFI$_{FM(K=2)}$	ε-MA-fBM	0.06	2	$\xi - x_0$	KL-σ	10.36	10.36	240.05	0.75	0.29	7.21
GFI$_{FM(K=3)}$	ε-MA-fBM	0.82	3	$\xi - x_0$	KL-σ	8.78	10.51	203.33	0.78	0.29	7.32
GFI$_{FM(K=4)}$	ε-MA-fBM	0.04	4	$\xi - x_0$	KL-σ	8.25	10.28	202.71	0.78	0.30	7.29
GFI$_{FM(K=5)}$	ε-MA-fBM	0.92	5	$\xi - x_0$	KL-σ	7.83	9.50	191.54	0.76	0.34	7.52
GFI$_{FM(K=6)}$	ε-MA-fBM	0.92	6	$\xi - x_0$	KL-σ	7.49	9.90	173.96	0.81	0.32	7.21
GFI$_{FM(K=7)}$	ε-MA-fBM	0.92	7	$\xi - x_0$	KL-σ	7.21	9.42	173.16	0.79	0.33	7.28
GFI$_{VP(K=1)}$	ε-MA-fBM	0.40	1	$\xi - x_0$	KL-σ	9.83	14.10	229.24	0.79	0.23	6.77
GFI$_{VP(K=2)}$	ε-MA-fBM	0.46	2	$\xi - x_0$	KL-σ	10.27	10.27	237.36	0.75	0.31	7.16
GFI$_{VP(K=3)}$	ε-MA-fBM	0.03†	3	$\xi - x_0$	KL-σ	8.76	8.76†	245.76†	0.74†	0.36†	7.43†
GFI$_{VP(K=4)}$	ε-MA-fBM	0.73	4	$\xi - x_0$	KL-σ	8.24	9.28	190.17	0.78	0.35	7.30
GFI$_{VP(K=5)}$	ε-MA-fBM	0.66	5	$\xi - x_0$	KL-σ	8.32	9.68	202.16	0.76	0.35	7.40
GFI$_{VP(K=6)}$	ε-MA-fBM	0.61	6	$\xi - x_0$	KL-σ	8.95	8.95	239.98	0.71	0.38	7.78
GFI$_{VP(K=7)}$	ε-MA-fBM	0.02	7	$\xi - x_0$	KL-σ	8.41	9.73	240.32	0.70	0.37	7.84

Table 13: Quantitative results on CelebA-HQ (latent space, DiT-B/2, 600k training steps, 50k samples), where we train GFI over the Hurst range $(0.01, 0.995)$. Every row reports the checkpoint with the lowest in-training FD-DINOv2; FID* is the FID of the checkpoint with the lowest in-training FID. GFI is evaluated at the $H \in (0.01, 0.4)$ in the rough regime with the smallest prior loss and at the $H \in (0.8, 0.99)$ in the smooth regime with the smallest prior loss; we report the H with the lowest FD-DINOv2.

						CelebA-HQ					
Schedule	Driving noise	H	K	Target	weighting	FID* ↓	FID ↓	FD-DINOv2 ↓	Precision ↑	Recall ↑	Vendi Score ↑
FM	random straight-lines	-	-	$\xi - x_0$	variance	7.92	9.24	213.45	0.75	0.34	7.54
Diffusion FM	BM	-	-	$\xi - x_0$	variance	8.02	9.84	192.23	0.78	0.33	7.25
VP	BM	-	-	noise	variance	8.18	10.20	193.93	0.80	0.32	7.07
GFI$_{FM(K=5)}$	ε-MA-fBM	0.92	5	$\xi - x_0$	KL-σ	8.09	9.55	194.53	0.76	0.34	7.55
GFI$_{FM(K=6)}$	ε-MA-fBM	0.92	6	$\xi - x_0$	KL-σ	7.76	11.14	180.19	0.81	0.30	7.11
GFI$_{FM(K=7)}$	ε-MA-fBM	0.92	7	$\xi - x_0$	KL-σ	7.23	8.92	176.63	0.80	0.34	7.33
GFI$_{FM(K=8)}$	ε-MA-fBM	0.93	8	$\xi - x_0$	KL-σ	8.06	8.97	177.20	0.79	0.34	7.33
GFI$_{FM(K=9)}$	ε-MA-fBM	0.94	9	$\xi - x_0$	KL-σ	7.74	10.27	180.03	0.79	0.31	7.23
GFI$_{FM(K=10)}$	ε-MA-fBM	0.95	10	$\xi - x_0$	KL-σ	7.21	10.38	177.46	0.80	0.31	7.21

J.7 ADDITIONAL EXPERIMENTS FOR COMPARISON AGAINST GFDM

Table 14: Quantitative results of GFDM (VP schedule driven by MA-fBM with $K = 3$ and a fixed Hurst index) on CIFAR-10 (DiT-B/2, 400k steps, 50k samples). FID* is the FID at the checkpoint with the lowest FID; all other metrics are taken at the checkpoint with the lowest FD-DINOv2 (both selected on a 5k-sample trajectory). Memory penalty (DINOv2) is the memorization penalty of MiFID at the same checkpoint: the reciprocal of the mean cosine distance between each sample and its nearest training image in DINOv2 feature space when that mean is below 0.1, and 1 otherwise. Precision, recall and Vendi score use Inception features. Rows with a memory penalty greater than 1 are excluded from bolding.

	CIFAR-10										
Schedule	Driving noise	H	K	Target	weighting	FID* ↓	FD-DINOv2 ↓	Mem. penalty (DINOv2) ↓	Precision ↑	Recall ↑	Vendi Score ↑
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	noise	variance	13.63	716.08	1.00	0.58	0.56	15.86
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	noise	variance	14.39	745.12	1.00	0.59	**0.57**	15.88
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	noise	KL-σ	**5.78**	**406.80**	1.00	**0.66**	0.55	17.30
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	noise	KL-σ	6.03	413.88	1.00	0.65	0.56	**17.39**
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	vpred	variance	12.89	705.57	1.00	0.58	0.56	15.83
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	vpred	variance	14.40	738.99	1.00	0.58	0.56	15.69
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	vpred	KL-σ	6.31	422.28	1.00	**0.66**	0.54	17.28
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	vpred	KL-σ	6.54	427.10	1.00	**0.66**	0.54	17.38

Table 15: Quantitative results of GFDM (VP schedule driven by MA-fBM with $K = 3$ and a fixed Hurst index) on CIFAR-10-LT (DiT-B/2, 400k steps, 50k samples). FID* is the FID at the checkpoint with the lowest FID; all other metrics are taken at the checkpoint with the lowest FD-DINOv2 (both selected on a 5k-sample trajectory). Memory penalty (DINOv2) is the memorization penalty of MiFID at the same checkpoint: the reciprocal of the mean cosine distance between each sample and its nearest training image in DINOv2 feature space when that mean is below 0.1, and 1 otherwise. Precision, recall and Vendi score use Inception features. Rows with a memory penalty greater than 1 are excluded from bolding.

	CIFAR-10-LT										
Schedule	Driving noise	H	K	Target	weighting	FID* ↓	FD-DINOv2 ↓	Mem. penalty (DINOv2) ↓	Precision ↑	Recall ↑	Vendi Score ↑
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	noise	variance	31.59	1023.04	1.00	0.50	0.47	15.31
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	noise	variance	32.86	1048.92	1.00	0.51	0.47	14.44
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	noise	KL-σ	11.12	159.17	1.00	0.92	0.37	19.47
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	noise	KL-σ	13.31	178.72	1.00	0.90	0.38	19.74
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	vpred	variance	28.75	987.78	1.00	0.53	**0.48**	15.04
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	vpred	variance	30.19	1015.13	1.00	0.54	0.47	14.32
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	vpred	KL-σ	**9.32**	**140.69**	1.00	**0.94**	0.37	19.20
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	vpred	KL-σ	12.92	173.41	1.00	0.91	0.37	19.91

Table 16: Quantitative results of GFDM (VP schedule driven by MA-fBM with $K = 3$ and a fixed Hurst index) on CelebA-HQ 256 (DiT-B/2, 600k steps, 50k samples). FID* is the FID at the checkpoint with the lowest FID; all other metrics are taken at the checkpoint with the lowest FD-DINOv2 (both selected on a 5k-sample trajectory). Memory penalty (DINOv2) is the memorization penalty of MiFID at the same checkpoint: the reciprocal of the mean cosine distance between each sample and its nearest training image in DINOv2 feature space when that mean is below 0.1, and 1 otherwise. Precision, recall and Vendi score use Inception features. Rows with a memory penalty greater than 1 are excluded from bolding.

	CelebA-HQ 256										
Schedule	Driving noise	H	K	Target	weighting	FID* ↓	FD-DINOv2 ↓	Mem. penalty (DINOv2) ↓	Precision ↑	Recall ↑	Vendi Score ↑
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	noise	variance	11.30	335.99	1.00	0.69	0.33	7.54
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	noise	variance	11.50	345.61	1.00	0.68	0.32	7.67
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	noise	KL-σ	8.23	198.90	1.00	**0.79**	0.32	7.25
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	noise	KL-σ	**8.11**	**198.54**	1.00	**0.79**	0.32	7.19
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	$\xi - x_0$	variance	11.58	344.25	1.00	0.67	**0.35**	7.75
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	$\xi - x_0$	variance	11.47	352.53	1.00	0.67	0.34	**7.84**
GFDM$_{FVP(K=3)}$	MA-fBM	0.7	3	$\xi - x_0$	KL-σ	9.29	206.42	1.00	0.78	0.32	7.22
GFDM$_{FVP(K=3)}$	MA-fBM	0.9	3	$\xi - x_0$	KL-σ	8.84	202.23	1.00	**0.79**	0.30	7.21