Generative Fractional Diffusion Models

Gabriel Nobis [1] **Maximilian Springenberg** [1] **Marco Aversa** [2] **Michael Detzel** [1] **Rembert Daems** [3] **Roderick Murray-Smith** [4] **Shinichi Nakajima** [5] **Sebastian Lapuschkin** [1] **Stefano Ermon** [6] **Tolga Birdal** [7] **Manfred Opper** [5] **Christoph Knochenhauer** [8] **Luis Oala** [2] **Wojciech Samek** [1 5]

Abstract

We introduce the first continuous-time score-based generative model that leverages fractional diffusion processes for its underlying dynamics. Although diffusion models have excelled at capturing data distributions, they still suffer from various limitations such as slow convergence, mode-collapse on imbalanced data, and lack of diversity. These issues are partially linked to the use of light-tailed Brownian motion (BM) with independent increments. In this paper, we replace BM with an approximation of its non-Markovian counterpart, fractional Brownian motion (fBM), characterized by correlated increments and Hurst index $H \in (0, 1)$, where $H = 1/2$ recovers the classical BM. To ensure tractable inference and learning, we employ a recently popularized Markov approximation of fBM (MA-fBM) and derive its reverse time model, resulting in *generative fractional diffusion models* (GFDMs). We characterize the forward dynamics using a continuous reparameterization trick and propose an augmented score matching loss to efficiently learn the score-function, which is partly known in closed form, at minimal added cost. The ability to drive our diffusion model via fBM provides flexibility and control. $H \leq 1/2$ enters the regime of *rough paths* whereas $H > 1/2$ regularizes diffusion paths and invokes long-term memory as well as a heavy-tailed behaviour (super-diffusion). The Markov approximation allows added control by varying the number of Markov processes linearly combined to approximate fBM. Our evaluations on real image datasets demonstrate that GFDM achieves greater pixel-wise diversity and enhanced image quality, as indicated by a lower FID, offering a promising alternative to traditional diffusion models.

1. Introduction

Recent years have witnessed a remarkable leap in generative diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song and Ermon, 2019), celebrated for their ability to accurately learn data distributions and generate high-fidelity samples. These models have made significant impact across a wide spectrum of application domains, including the generation of complex molecular structures (Avdeyev et al., 2023; Hoogeboom et al., 2022) for material (Manica et al., 2023) or drug discovery (Corso et al., 2022), realistic audio samples (Copet et al., 2023; Kreuk et al., 2022), 3D objects (Zeng et al., 2022), medical images (Aversa et al., 2023) and aerospace (Espinosa and Crowley, 2023).

Despite these successes, modern score-based generative models (SBGMs) formulated in continuous time (Song et al., 2021) face limitations due to their reliance on a simplistic driving noise, the Brownian motion (BM) (Brown, 1828; Einstein, 1905; Wiener, 1923). As a light-tailed process, using BM often results in slow convergence rates and susceptibility to mode-collapse, especially with imbalanced data (Yoon et al., 2023). Additionally, it's purely Markovian nature may also make it hard to capture the full complexity and richness of real-world data. All these attracted a number of attempts for involving different noise types (Yoon et al., 2023; Ma et al., 2023). In this paper, we propose leveraging fractional noises, particularly the renowned non-Markovian fractional BM (fBM) (Lévy, 1953; Mandelbrot and Van Ness, 1968) to drive diffusion models. fBM extends BM to stationary increments with a more complex dependence structure, *i.e.*, long-range dependence vs. roughness/regularity controlled by a Hurst index, a measure of "mild" or "wild" randomness (Stinson, 2005). This all come at the expense of computational challenges

[1]Department of Artificial Intelligence, Fraunhofer HHI, Berlin, Germany [2]Dotphoton, Zug, Switzerland [3]Ghent University, Ghent, Belgium [4]University of Glasgow, Glasgow, United Kingdom [5]Technical University of Berlin, Berlin, Germany [6]Stanford University University, Stanford, United States [7]Imperial College London, Faculty of Engineering, Department of Computing, London, United Kingdom [8]Technical University of Munich, Machine Learning Lab for Finance & Insurance, Munich, Germany. Correspondence to: Gabriel Nobis <gabriel.nobis@hhi.fraunhofer.de>.

Accepted by the Structured Probabilistic Inference & Generative Modeling workshop of ICML 2024, Vienna, Austria.

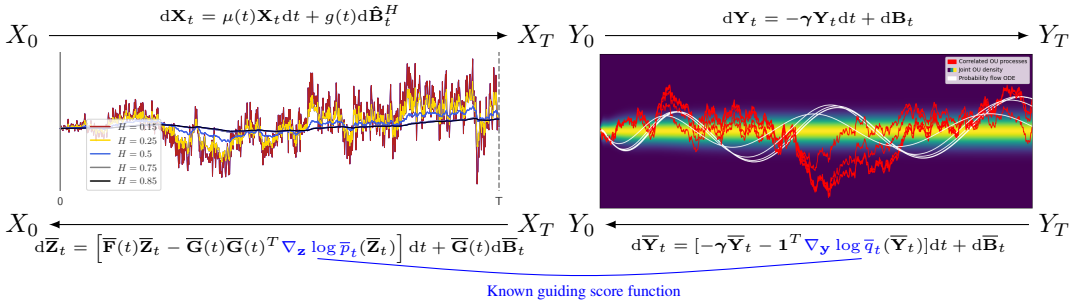


Figure 1: The score function of the augmenting processes is known in closed form and serves as guidance for the unknown score function. A weighted sum of the correlated augmenting processes approximates a driving fractional diffusion process.

and intractability of inference, mostly stemming from its non-Markovian nature. To overcome these limitations, we leverage the recent works in Markov approximations of fBM (MA-fBM) (Daems et al., 2024; Harms and Stefanovits, 2019) and establish a framework for training continuous-time score models using an approximate fractional diffusion process, as well as generating samples from the corresponding tractable reverse process. Notably, our method maintains the same number of score model evaluations during both training and data generation, with only a minimal increase in computational load. Our contributions are:

- We analytically derive the time-reversal of forward dynamics driven by Markov-approximate fractional Brownian motion in a way that the dimensionality of the score matches that of the data.
- We derive an explicit formulae for the marginals of the conditional forward process via a continuous reparameterization trick.
- We introduce a novel augmented score matching loss for learning the score function in our fractional diffusion model, which can be minimized by a score model of data-dimension.

Our experimental evaluation validates our contributions, showing the gains brought by opting for a correlated-noise with long-term memory, approximated by a combination of a number of Markov processes, where the amount of processes further control the diverstiy.

Differentiation from existing work. Yoon et al. (2023) generalizes SBGM from an underlying BM to a driving Lévy process, a stochastic process with independent and stationary increments. A driving noise with correlated increments is not included in the framework of Yoon et al. (2023). Conceptually, every Lévy process is a semimartingale (Protter, 2013) and hence fBM is not a Lévy process. As far as we know, we are the first to build a SBGM with driving noise converging to non-Markovian process of infinite quadratic variation.

The closest work to ours is Tong et al. (2022) constructing a neural-SDE based on correlated noise and using the neural SDE as a forward process of a SBGM. Our framework with exact reverse time model is based on the integral representation of fBM derived in Harms and Stefanovits (2019) and the optimal approximation coefficients of Daems et al. (2024), while the fractional noise in (Tong et al., 2022) is sparsely approximated by a linear combination of independent standard normal random variables without exact reverse time model. Moreover, the framework of Tong et al. (2022) is limited to $H > 1/3$ and only compatible with the Euler-Maruyama sample schema (Cohen and Elliott, 2015) while our framework is up to numerical stability applicable for any $H \in (0, 1)$ and compatible with any suitable SDE or ODE solver. To the best of our knowledge, we are the first to build a SBGM in continuous time converging to a fractional diffusion process.

2. Background

Modeling the distribution transforming process of a SBGM through stochastic differential equations (SDEs) (Song et al., 2021) offers a unifying framework to generate data from an unknown probability distribution. Instead of injecting a finite number of fixed noise scales via a Markov chain, infinitely many noise scales tailored to the continuous dynamics of the Markov process $\mathbf{X} = (\mathbf{X}_t)_{t\in[0,T]}$ are utilized during the distribution transformation, offering considerable practical advantages over discrete time diffusion models (Song et al., 2021). The forward dynamics, transitioning from a data sample $\mathbf{X}_0 \sim p_0$ to a tractable noise sample $\mathbf{X}_T \sim p_T$ are specified by a continuous drift function $\mathbf{f}$ and a continuous diffusion coefficient g. These dynamics define a diffusion process that solves the SDE

$$\mathrm{d}\mathbf{X}_t = \mathbf{f}(\mathbf{X}_t, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{B}_t, \quad \mathbf{X}_0 \sim p_0 \tag{1}$$

driven by a multivariate BM $\mathbf{B}$. To sample data from noise, a reverse time model is needed that defines the backward transformation from the tractable noise distribution to the data distribution. Whenever $\mathbf{X} = (\mathbf{X}_t)_{t\in[0,T]}$ is a stochastic process and g is a function on $[0, T]$, we write $\overline{\mathbf{X}}_t = \mathbf{X}_{T-t}$ for the reverse time model and $\bar{g}(t) = g(T - t)$ for the reverse time function. The marginal density of the stochastic process $\mathbf{X}$ at time t is denoted by p_t throughout this

work[1]. Remarkably, an exact reverse time model to the forward model in eq. (1) is given by the backward dynamics (Stratonovich, 1960; Anderson, 1982; Föllmer, 1986)

$$\mathrm{d}\overline{\mathbf{X}}_t = \left[\overline{\mathbf{f}}(\overline{\mathbf{X}}_t, t) - \bar{g}^2(t)\nabla_{\mathbf{x}} \log \bar{p}_t(\overline{\mathbf{X}}_t)\right] \mathrm{d}t + \bar{g}(t)\mathrm{d}\overline{\mathbf{B}}_t, \tag{2}$$

where the only unknown is the score-function $\nabla_{\mathbf{x}} \log p_t$, inheriting the intractability from the unknown initial distribution p_0. In addition to the stochastic dynamics, the reverse time model provides deterministic backward dynamics via an ordinary differential equation (ODE) by the so called probability flow ODE (PF ODE) (Song et al., 2021)

$$\mathrm{d}\bar{\mathbf{x}}_t = \left[\overline{\mathbf{f}}(\bar{\mathbf{x}}_t, t) - \frac{1}{2}\bar{g}^2(t)\nabla_{\mathbf{x}} \log \bar{p}_t(\bar{\mathbf{x}}_t, t)\right] \mathrm{d}t. \tag{3}$$

Stochasticity is only injected into the system through the random initialization $\mathbf{x}_T \sim p_T$, implying a deterministic and bijective map from noise to data (Song et al., 2021). Conditioning the forward process on a data sample $\mathbf{x}_0 \sim p_0$ results for linear $\mathbf{f}(\cdot, t)$ in a tractable Gaussian forward process with conditional score function $\nabla_{\mathbf{x}} \log p_{0t}(\mathbf{x}|\mathbf{x}_0)$ in closed form. To approximate the exact reverse time model, this tractable score function is used to train a time-dependent score-model $S_{\boldsymbol{\theta}}$ via score matching (Hyvärinen, 2005; Song et al., 2019). Upon training, any solver for SDEs or ODEs can be utilized to generate data from noise by simulating the stochastic or deterministic backward dynamics of the reverse time model with $S_{\boldsymbol{\theta}} \approx \nabla_{\mathbf{x}} \log p$.

Simulation error of the reverse time model. The two main sources of error when simulating the reverse time model are the approximation error due to $S_{\boldsymbol{\theta}}$ only approximating $\nabla_{\mathbf{x}} \log p$, and the discretization error, which arises from transitioning from continuous time to discrete steps. Simulating the PF ODE with the Euler method over $N \in \mathbb{N}$ equidistant time steps results in a global error of order N^{-1} (Bayer et al., 2021). In contrast, the expected global error for simulating the SDE using the Euler-Maruyama method is of a lower order $N^{-\frac{1}{2}}$, indicating a larger error for the same number of steps (Cohen and Elliott, 2015; Bayer et al., 2021). From this perspective it is reasonable that sampling from the PF ODE requires fewer steps. Yet, the source of qualitative differences between sampling from the ODE and the SDE (Song et al., 2021) remains unclear.

A pathwise perspective on sampling. The roughness of a path can be measured by its Hölder exponent $0 < \delta \leq 1$ (Lyons, 1998). For example, BM as the integrator in the backward dynamics eq. (2) has δ-Hölder continuous paths for any $0 < \delta < \frac{1}{2}$, whereas the integrator $t \mapsto t$ of the PF ODE eq. (3) can be regarded as a Hölder continuous path with exponent $\delta = 1$. Therefore, from a pathwise perspective, we move away from a rough path when we sample using the PF ODE. An unexplored topic in SGBMs is the interpolation between the SDE and the PF ODE in terms of the Hölder exponent. It remains to be examined whether there is, to some extent, an optimal degree of Hölder continuity in between, or if an even rougher path with $\delta << \frac{1}{2}$ could yield an advantageous data generator.

The process that naturally arises from this line of thought is fractional BM with Hurst index $H \in (0, 1)$, where almost all paths are Hölder continuous for any exponent $\delta < H$, controlled by H. In terms of roughness, the Hurst index interpolates between the paths of Brownian-driven SDEs and those of the underlying integration in PF ODEs, while also offering the potential for even rougher paths. Motivated by these observations we define a new SBGM converging to a fractional diffusion process, generalizing SBGMs from an underlying BM (Brown, 1828; Einstein, 1905; Wiener, 1923) to a fractional BM (Lévy, 1953; Mandelbrot and Van Ness, 1968).

3. Fractional driving noise

Before describing the challenges in defining a score-based generative model with control over the roughness of the distribution-transforming path, we introduce fractional Brownian motion (fBM). The literature distinguishes between "Type I" fBM and "Type II" fBM (Davidson and Hashimzade, 2009) having stationary and non-stationary increments, respectively. The type II fBM, also called Riemann-Liouville fBM, possesses smaller deviations from its mean, potentially an advantageous property for a driving noise of a score-based generative model, since large deviations of the sampling process to the data mean can lead to sample artifacts (Lou and Ermon, 2023). Here and in the experiments we focus on type II fBM. However, our theoretical framework generalizes to both types as detailed in Appendix A. The empirical study of a score-based generative model with driving noise converging to type I fBM is dedicated to future work. We begin with the definition of Riemann-Liouville fBM (Lévy, 1953), a generalization of BM permitting correlated increments.

Definition 3.1 (Type II Fractional Brownian Motion (Lévy, 1953))**.** *Let $B = (B_t)_{t\geq 0}$ be a standard BM and Γ the Gamma function. The centered Gaussian process*

$$B_t^H = \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t - s)^{H-\frac{1}{2}} \mathrm{d}B_s, \quad t \geq 0, \tag{4}$$

uniquely characterized in law by its covariances

$$\frac{1}{\Gamma^2(H + 1/2)} \int_0^{\min\{t,s\}} ((t - u)(s - u))^{H-\frac{1}{2}} \mathrm{d}u, \tag{5}$$

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

BM being the unique continuous and centered Gaussian process with covariance $\min\{t, s\}$ is recovered for $H =$

[1] See Appendix G for the notational conventions of this work.

$1/2$, since $\Gamma(1) = 1$. In comparison to the purely Brownian setting of $H = \frac{1}{2}$ with independent increments (diffusion), the path of B^H becomes more smooth for $H > \frac{1}{2}$ due to positively correlated increments (super-diffusion) and more rough for $H < \frac{1}{2}$ due to negatively correlated increments (sub-diffusion). These three regimes are reflected in the Hölder exponent of $\delta < H$ for almost all paths.

Generalization challenges. The most challenging part in defining a score-based generative model driven by fBM is the derivation of a reverse time model. Due to its covariance structure, fBM is not a Markov process (Huy, 2003) and the shift in the roughness of the sample path leads to changes in its quadratic variation: from t in the purely Brownian (diffusion) regime to zero in the smooth regime, and to infinite in the rough regime (Cohen and Elliott, 2015). For that reason fBM is neither a Markov process nor a semimartingale (Biagini et al., 2008) for all $H \neq \frac{1}{2}$. Hence, we cannot make use of the Markov property or the Kolmogorov equations (Fokker-Planck) that are used to derive the reverse time model of BM-driven SDEs (Stratonovich, 1960; Anderson, 1982; Föllmer, 1986). See Appendix F for a more detailed illustration of the problem. The existence of a reverse time model can be proven in the smooth regime of fBM (Darses and Saussereau, 2007). However, due to the absence of an explicit score function in Darses and Saussereau (2007) it does not provide a sufficient structure to train a SBGM.

To overcome this difficulty we follow (Daems et al., 2024; Harms and Stefanovits, 2019) and define the driving noise of our generative model by a linear combination of Markovian semimartingales converging to fBM. The approximation is based on the exact infinite-dimensional Markovian representation of fBM given in Theorem A.2.

Definition 3.2 (Markov approximation of fBM (Daems et al., 2024; Harms and Stefanovits, 2019))**.** *Choose $K \in \mathbb{N}$ Ornstein–Uhlenbeck (OU) processes with speeds of mean reversion $\gamma_1, ..., \gamma_K$ and dynamics $\mathrm{d}Y_t^k = -\gamma_k Y_t^k \mathrm{d}t + \mathrm{d}B_t$:*

$$Y_t^k = \int_0^t e^{-\gamma_k(t-s)} \mathrm{d}B_s, \quad k \in \mathbb{N}, \quad t \geq 0, \tag{6}$$

Given a Hurst index $H \in (0,1)$ and a geometrically spaced grid $\gamma_k = r^{k-n}$ with $r > 1$ and $n = \frac{K+1}{2}$ we call the process

$$\hat{B}_t^H := \sum_{k=1}^K \omega_k Y_t^k, \quad H \in (0,1), \quad t \geq 0, \tag{7}$$

Markov-approximate fractional Brownian motion (MA-fBM) with approximation coefficients $\omega_1, ..., \omega_K \in \mathbb{R}$ and denote by $\hat{\mathbf{B}}^H = (\hat{B}_1^H, ..., \hat{B}_D^H)$ the corresponding D-dimensional process where $\hat{B}_i^H$ and $\hat{B}_j^H$ are independent for $i \neq j$ inheriting independence from the underlying standard BMs B_i and B_j.

While Harms (2019) defines the above approximation coefficient aiming for strong convergence rates of high polynomial order in K, we follow the approach of Daems et al. (2024) to choose the $L^2(\mathbb{P})$ optimal approximation coefficients for a given K.

Proposition 3.3 (Optimal Approximation Coefficients (Daems et al., 2024))**.** *The optimal approximation coefficients $\boldsymbol{\omega} = (\omega_1, ..., \omega_K) \in \mathbb{R}^K$ for a given Hurst index $H \in (0,1)$, a terminal time $T > 0$ and a fixed geometrically spaced grid to minimize the $L^2(\mathbb{P})$-error*

$$\mathcal{E}(\boldsymbol{\omega}) := \int_0^T \mathbb{E}\left[\left(B_t^H - \hat{B}_t^H\right)^2\right] \mathrm{d}t \tag{8}$$

are given by the closed-form expression $\boldsymbol{A}\boldsymbol{\omega} = \boldsymbol{b}$ with

$$\begin{aligned} \boldsymbol{A}_{i,j} &:= \frac{2T + \frac{e^{-(\gamma_i+\gamma_j)T}-1}{\gamma_i+\gamma_j}}{\gamma_i + \gamma_j}, \\ \boldsymbol{b}_k &:= \frac{T}{\gamma_k^{H+1/2}} P(H+1/2, \gamma_k T) - \frac{H+1/2}{\gamma_k^{H+3/2}, \gamma_k T} \end{aligned} \tag{9}$$

and where $P(z,x) = \frac{1}{\Gamma(z)} \int_0^x t^{z-1} e^{-t} \mathrm{d}t$ is the regularized lower incomplete gamma function.

MA-fBM serves as the driving noise of our generative model, replacing BM in the distribution transforming process solving eq. (1). See Figure 1 for an illustration of the underlying processes.

4. A score-based generative model based on fractional noise

In this section, we define a continuous-time SBGM driven by MA-fBM. A detailed treatment of the theory can be found in Appendix A. We begin with the forward dynamics, transitioning data to noise.

Definition 4.1 (Forward process)**.** *Let $\hat{\mathbf{B}}^H$ be a D-dimensional MA-fBM with Hurst index $H \in (0,1)$. For continuous functions $\mu : [0,T] \to \mathbb{R}$ and $g : [0,T] \to \mathbb{R}$ we define the forward process $\mathbf{X} = (\mathbf{X}_t)_{t\in[0,T]}$ of a generative fractional diffusion model (GFDM) by*

$$\mathrm{d}\mathbf{X}_t = \mu(t)\mathbf{X}_t \mathrm{d}t + g(t)\mathrm{d}\hat{\mathbf{B}}_t^H, \quad \mathbf{X}_0 = \boldsymbol{x}_0 \sim p_0, \tag{10}$$

where p_0 is the unknown data distribution from which we aim to sample from.

The forward process converges to a fractional diffusion process (Daems et al., 2024), and yields empirically good results even for a small number of OU processes (Daems et al., 2024). Considering both the forward process as well as the OU processes defining the driving noise $\hat{\mathbf{B}}^H$, we have for every data dimension an augmented vector of correlated processes $(X, Y^1, \ldots, Y^K)$, driven by the same BM, approximating the time-correlated behavior of

a one-dimensional forward process driven by fBM. We denote the stacked process of the D augmented vectors as $\mathbf{Z} \equiv (\mathbf{X}, \mathbf{Y}^1, \ldots, \mathbf{Y}^K) = (\mathbf{Z}_t)_{t\in[0,T]}$ and refer to the resulting $D(K+1)$-dimensional process as the *augmented forward process*. Rewriting the dynamics of the forward process we observe that the augmented forward process $\mathbf{Z}$ solves a linear SDE

$$\mathrm{d}\mathbf{Z}_t = \mathbf{F}(t)\mathbf{Z}_t\mathrm{d}t + \mathbf{G}(t)\mathrm{d}\mathbf{B}_t, \quad t \in [0,T], \tag{11}$$

where $\mathbf{F}$ and $\mathbf{G}$ are the matrix valued functions defined in Appendix A.2. Hence, $\mathbf{Z}|\mathbf{x}_0$, the augmented forward process conditioned on a data sample $\mathbf{x}_0 \sim p_0$, is a linear transformation of BM. Thus $\mathbf{Z}|\mathbf{x}_0$ is a Gaussian process and so is $\mathbf{X}|\mathbf{x}_0$ (Särkkä and Solin, 2019). To efficiently sample for every $t \in (0,T]$ from the conditional augmented forward distribution during training, we characterize its marginal statistics.

Derivation of marginal statistics. The marginal mean $\mathbb{E}[\mathbf{X}_t|\mathbf{x}_0] = \mathbf{x}_0 \exp(\int_0^t \mu(s)\mathrm{d}s)$ of the conditional forward process is not affected by changing the driving noise to MA-fBM. Since the integral with respect to BM has zero mean, the mean vector of the augmenting OU processes is zero. Additionally, Itô isometry provides a complete characterization of their covariances. See Appendix A.2 for a detailed derivation of the marginal statistics of the augmenting processes. The missing components in the conditional covariance matrix $\mathbf{\Sigma}_t$ of the augmented forward process are the conditional marginal variance of the forward process and the conditional marginal correlation between the forward process and the augmenting processes. In compliance with the reparameterization trick $\mathbf{X}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$ used in discrete time (Ho et al., 2020), we derive by reparameteriziation an explicit formula for the marginal variance of the conditional forward process. This generalizes the explicit formula for the perturbation kernel $p_{0t}(\mathbf{x}|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}; c(t)\mathbf{x}_0, c^2(t)\sigma^2(t)\mathbf{I}_D)$ given in Karras et al. (2022).

Proposition 4.2 (Continuous Reparameterization Trick)**.** *The forward process $\mathbf{X}$ of GFDM conditioned on $\mathbf{x}_0 \in \mathbb{R}^d$ admits the continuous reparameterization*

$$\mathbf{X}_t = c(t)\left(\mathbf{x}_0 + \int_0^t \alpha(t,s)\mathrm{d}\mathbf{B}_s\right) \tag{12}$$

such that $\mathbf{X}_t \sim \mathcal{N}(c(t)\mathbf{x}_0, c^2(t)\sigma^2(t)\mathbf{I}_d)$ with $c(t) = \exp\left(\int_0^t \mu(s)\mathrm{d}s\right)$ and $\sigma^2(t) = \int_0^t \alpha^2(t,s)\mathrm{d}s$ where α is given by

$$\alpha(t,s) = \sum_{k=1}^K \omega_k \left[\frac{g(s)}{c(s)} - \gamma_k \int_s^t f_k(u,s)\mathrm{d}u\right]. \tag{13}$$

Sketch of Proof. Reparameterization of the forward dynamics in Equation (10) and the Stochastic Fubini Theorem yields the Gaussian process $\mathbf{X}_t = c(t)(\mathbf{x}_0 + \int_0^t \alpha(t,s)\mathrm{d}\mathbf{B}_s)$ with variance $\mathbb{V}[\mathbf{X}_t] = c^2(t)\int_0^t \alpha^2(t,s)\mathrm{d}s$ by Itô isometry. See Proposition A.3 for the proof. □

For $K=1$, $\gamma_1 = 0$ and $\omega_1 = 1$ we retrieve by the above definition of α the perturbation kernel of the purely Brownian setting given in Karras et al. (2022, Equation 12). When, depending on the choice of forward dynamics, $\int_0^t \alpha(t,s)\mathrm{d}s$ is not accessible in closed form, $\mathbf{\Sigma}_t$ can be described by an ODE and solved numerically (Särkkä and Solin, 2019) as described in Appendix B. Thus our method admits any choice of forward dynamics in terms of μ and g.

Explicit fractional forward dynamics. Although our framework is not bound to any specific dynamics, this work's empirical evaluation focuses on *Fractional Variance Exploding* (FVE) dynamics given by

$$\mathrm{d}\mathbf{X}_t = \sigma_{min}\left(\frac{\sigma_{max}}{\sigma_{min}}\right)^t \sqrt{2\log\frac{\sigma_{max}}{\sigma_{min}}}\mathrm{d}\hat{\mathbf{B}}_t^H, \tag{14}$$

with $(\sigma_{min}, \sigma_{max}) = (0.01, 50)$ and *Fractional Variance Preserving* (FVP) dynamics given by

$$\mathrm{d}\mathbf{X}_t = -\frac{1}{2}\beta(t)\mathbf{X}_t\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}\hat{\mathbf{B}}_t^H, \tag{15}$$

with $\beta(t) = \beta(t) = \bar{\beta}_{\min} + t\left(\bar{\beta}_{max} - \bar{\beta}_{min}\right)$ and $(\bar{\beta}_{min}, \bar{\beta}_{max}) = (0.1, 20)$ (Song et al., 2021). Leveraging the continuous reparameterization trick we derive in Appendix B the conditional marginal covariance matrix of FVE in closed form. The integral $\int_s^t f_k(u,s)\mathrm{d}u$ in the setting of FVP is to the best of our knowledge not accessible in closed form and we deploy a numerical ODE solver to estimate the same quantity corresponding to FVP dynamics. See Appendix B for details on the computation of the marginal variances and Figure 4 for an illustration of the resulting variance schedules.

The reverse time model. We observe that the augmented forward dynamics of GFDM are already encompassed in the general framework presented in Song et al. (2021, Appendix A), although they differ from the Variance Exploding (VE), Variance Preserving (VP), and sub-VP dynamics discussed therein. To simplify notation, we use p_t here to denote the marginal density of both $\mathbf{Z}_t$ and $\mathbf{X}_t$. The specific density referred to will be clear from the context. By the significant results of (Stratonovich, 1960; Anderson, 1982; Föllmer, 1986), the reverse time model of GFDM is given by the backward dynamics

$$\mathrm{d}\overline{\mathbf{Z}}_t = \left[\overline{\mathbf{F}}(t)\overline{\mathbf{Z}}_t - \overline{\mathbf{G}}(t)\overline{\mathbf{G}}(t)^T\nabla_{\mathbf{z}}\log\overline{p}_t(\overline{\mathbf{Z}}_t)\right]\mathrm{d}t + \overline{\mathbf{G}}(t)\mathrm{d}\overline{\mathbf{B}}_t. \tag{16}$$

A direct application of (Song et al., 2021) would require to train a score-model with input and output dimension of

$D(K+1)$. However, we show that a model with dimension D is sufficient to learn the augmented score function, informed by the available score of the augmented processes. We further show that our method requires the same number of score model evaluations during training and data generation, incurring only minimal increase of the computational load, due to the simulation of the additional processes during training and sampling without increasing the necessary score model evaluation.

Augmented score matching. We condition the score function $\nabla_{\mathbf{z}} \log p_t$ on a data sample $\mathbf{x}_0 \sim p_0$ *and additionally* on the states of the stacked vector $\mathbf{Y}_t^{[K]} := (\mathbf{Y}_t^1, ..., \mathbf{Y}_t^K)$ of augmenting processes. To train our time-dependent score-model $s_{\boldsymbol{\theta}}$ we propose the *augmented score matching loss*

$$\mathcal{L}(\boldsymbol{\theta}) := \mathbb{E}_t \left\{ \mathbb{E}_{(\mathbf{X}_0, \mathbf{Y}_t^{[K]})} \mathbb{E}_{(\mathbf{X}_t | \mathbf{Y}_t^{[K]}, \mathbf{X}_0)} \left[L_{\boldsymbol{\theta}} \left(\mathbf{X}_t, \mathbf{Y}_t^{[K]}, \mathbf{X}_0 \right) \right] \right\} \tag{17}$$

where $L_{\boldsymbol{\theta}} \left(\mathbf{X}_t, \mathbf{Y}_t^{[K]}, \mathbf{X}_0 \right)$ is given by

$$\| s_{\boldsymbol{\theta}}(\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k, t) - \nabla_{\mathbf{x}} \log p_{0t}(\mathbf{X}_t | \mathbf{Y}_t^{[K]}, \mathbf{X}_0) \|_2^2. \tag{18}$$

The weights $\eta_t^1, ..., \eta_t^K$ arise from conditioning $\mathbf{Z}_t$ on $\mathbf{Y}_t^{[K]}$ and the time points t are uniformly sampled from $\mathcal{U}[0, T]$. We show in the following that the optimal $s_{\boldsymbol{\theta}}$ w.r.t. the *augmented score matching loss* is the L^2-optimal approximation of the score function of our reverse time model.

Proposition 4.3 (Optimal Score-Model)**.** *Assume that $s_{\boldsymbol{\theta}}$ is optimal w.r.t. the augmented score matching loss $\mathcal{L}$. The score-model defined in Proposition A.4 yields the optimal $L^2(\mathbb{P})$ approximation of $\nabla_{\mathbf{z}} \log p_t(\mathbf{Z}_t)$ via*

$$S_{\boldsymbol{\theta}}(\mathbf{Z}_t, t) + \nabla_{\mathbf{z}} \log q_t(\mathbf{Y}_t^{[K]}) \approx \nabla_{\mathbf{z}} \log p_t(\mathbf{Z}_t). \tag{19}$$

Sketch of Proof. Using the relation $\nabla_{\mathbf{x}} \log p_{0t} = -\eta_t^k \nabla_{\mathbf{y}^k} \log p_{0t}$ and the independence of $\mathbf{X}_0$ and $\mathbf{Y}_t^{[K]}$ yields the claim. See Appendix A.3 for the proof. □

Remark 4.4. We show here that it suffices to approximate a D-dimensional score to reverse the $D(K+1)$-dimensional MA-fBM driven SDE with unknown starting distribution.

Sampling from reverse time model. Once we trained our score model $S_{\boldsymbol{\theta}}$ via augmented score matching, we simulate the reverse time model backward in time and sample from the reverse time model via the SDE

$$\begin{aligned} \mathrm{d}\overline{\mathbf{Z}}_t = \{ \overline{\mathbf{F}}(t)\overline{\mathbf{Z}}_t - \overline{\mathbf{G}}(t)\overline{\mathbf{G}}(t)^T [\overline{S}_{\boldsymbol{\theta}}(\overline{\mathbf{Z}}_t, t) + \nabla_{\mathbf{z}} \log \overline{q}_t(\overline{\mathbf{Y}}_t^{[K]})] \} \mathrm{d}t \\ + \overline{\mathbf{G}}(t) \mathrm{d}\overline{\mathbf{B}}_t, \end{aligned} \tag{20}$$

or the corresponding augmented *probability flow ODE* (PF ODE) (Song et al., 2021) by

$$\mathrm{d}\overline{\mathbf{z}}_t = \{ \overline{\mathbf{F}}(t)\overline{\mathbf{z}}_t - \frac{1}{2} \overline{\mathbf{G}}(t)\overline{\mathbf{G}}(t)^T [\overline{S}_{\boldsymbol{\theta}}(\overline{\mathbf{z}}_t, t) + \nabla_{\mathbf{z}} \log \overline{q}_t(\overline{\mathbf{y}}_t^{[K]})] \} \mathrm{d}t \tag{21}$$

where we initialize in both cases the reverse dynamics with the centered (non-isotropic) Gaussian $\overline{\mathbf{Z}}_0$ with covariance matrix $\boldsymbol{\Sigma}_T$. To traverse backward from noise to data, we may deploy any suitable SDE or ODE solver. The PF ODE enables in addition negative log-likelihoods (NLLs) estimation of test data under the learned density (Song et al., 2021). See Appendix E for the computation details of NLLs.

5. Experiments

We conduct two rounds of experiments. First, we train a on CIFAR and MNIST without using an exponential moving average (EMA) (Song et al., 2021) to ensure the fairest possible comparison between different SDE dynamics and avoid favoring one dynamic over another. Secondly we use the most promising configurations to train with EMA on CIFAR to achieve good qualitative results. We evaluate GFDM on three different axis: image quality, test distribution coverage and pixel-wise diversity of the generated data. To measure the quality of generated images we use the two most common metrics: the Frechét-Inception Distance (FID) (Heusel et al., 2017) and the Inception Score (IS) (Salimans et al., 2016). To estimate the log-likelihood of test data under the learned density we calculate NLLs according to Appendix E. The pixel-wise diversity is measured by the pixel Vendi Score VS_p (Friedman and Dieng, 2022) and the minimal VS_p per class denoted by VS_p^{min}. To clarify our terminology, we refer to the (visual) effect of a higher VS_p or VS_p^{min} as "pixel-wise diversity" to distinguish it from the general term "diversity", thus trying to preventing a misunderstanding or mix-up. In line with (Song et al., 2021) we observe that choosing a good $\epsilon > 0$ for the training and sampling time interval $[\epsilon, T]$ for the forward process makes a difference for the measured scores. See Appendix C for the implementation details and the full quantitative results. We begin with the empirical evaluation of how the augmenting processes affect performance.

Effect of augmentation on MNIST. We fix $H = 0.5$ throughout this experiment such that the purely BM is approximated instead of fBM by the weighted sum of the augmenting processes. We observe an increase of the pixel-wise diversity for both FVE and FVP dynamics, with increasing K. In Table 1 we see that the minimal pixel-wise diversity observed per class VS_p^{min} goes from 6.46 to 13.49 for FVE and from 6.25 to 9.23 for FVP. In terms of quality, we see for FVE that $K \in \{2, 3\}$ yields better FID compared to VE and we observe a slight quality degradation for FVP and increasing K. To summarize, the number of augmenting processes enhances pixel-wise diversity on MNIST. However, this comes at the cost of a reduced likelihood of test data under the learned density, indicated by a higher NLLs for more augmenting processes.

Effect of varying Hurst index on MNIST. By varying the Hurst index we observe in Table 2a that $H > 0.5$ with FVP

FVE($H = 0.5$)	FID ↓	NLLs Test ↓	VS_p ↑	VS_p^{min} ↑
VE (retrained)	10.82	2.73	24.20	6.46
$K = 1$	10.30	**2.55**	24.22	6.38
$K = 2$	9.89	3.03	24.15	6.52
$K = 3$	**9.74**	2.93	24.42	6.92
$K = 4$	11.25	3.10	**24.54**	9.81
$K = 5$	25.51	3.94	23.08	**13.49**

FVP($H = 0.5$)	FID ↓	NLLs Test ↓	VS_p ↑	VS_p^{min} ↑
VP (retrained)	**1.44**	**2.38**	23.64	6.25
$K = 1$	2.81	3.90	23.69	6.17
$K = 2$	2.92	4.57	23.63	6.35
$K = 3$	3.51	7.02	23.78	6.35
$K = 4$	1.86	5.71	24.50	6.95
$K = 5$	4.89	7.09	**24.56**	**9.23**

Table 1: Effects of augmenting processes to FVE and FVP dynamics trained on the MNIST dataset.

dynamics clearly performs better in terms of FID achieving a SOTA FID of 0.72 for FVP with $K = 3$ and $H = 0.9$. We conjecture that this is due to the long-term memory regime, smoothing the sample paths, making the dynamics easier to learn. The augmenting processes increase the pixel-wise diversity in terms of VS_p and VS_p^{min} as well for $H \in \{0.9, 0.7\}$ compared to the original VP dynamics, again at the cost of a higher NLLs for more augmenting processes. We conjecture that the increase in VS_p^{min} is a consequence of the heavy-tailed behaviour, which may further enhance robustness to data imbalance, similar to the findings of Yoon et al. (2023).

Effect of augmentation on CIFAR10. For FVE dynamics we empirically observe in Table 6 that $K = 1$ with $H = 0.1$ and $K = 2$ with $H = 0.9$ slightly increases the performance in terms of FID compared to the original VE dynmaics. For FVP dynamics we observe better performance in terms of quality and pixel-wise diversity across all considered Hurst indices $H \in \{0.9, 0.7, 0.1\}$ in Table 2b.

Qualitative results on CIFAR10. Training with EMA on CIFAR10 according to Appendix C we observe a slightly higher pixel-wise diversity score across all H comparing FVP to the original VP dynamics in Table 3a. For FVE the best pixel-wise diversity scores are achieved for $K = 1$ and $H = 0.7$. In line with out results on the MNIST dataset we achieve the best quality results in the sub-diffusion regime of fBM with a FID of 5.52 for $H = 0.9$ and $K = 2$. Remarkably, when we sample with the PF ODE for FVP dynamics GFDM achieves $\text{VS}_p = 4.55$ and a $\text{VS}_p^{min} = 3.10$ for $H = 0.7$ and $K = 3$. The quantitatively superior VS_p index is reflected also in the perceptual quality in Figure 2.

6. Related work

Diffusion models in continuous time. The seminal work of Song et al. (2021) offers a unifying framework modeling the distribution transforming process by a stochastic processes in continuous time with exact reverse time model. Extensive research has been carried out to examine (Karras et al., 2022; Chen et al., 2023; Singhal et al., 2023) and extend (Lou and Ermon, 2023; Jing et al., 2022; Kim et al., 2022; Huang et al., 2022; Bunne et al., 2023; Song et al., 2023) the continuous time view on generative models through the lens of SDEs, including deterministic corruptions (Daras et al., 2023) and blurring diffusion (Hoogeboom and Salimans, 2023). While critic on this view question the usefulness of the theoretical superstructure (Bansal et al., 2023), others extend in line with our work the theoretical framework to new types of underlying diffusion processes (Rissanen et al., 2023). Conceptually similar to our work,Yoon et al. (2023) generalizes the score-based generative model from an underlying Brownian motion to a driving Lévy process, thereby dropping the Gaussian assumptions on the increments. In contrast to our work, the framework of Yoon et al. (2023) does not include correlated increments. Importantly, every Lévy process is a semimartingale, which means that fBM is not a Lévy process.

Fractional noises in machine learning. Recently, Hayashi and Nakagawa (2022) considered neural-SDEs driven by

FVP	$H = 0.9$				$H = 0.7$			
	FID ↓	NLLs Test ↓	VS_p ↑	VS_p^{min} ↑	FID ↓	NLLs Test ↓	VS_p ↑	VS_p^{min} ↑
$K = 2$	1.93	**2.24**	24.00	6.21	2.30	2.95	23.82	6.22
$K = 3$	**0.72**	4.58	24.18	6.53	2.67	5.77	23.96	6.40
$K = 4$	1.22	4.09	24.76	7.31	0.86	4.85	24.39	6.71
$K = 5$	2.17	5.38	**25.15**	8.67	1.36	5.13	24.63	7.89

(a)

FVP dynamics	FID ↓	IS ↑	VS_p ↑	VS_p^{min} ↑
VP (retrained)	17.29	8.74	2.24	1.8
$K = 3, H = 0.9$	12.63	8.34	**3.23**	**2.42**
$K = 3, H = 0.7$	**8.72**	8.86	3.11	2.32
$K = 3, H = 0.5$	9.94	**8.87**	3.03	2.24
$K = 3, H = 0.1$	9.51	8.54	3.01	2.27

(b)

Table 2: (a) Quantitative results for FVP dynamics in the smooth sub-diffusion regime $H > 0.5$ on the MNIST dataset. (b) Quantitative results for FVP dynamics and varying Hurst index on CIFAR10.

Euler-Maruyama SDE	FID ↓	IS ↑	NLLs Test ↓	VS_p ↑	VS_p^{min} ↑
VE (retrained)	6.70	**9.71**	3.75	3.98	2.36
VP (retrained)	11.74	9.13	3.56	2.50	1.93
FVE($K = 1, H = 0.7$)	7.93	8.91	3.78	**4.06**	**2.4**
FVP($K = 2, H = 0.7$)	6.51	9.46	**3.37**	3.00	2.22
FVP($K = 2, H = 0.9$)	**5.52**	9.50	3.43	3.07	2.21
FVP($K = 3, H = 0.7$)	7.10	9.07	3.91	3.18	2.39

(a) Sampled from the SDE.

Euler PF ODE	FID ↓	IS ↑	VS_p ↑	VS_p^{min} ↑
VE (retrained)	**7.23**	9.39	2.97	2.32
VP (retrained)	11.10	8.98	2.7	2.09
FVE($K = 1, H = 0.7$)	8.04	8.77	3.15	2.4
FVP($K = 2, H = 0.7$)	17.74	9.51	3.89	2.60
FVP($K = 2, H = 0.9$)	16.38	9.36	4.10	2.79
FVP($K = 3, H = 0.7$)	26.76	**9.51**	**4.55**	**3.10**

(b) Sampled from the PF ODE.

Table 3: Quantitative results on CIFAR10 for different Hurst indices compared to the purely Brownian dynamics.

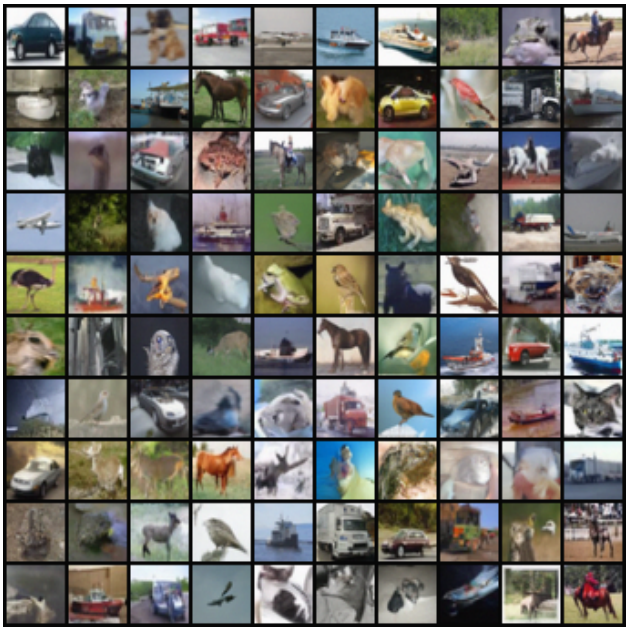
fractional noise. Yet they do not study diffusion models. The closest work to our work, Tong et al. (2022) approximated the type-II fBM with sparse Gaussian processes constructing a neural SDE as a forward process of a score-based generative model, without exact reverse time model. Unfortunately, they are also limited to Euler-Maruyama solvers and to the case of $H > 1/3$, while our framework is up to numerical stability applicable for any $H \in (0, 1)$ and compatible with any suitable SDE or ODE solver. Daems *et al.* (Daems et al., 2024), who inspired our Markov-approximate noise, includes a more elaborate discussion as well as a variational inference framework for MA-fBM.

Rough path theory. The pathwise analysis of SDEs driven by processes with a Hölder exponent less than $1/2$, including fBM for $H < 1/2$ and BM, is encompassed by rough path theory (Lyons, 1998). Rough path theory is applied in machine learning (i) to derive stability bounds for the trained weights of a residual neural network (Bayer et al., 2023), (ii) for rough control of neural ODEs (Kidger, 2021), and (iii) to model long time series behavior via neural rough differential equations (Liao et al., 2019; Morrill et al., 2021). In finance the famous Black-Scholes model (Black and Scholes, 1973) is driven by BM, while more recent continuous-time models employ fractional noise to model price processes (Czichowsky et al., 2018; Guasoni et al., 2019) or rough volatility (Bayer et al., 2016; Gatheral et al., 2022).

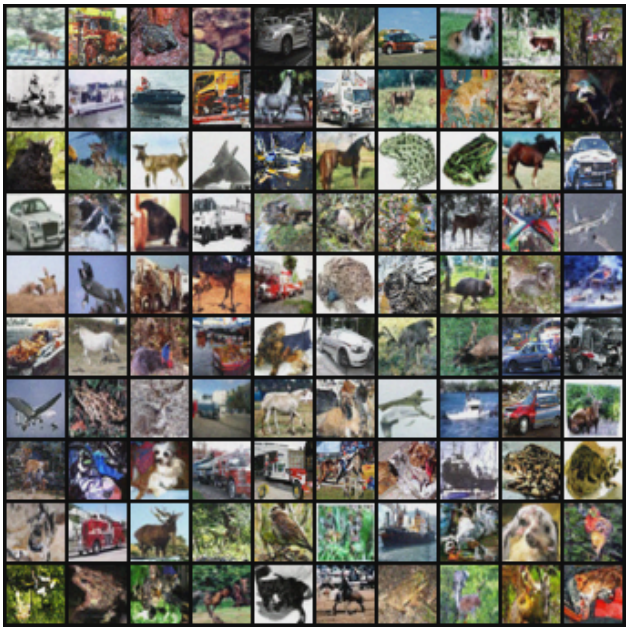
7. Conclusion

In this work, we determined the extent to which the continuous time framework of score-based generative models can be generalized to an underlying fBM, introducing a novel generative model driven by MA-fBM with control over the roughness of distribution transformation paths via augmenting processes. We show that despite the increased dimensionality of the forward process, it is sufficient to learn a score model with the dimensionality of the data distribution. The resulting score function is guided by the marginal score of the augmenting processes, achieving better performance in terms of pixel-wise diversity measured by the vendi score and a lower FID, indicating better image quality.

Limitations & future work. Our work offers a new framework for using fractional noises within diffusion models. With that, several practical and theoretical issues remain open. Our future work will aim to empirically and theoretically characterize the optimal degree of correlated noise during the training and sampling of continuous time SBGMs. It also remains open to examine the extent to which our reverse time model converges to the backward SDE driven by time-reversed fBM. An interesting potential application would be to leverage the augmented probability flow ODE of GFDM directly to train a one-step diffusion model or to use GFDM to generate rough time-series data.



(a) Purely Brownian VE samples.



(b) Positively correlated regime of MA-fBM.

Figure 2: (**a**) Randomly chosen images generated by the PF ODE corresponding to the purely Brownian VE dynamics with a pixel diversity of $\text{VS}_p = 2.97$. (**b**) randomly chosen images generated by our GFDM PF ODE in the sub-diffusion regime of $H = 0.7$ with $K = 3$ augmenting processes and a higher pixel-wise diversity of $\text{VS}_p = 4.55$ confirmed perceptually.

References

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 2256–2265. JMLR.org, 2015.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, pages 11895–11907, 2019.

Pavel Avdeyev, Chenlai Shi, Yuhao Tan, Kseniia Dudnyk, and Jian Zhou. Dirichlet Diffusion Score Model for biological sequence generation. In *International Conference on Machine Learning*, 2023. URL https://arxiv.org/abs/2305.10699.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *International conference on machine learning*, pages 8867–8887. PMLR, 2022.

Matteo Manica, Jannis Born, Joris Cadow, Dimitrios Christofidellis, Ashish Dave, Dean Clarke, Yves Gaetan Nana Teukam, Giorgio Giannone, Samuel C Hoffman, Matthew Buchan, et al. Accelerating material design with the generative toolkit for scientific discovery. *npj Computational Materials*, 9(1):69, 2023.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. Simple and controllable music generation. *arXiv preprint arXiv:2306.05284*, 2023.

Felix Kreuk, Gabriel Synnaeve, Adam Polyak, Uriel Singer, Alexandre Défossez, Jade Copet, Devi Parikh, Yaniv Taigman, and Yossi Adi. Audiogen: Textually guided audio generation. *arXiv preprint arXiv:2209.15352*, 2022.

Xiaohui Zeng, Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, and Karsten Kreis. LION: Latent point diffusion models for 3D shape generation. *arXiv preprint arXiv:2210.06978*, 2022.

Marco Aversa, Gabriel Nobis, Miriam Hägele, Kai Standvoss, Mihaela Chirica, Roderick Murray-Smith, Ahmed Alaa, Lukas Ruff, Daniela Ivanova, Wojciech Samek, Frederick Klauschen, Bruno Sanguinetti, and Luis Oala. DiffInfinite: Large mask-image synthesis via parallel random patch diffusion in histopathology. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023. URL https://openreview.net/forum?id=QXTjde8evS.

Miguel Espinosa and Elliot J Crowley. Generate your own Scotland: Satellite image generation conditioned on maps. *arXiv preprint arXiv:2308.16648*, 2023.

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. URL https://openreview.net/forum?id=PxTIG12RRHS.

Robert Brown. Xxvii. a brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *The Philosophical Magazine*, 4(21):161–173, 1828. doi: 10.1080/14786442808674769. URL https://doi.org/10.1080/14786442808674769.

Albert Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, pages 549–560, 1905.

Norbert Wiener. Differential-space. *Journal of Mathematics and Physics*, 2:131–174, 1923. URL https://doi.org/10.1002/sapm192321131.

Eunbi Yoon, Keehun Park, Sungwoong Kim, and Sungbin Lim. Score-based generative models with Lévy Processes. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=0Wp3VHX0Gm.

Hengyuan Ma, Li Zhang, Xiatian Zhu, and Jianfeng Feng. Approximated anomalous diffusion: Gaussian mixture score-based generative models, 2023. URL https://openreview.net/forum?id=yc9xen7EAzd.

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

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

10.1137/1010093. URL https://doi.org/10.1137/1010093.

Jerry Stinson. The (mis) behavior of markets. *Journal of Personal Finance*, 4(4):99, 2005.

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.

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. doi: https://doi.org/10.1016/j.spa.2018.04.010. URL https://www.sciencedirect.com/science/article/pii/S0304414918301418.

Philip E. Protter. *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 2st edition, 2013. ISBN 978-3-662-10061-5. doi: 10.1007/978-3-662-10061-5.

Anh Tong, Thanh Nguyen-Tang, Toan Tran, and Jaesik Choi. Learning fractional white noises in neural stochastic differential equations. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 37660–37675. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/f51df088779c27cbb25b8f094a346544-Paper-Conference.pdf.

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. doi: https://doi.org/10.1007/978-1-4939-2867-5.

R. L. Stratonovich. Conditional Markov Processes. *Theory of Probability & Its Applications*, 5(2):156–178, 1960. doi: 10.1137/1105015. URL https://doi.org/10.1137/1105015.

Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(82)90051-5. URL https://www.sciencedirect.com/science/article/pii/0304414982900515.

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

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005. URL http://jmlr.org/papers/v6/hyvarinen05a.html.

Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: A scalable approach to density and score estimation. In *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22-25, 2019*, page 204, 2019. URL http://auai.org/uai2019/proceedings/papers/204.pdf.

Christian Bayer, Antonis Papapantoleon, and Raul Tempone. *Computational Finance*. Technical University of Berlin, 2021. URL https://www.wias-berlin.de/people/bayerc/files/lecture.pdf. Lecture notes.

Terry J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998. URL http://eudml.org/doc/39555.

James Davidson and Nigar Hashimzade. Type i and type ii fractional Brownian motions: A reconsideration. *Computational Statistics & Data Analysis*, 53(6):2089–2106, 2009. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2008.11.008. URL https://www.sciencedirect.com/science/article/pii/S0167947308005458. The Fourth Special Issue on Computational Econometrics.

Aaron Lou and Stefano Ermon. Reflected diffusion models. In *International Conference on Machine Learning*. PMLR, 2023.

Dang Huy. A remark on non-Markov property of a fractional Brownian motion. *Vietnam Journal of Mathematics*, 31, 01 2003.

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.

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. URL https://doi.org/10.1214/EJP.v12-439.

Philipp Harms. Strong convergence rates for Markovian representations of fractional processes, 2019. URL https://arxiv.org/abs/1902.01471.

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

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Proc. NeurIPS*, 2022.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. *Advances in neural information processing systems*, 29, 2016.

Dan Friedman and Adji Bousso Dieng. The Vendi Score: A diversity evaluation metric for machine learning. *arXiv preprint arXiv:2210.02410*, 2022.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=osei3IzUia.

Raghav Singhal, Mark Goldstein, and Rajesh Ranganath. Where to diffuse, how to diffuse, and how to get back: Automated learning for multivariate diffusions. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=osei3IzUia.

Bowen Jing, Gabriele Corso, Renato Berlinghieri, and Tommi Jaakkola. Subspace diffusion generative models. In *Lecture Notes in Computer Science*, pages 274–289. Springer Nature Switzerland, 2022. doi: 10.1007/978-3-031-20050-2_17. URL https://doi.org/10.1007%2F978-3-031-20050-2_17.

Dongjun Kim, Byeonghu Na, Se Jung Kwon, Dongsoo Lee, Wanmo Kang, and Il-Chul Moon. Maximum likelihood training of implicit nonlinear diffusion model. *Advances in Neural Information Processing Systems*, 35:32270–32284, 2022.

Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron C Courville. Riemannian diffusion models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 2750–2761. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/123d3e814e257e0781e5d328232ead9b-Paper-Conference.pdf.

Charlotte Bunne, Ya-Ping Hsieh, Marco Cuturi, and Andreas Krause. The Schrödinger Bridge between Gaussian Measures has a closed form. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023.

Giannis Daras, Mauricio Delbracio, Hossein Talebi, Alex Dimakis, and Peyman Milanfar. Soft diffusion: Score matching with general corruptions. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=W98rebBxlQ.

Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OjDkC57x5sz.

Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie S. Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=XH3ArccntI.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *International Conference on Learning Representations (ICLR)*, 2023.

Kohei Hayashi and Kei Nakagawa. Fractional SDE-Net: Generation of time series data with long-term memory, 2022.

Christian Bayer, Peter K. Friz, and Nikolas Tapia. Stability of deep neural networks via discrete rough paths. *SIAM Journal on Mathematics of Data Science*, 5(1):50–76, 2023. doi: 10.1137/22M1472358. URL https://doi.org/10.1137/22M1472358.

P Kidger. *On neural differential equations*. PhD thesis, University of Oxford, 2021.

Shujian Liao, Terry Lyons, Weixin Yang, and Hao Ni. Learning stochastic differential equations using rnn with log signature features. *arXiv preprint arXiv:1908.08286*, 2019.

James Morrill, Cristopher Salvi, Patrick Kidger, James Foster, and Terry Lyons. Neural rough differential equations for long time series, 2021.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, 1973. URL http://www.jstor.org/stable/1831029.

Christoph Czichowsky, Rémi Peyre, Walter Schachermayer, and Junjian Yang. Shadow prices, fractional Brownian motion, and portfolio optimisation under transaction costs. *Finance and Stochastics*, 22:161–180, 2018.

Paolo Guasoni, Zsolt Nika, and Miklós Rásonyi. Trading fractional Brownian motion. *SIAM journal on financial mathematics*, 10(3):769–789, 2019.

Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. In *Commodities*, pages 659–690. Chapman and Hall/CRC, 2022.

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*, pages 234–241. Springer, 2015.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, December 2014. arXiv:1412.6980 [cs.LG].

Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates, 2018.

Marjorie G. Hahn, Kei Kobayashi, and Sabir Umarov. Fokker-Planck-Kolmogorov equations associated with time-changed fractional Brownian motion. *arXiv: Mathematical Physics*, 139:691–705, 2011.

Appendix

A The mathematical framework of generative fractional diffusion models **14**

A.1 A Markovian representation of fractional Brownian motion 14

A.2 The forward model 15

A.3 Estimating the score via augmented score matching loss 18

B Forward sampling **20**

C Experimental results **23**

D Illustration of generated data **25**

E Likelihood computation **25**

F Challenges in the attempt to generalize **26**

G Notational conventions **27**

A. The mathematical framework of generative fractional diffusion models

In this section we provide the mathematical details of the score-based generative model defined in the main paper. The driving noise of the underlying stochastic process is based on the affine representation of fractional processes from Harms and Stefanovits (2019) and further simplified by the closed-form expression to determine optimal approximation coefficients of Daems et al. (2024).

A.1. A Markovian representation of fractional Brownian motion

We begin with the definition of type I fractional Brownian motion, defined on the whole real line, possessing correlated increments that are in contrast to type II fractional Brownian motion stationary.

Definition A.1 (Type I Fractional Brownian Motion (Mandelbrot and Van Ness, 1968))**.** *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 two standard independent $\{\mathcal{F}_t\}$-Brownian motions (BMs) $\tilde{B}$ and B the centered Gaussian process $W^H = (W_t^H)_{t\in\mathbb{R}}$ with*

$$W_t^H := \frac{1}{\Gamma(H+\frac{1}{2})}\int_{-\infty}^{0}((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}})\mathrm{d}\tilde{B}_s + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-s)^{H-\frac{1}{2}}\mathrm{d}B_s \tag{22}$$

uniquely characterized in law by its covariances

$$\mathbb{E}\left[W_t^H W_s^H\right] = \frac{1}{2}\left[t^{2H} + s^{2H} - (t-s)^{2H}\right], \quad t \geq s > 0 \tag{23}$$

is called type I fractional Brownian motion (fBM) with Hurst index $H \in (0,1)$.

Type II fBM from the main paper is retrieved by setting the additionally defined BM $\tilde{B}$ on the negative real line to zero. Therefore, the difference to type II fBM is the stochastic integral w.r.t. $\tilde{B}$ that yields stationary increments and a non trivial distribution at $t = 0$. For $H = \frac{1}{2}$, the process is a BM and has thus independent increments. For $H \in (0,1) \setminus \{\frac{1}{2}\}$, the process possesses correlated increments and, compared to BM, smoother paths for $H > \frac{1}{2}$ due to positively correlated increments (super-diffusion) and rougher paths for $H < \frac{1}{2}$ due to negatively correlated increments (sub-diffusion). These three regimes reflect for type I fBM in the same change of quadratic variation from t (Cohen and Elliott, 2015) to zero quadratic variation in the smooth regime and to infinite quadratic variation in the rough regime. To prepare the approximation of the non-Markovian and non-semimartingale fBM (Biagini et al., 2008) via Markovian semimartingales, define for every $\gamma \in (0,\infty)$ the Ornstein-Uhlenbeck process Y^γ given by

$$Y_t^\gamma := Y_0^\gamma e^{-t\gamma} + \int_0^t e^{-\gamma(t-s)}dB_s, \quad t \geq 0, \quad Y_0 := \int_{-\infty}^0 e^{s\gamma}dW_s, \tag{24}$$

with speed of mean reversion γ and non trivial starting value in contrast to the OU processes defined in eq. (6) of the main paper. By Itô's product rule (Cohen and Elliott, 2015), the process Y^γ solves the same SDE

$$dY_t^\gamma = -\gamma Y_t^\gamma dt + dW_t, \quad Y_0 = \int_{-\infty}^0 e^{s\gamma}dW_s, \tag{25}$$

with different starting value. According to Harms and Stefanovits (2019) we represent fBm by an integral over the predefined family of Ornstein-Uhlenbeck processes.

Theorem A.2 (Markovian Representation of fBM (Daems et al., 2024; Harms and Stefanovits, 2019))**.** *The non-Markovian process W^H permits the infinite-dimensional Markovian representation*

$$W_t^H = \begin{cases} \int_0^\infty (Y_t^\gamma - Y_0^\gamma)\,\nu_1(\gamma)\mathrm{d}\gamma, & H \leq \frac{1}{2} \\ -\int_0^\infty \partial_\gamma (Y_t^\gamma - Y_0^\gamma)\,\nu_2(\gamma)\mathrm{d}\gamma, & H > \frac{1}{2} \end{cases} \tag{26}$$

where $\nu_1(\gamma) = \gamma^{-(H+1/2)}/\Gamma(H+1/2)\Gamma(1/2-H)$ and $\nu_2(\gamma) = \gamma^{-(H-1/2)}/(\Gamma(H+1/2)\Gamma(3/2-H))$.

Note that we follow Daems et al. (2024) in replacing the process $Z_t^\gamma := Z_0^\gamma e^{-t\gamma} + \int_0^t e^{-(t-s)\gamma}Y_s^\gamma ds$ from the original theorem throughout this work by $Z_t^\gamma = -\partial_\gamma Y_t^\gamma + (\partial_\gamma Y_0^\gamma + Z_0^\gamma)\,e^{-t\gamma}$. This is justified by Harms and Stefanovits (2019,

Remark 3.5) and simplifies for $H > \frac{1}{2}$ the approximation of fBM and the definition of our generative model, since we only have to reverse the Y^γ processes instead of the pairs (Y^γ, Z^γ). For $Y_0^\gamma = 0$ eq. (26) yields an infinite-dimensional Markovian representation of type II fBM (Daems et al., 2024). The MA-fBM from the main paper becomes for type I fBM

$$\hat{B}_t^H = \sum_{k=1}^{K} \omega_k \left(Y_t^k - Y_0^k\right), \quad H \in (0,1), \quad t \geq 0 \tag{27}$$

with non trivial $\mathbf{Y}_0 = (Y_0^1, ..., Y_0^1)$ that is a centered multivariate Gaussian with covariances $\mathbb{E}\left[Y_0^k Y_0^l\right] = 1/(\gamma_k + \gamma_l)$ (Daems et al., 2024). Proposition 3.3 holds true for type I fBM as well with optimal approximation coefficients given in Daems et al. (2024, Proposition 5). For more details on the properties and distinction of type I and type II fBM we refer the reader to Daems et al. (2024).

A.2. The forward model

We define in the following a score-based generative model (SBGM) converging to a fractional diffusion process driven by type I fBM. For the remainder of Appendix A we assume $Y_0^k = \int_{-\infty}^0 e^{s\gamma_k} \mathrm{d}s$ for all $1 \leq k \leq K$ where the setting from the main paper with type II fBM is recovered by choosing $Y_0^k = 0$ instead. Let $\hat{\mathbf{B}}^H$ be a d-dimensional MA-fBM with Hurst index $H \in (0,1)$. For continuous functions $\mu : [0,T] \to \mathbb{R}$ and $\sigma : [0,T] \to \mathbb{R}$ we define the forward process $\mathbf{X} = (\mathbf{X}_t)_{t\in[0,T]}$ by

$$\mathrm{d}\mathbf{X}_t = \mu(t)\mathbf{X}_t \mathrm{d}t + g(t)\mathrm{d}\hat{\mathbf{B}}_t^H, \quad \mathbf{X}_0 = \mathbf{x}_0 \sim p_0, \quad t \in [0,T] \tag{28}$$

where p_0 is an unknown data distribution from which we aim to sample from. Using eq. (25) we note

$$\mathrm{d}\hat{\mathbf{B}}_t^H = -\sum_{i=1}^{K} \omega_k \gamma_k Y_t^k dt + \sum_{k}^{K} \omega_k \mathrm{d}\mathbf{B}_t, \tag{29}$$

where $\mathbf{B} = (B_1, ..., B_d)$ is a multivariate BM. With $\bar{\boldsymbol{\omega}} := \sum_{k=1}^K \omega_k$ we rewrite the dynamics of the forward process as

$$\mathrm{d}\mathbf{X}_t = \left[\mu(t)\mathbf{X}_t - g(t)\sum_{k=1}^{K} \omega_k \gamma_k \mathbf{Y}_t^k\right] \mathrm{d}t + \bar{\boldsymbol{\omega}} g(t)\mathrm{d}\mathbf{B}_t, \quad t \in [0,T], \tag{30}$$

Taking into account the dynamics of the OU processes, we define the *augmented forward process* $\mathbf{Z} = (\mathbf{Z}_t)_{t\in[0,T]}$ by

$$\mathbf{Z}_t = (X_{t,1}, Y_{t,1}^1, ..., Y_{t,1}^K, X_{t,2}, Y_{t,2}^1, ..., Y_{t,2}^K, ..., ..., ..., X_{t,D}, Y_{t,D}^1, ...Y_{t,D}^K) \in \mathbb{R}^{D(K+1)} \tag{31}$$

following the dynamics

$$\mathrm{d}\mathbf{Z}_t = \mathbf{F}(t)\mathbf{Z}_t dt + \mathbf{G}(t)\mathrm{d}\mathbf{B}_t \tag{32}$$

with $\mathbf{F}(t) = diag(\mathbf{D}(t), ..., \mathbf{D}(t)) \in \mathbb{R}^{d(K+1),d(K+1)}$,

$$\mathbf{D}(t) = \begin{pmatrix} \mu(t) & -g(t)\omega_1\gamma_1 & \cdots & -g(t)\omega_K\gamma_K \\ \mathbf{0}_K & & -diag(\gamma_1, ..., \gamma_K) & \end{pmatrix} \in \mathbb{R}^{K+1,K+1} \tag{33}$$

and

$$\boldsymbol{G}(t) = \begin{pmatrix} \bar{\boldsymbol{\omega}} g(t)\boldsymbol{I}_d & \boldsymbol{I}_d & \ldots & \boldsymbol{I}_d \end{pmatrix}^T \in \mathbb{R}^{d(K+1),d} \tag{34}$$

The augmented forward process $\mathbf{Z}$ conditioned on $\mathbf{y}_0^1, ..., \mathbf{y}_0^K$ and a data sample $\mathbf{x}_0 \sim p_0$ is a linear transformation of BM and hence a Gaussian process and so is $\mathbf{X}$ (Särkkä and Solin, 2019). Since the integral w.r.t BM has zero mean, the mean vector of the augmenting processes is $\mathbb{E}\left[\mathbf{Y}_t^k\right] = \mathbf{0}_d$ for all $1 \leq k \leq K$ and the mean of the conditional forward process is the solution of the ODE

$$\partial_t \mathbb{E}\left[\mathbf{X}_t | \mathbf{x}_0\right] = \mu(t)\mathbb{E}\left[\mathbf{X}_t | \mathbf{x}_0\right] \tag{35}$$

and hence the marginal mean

$$\mathbb{E}\left[\mathbf{X}_t | \mathbf{x}_0\right] = c(t)\mathbf{x}_0 \quad \textit{with} \quad c(t) = \exp\left(\int_0^t \mu(s)ds\right) \tag{36}$$

is not affected by changing the driving noise to MA-fBM. The marginal covariance matrix $\mathbf{\Sigma}_t$ of the conditional augmented forward process can be approximated numerically by solving an ODE, see Appendix B for details. In addition we present a continuous reparameterization of the forward process, resulting for some forward dynamics in a closed form solution of the marginal covariance matrix. Our result generalizes the explicit formula for the perturbation kernel $p_{0t}(\mathbf{x}|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}; c(t)\mathbf{x}_0, c^2(t)\sigma^2(t)\mathbf{I}_d)$ given in (Karras et al., 2022).

Proposition A.3 (Continuous Reparameterization Trick)**.** *Let $\mathbf{x}_0$ be a fixed realisation drawn from p_0. The forward process $\mathbf{X} = (\mathbf{X}_t)_{t\in[0,T]}$ conditioned on $\mathbf{x}_0$ admits the continuous reparameterization*

$$\mathbf{X}_t = c(t)\left(\mathbf{x}_0 + \int_0^t \alpha(t,s)\mathrm{d}\mathbf{B}_s\right) + \underbrace{c(t)\sum_{k=1}^K \omega_k\gamma_k \int_0^t \frac{g(s)}{c(s)} e^{-s\gamma_k}\mathrm{d}s\mathbf{Y}_0^k}_{=0 \text{ for type II fBM since } \mathbf{Y}_0^k=0} \tag{37}$$

with $c(t) = \exp\left(\int_0^t \mu(s)\mathrm{d}s\right)$ and

$$\alpha(t,s) = -\sum_{k=1}^K \omega_k\gamma_k \int_s^t \frac{g(u)}{c(u)} e^{-\gamma_k(u-s)}\mathrm{d}u + \bar{\boldsymbol{\omega}}\frac{g(s)}{c(s)} \tag{38}$$

such that $\mathbf{X}_t|\mathbf{x}_0 \sim \mathcal{N}\left(c(t)\mathbf{x}_0, \left[c^2(t)\sigma^2(t) + \sigma_K^2(t)\right]\mathbf{I}_d\right)$ is a Gaussian random vector for all $t \in (0,T]$ with

$$\sigma^2(t) = \int_0^t \alpha^2(t,s)ds \tag{39}$$

and

$$\sigma_K^2 = c^2(t)\sum_{k=1}^K \frac{\gamma_k}{2}\left[\omega_k \int_0^t \frac{g(s)}{c(u)}\mathrm{d}u\right]^2 \tag{40}$$

$$+ 2c^2(t)\sum_{k<l} \frac{\omega_k\omega_l\gamma_k\gamma_l}{\gamma_k+\gamma_l}\int_0^t \frac{g(s)}{c(s)}e^{-s\gamma_k}\mathrm{d}s\int_0^t \frac{g(s)}{c(s)}e^{-s\gamma_l}\mathrm{d}s \tag{41}$$

vanishing for an underlying type II fBM.

Proof. By continuity, the functions μ and σ are bounded. Moreover, the processes $Y_j^1, ..., Y_j^K$ posses continuous, hence bounded, paths and thus

$$\int_0^t |\mu(u)|\mathrm{d}u < \infty, \quad \int_0^t \sigma^2(u)\mathrm{d}u < \infty \quad \textit{and} \quad \int_0^t |\sum_k^K \omega_k\gamma_k\mathbf{Y}_t^k|\mathrm{d}u < \infty \quad \mathbb{P}-a.s., \tag{42}$$

where the last integral is understood entrywise. Hence, by Theorem 16.6.1 in (Cohen and Elliott, 2015), the unique solution of the SDE eq. (30) is given explicitly as

$$\mathbf{X}_t = c(t)\left(\mathbf{x}_0 - \int_0^t \frac{g(u)}{c(u)}\left[\sum_{k=1}^K \omega_k\gamma_k\mathbf{Y}_u^k\right]\mathrm{d}u + \bar{\boldsymbol{\omega}}\int_0^t \frac{g(u)}{c(u)}\mathrm{d}\mathbf{B}_u\right), \tag{43}$$

with $c(t) = \exp\left(\int_0^t \mu(s)ds\right)$. Define

$$\mathbf{J}\left(\mathbf{Y}_0^{[K]}, t\right) := \sum_{k=1}^K \omega_k\gamma_k \int_0^t \frac{g(s)}{c(s)}e^{-s\gamma_k}\mathrm{d}s\mathbf{Y}_0^k \tag{44}$$

and by the definition of Y_j^k in (24) we calculate using the Stochastic Fubini Theorem (Harms and Stefanovits, 2019)

$$\int_0^t \frac{g(u)}{c(u)}\left[\sum_{k=1}^K \omega_k\gamma_k\mathbf{Y}_u^k\right]\mathrm{d}u = \sum_{k=1}^K \omega_k\gamma_k\int_0^t\int_0^u \frac{g(u)}{c(u)}e^{-\gamma_k(u-s)}\mathrm{d}\mathbf{B}_s\mathrm{d}u + \mathbf{J}(\mathbf{Y}_0^{[K]}, t) \tag{45}$$

$$= \int_0^t \sum_{k=1}^K \omega_k\gamma_k\int_s^t \frac{g(u)}{c(u)}e^{-\gamma_k(u-s)}\mathrm{d}u\mathrm{d}\mathbf{B}_s + \mathbf{J}\left(\mathbf{Y}_0^{[K]}, t\right) \tag{46}$$

and hence

$$\mathbf{X}_t = c(t)\left(\mathbf{x}_0 - \int_0^t \frac{g(u)}{c(u)}\left[\sum_{k=1}^K \omega_k\gamma_k \mathbf{Y}_u^k\right]\mathrm{d}u + \bar{\boldsymbol{\omega}}\int_0^t \frac{g(u)}{c(u)}\mathrm{d}\mathbf{B}_u\right) \tag{47}$$

$$= c(t)\left(\mathbf{x}_0 - \int_0^t \sum_{k=1}^K \omega_k\gamma_k \int_s^t \frac{g(u)}{c(u)} e^{-\gamma_k(u-s)}\mathrm{d}u\mathrm{d}\mathbf{B}_s + \bar{\boldsymbol{\omega}}\int_0^t \frac{g(u)}{c(u)}\mathrm{d}\mathbf{B}_u - \mathbf{J}\left(\mathbf{Y}_0^{[K]}, t\right)\right)$$

$$= c(t)\left(\mathbf{x}_0 + \int_0^t \left[-\sum_{k=1}^K \omega_k\gamma_k \int_s^t \frac{g(u)}{c(u)} e^{-\gamma_k(u-s)}\mathrm{d}u + \bar{\boldsymbol{\omega}}\frac{g(s)}{c(s)}\right]\mathrm{d}\mathbf{B}_s - \mathbf{J}\left(\mathbf{Y}_0^{[K]}, t\right)\right)$$

$$= c(t)\mathbf{x}_0 + c(t)\int_0^t \int_0^t \alpha(t,s)\mathrm{d}\mathbf{B}_s - c(t)\mathbf{J}\left(\mathbf{Y}_0^{[K]}, t\right) \tag{48}$$

with

$$\alpha(t,s) = -\sum_{k=1}^K \omega_k\gamma_k \int_s^t \frac{g(u)}{c(u)} e^{-\gamma_k(u-s)}\mathrm{d}u + \bar{\boldsymbol{\omega}}\frac{g(s)}{c(s)}. \tag{49}$$

Since $\alpha(t,\cdot)$ is continuous for every fixed $t \in [0,T]$ we have $\int_0^t \alpha^2(t,s)\mathrm{d}s < \infty$. Using that the integral of a bounded deterministic function w.r.t. Brownian motion is a Gaussian process we have by Itô's isometry

$$\int_0^t \alpha(t,s)\mathrm{d}\mathbf{B}_s \sim \mathcal{N}\left(\mathbf{0}_d, \sigma^2(t)\mathbf{I}_d\right) \quad \textit{with} \quad \sigma^2(t) = \int_0^t \alpha^2(t,s)\mathrm{d}s. \tag{50}$$

Therefore, conditional on $\mathbf{x}_0$, the random vector $\mathbf{X}_t$ is Gaussian with mean vector

$$\boldsymbol{m}_t^{\mathbf{x}} = c(t)\boldsymbol{x}_0 + \underbrace{\mathbb{E}\left[\mathbf{J}(\mathbf{Y}_0^{[K]})\right]}_{=0} = \mathbf{x}_0 \exp\left(\int_0^t \mu(s)\mathrm{d}s\right). \tag{51}$$

Moreover, $\tilde{B}_j$ and B_j corresponding to the entries of $\tilde{\mathbf{B}} = (\tilde{B}_1, ..., \tilde{B}_d)$ and $\mathbf{B} = (B_1, ..., B_d)$ are independent by Definition A.1 resulting in the entrywise variance

$$\boldsymbol{\Sigma}_{t,j,j}^{\mathbf{x}} = c^2(t)\int_0^t \alpha^2(t,s)\mathrm{d}s + \sigma_K^2(t) \tag{52}$$

with

$$\sigma_K^2(t) = \mathbb{V}\left[\mathbf{J}(\mathbf{Y}_0^{[K]})_j\right] = c^2(t)\sum_{k=1}^K \frac{\gamma_k}{2}\left[\omega_k \int_0^t \frac{g(s)}{c(u)}\mathrm{d}u\right]^2 \tag{53}$$

$$+ 2c^2(t)\sum_{k<l} \frac{\omega_k\omega_l\gamma_k\gamma_l}{\gamma_k+\gamma_l}\int_0^t \frac{g(s)}{c(s)}e^{-s\gamma_k}\mathrm{d}s\int_0^t \frac{g(s)}{c(s)}e^{-s\gamma_l}\mathrm{d}s, \tag{54}$$

where we used again Itô's isometry to calculate

$$\mathbb{E}\left[Y_{0,j}^k Y_{0,j}^l\right] = \mathbb{E}\left[\int_{-\infty}^0 e^{\gamma_k s}\mathrm{d}\tilde{B}_{s,j}\int_{-\infty}^0 e^{\gamma_l s}\mathrm{d}\tilde{B}_{s,j}\right] = \int_{-\infty}^0 e^{(\gamma_k+\gamma_l)s}\mathrm{d}s = \frac{1}{\gamma_k+\gamma_l}. \tag{55}$$

Since the entries of $\mathbf{B}$ are independent, we find the covariance matrix

$$\boldsymbol{\Sigma}_t^{\mathbf{x}} = \left[c^2(t)\sigma^2(t) + \sigma_K^2(t)\right]\mathbf{I}_d. \tag{56}$$

□

The preceding proposition generalizes the "reparameterization trick"[2] from discrete time to continuous time in the sense that

$$\mathbf{X}_{t_n} = \sqrt{\bar{\alpha}_{t_n}}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_{t_n}}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d) \tag{57}$$

[2]See https://lilianweng.github.io/posts/2021-07-11-diffusion-models/ for the derivation in discrete time.

used in discrete time (Ho et al., 2020) with time steps $0 = t_0 < ... < t_N = T$ is replaced by our continuous time reparameterization

$$\mathbf{X}_t = c(t)\left(\mathbf{x}_0 + \int_0^t \alpha(t,s)\mathrm{d}\mathbf{B}_s\right) + c(t)\sum_{k=1}^K \omega_k\gamma_k \int_0^t \frac{g(s)}{c(s)} e^{-s\gamma_k}\mathrm{d}s\mathbf{Y}_0^k, \tag{58}$$

enabling to directly sample $\mathbf{X}_t|\mathbf{x}_0 \sim \mathcal{N}(c(t)\mathbf{x}_0 + \left[c^2(t)\sigma^2(t) + \sigma_K^2(t)\right]\mathbf{I}_D)$ for a given data sample $\mathbf{x}_0$ and time point $t \in (0,T]$, in case that $\sigma^2(t)$ and $\sigma_K^2(t)$ have a closed form solution. For a complete characterization of the marginal covariance matrix $\mathbf{\Sigma}_t$ of the conditioned augmented forward process we calculate by Itô isometry with $X = X_j$ and $Y^l = Y_j^l$ for all $1 \le j \le D, 1 \le l \le K$ and any $t \in [0,T]$

$$\mathbb{E}\left[X_t Y_t^l\right] = c(t)\int_0^t \alpha(t,s)e^{-\gamma_k(t-s)}\mathrm{d}s + c(t)\sum_{l=1}^K \frac{\omega_k\gamma_k}{\gamma_k+\gamma_l}e^{-\gamma_l t}\int_0^t \frac{g(s)}{c(s)}e^{-s\gamma_k}ds \tag{59}$$

and

$$\mathbb{E}\left[Y_t^k Y_t^l\right] = \frac{e^{-(\gamma_k+\gamma_l)s}}{\gamma_k+\gamma_l} + \frac{1-e^{-(\gamma_k+\gamma_l)t}}{\gamma_k+\gamma_l} = \frac{1}{\gamma_k+\gamma_l} \tag{60}$$

reducing for type II fBM to

$$\mathbb{E}\left[X_t Y_t^l\right] = c(t)\int_0^t \alpha(t,s)e^{-\gamma_k(t-s)}\mathrm{d}s \quad \textit{and} \quad \mathbb{E}\left[Y_t^k Y_t^l\right] = \frac{1-e^{-(\gamma_k+\gamma_l)t}}{\gamma_k+\gamma_l}. \tag{61}$$

We denote in the following the stacked vector of the augmenting processes by

$$\mathbf{Y}_t^{[K]} = (Y_{t,1}^1, Y_{t,1}^2, ..., Y_{t,1}^K, Y_{t,2}^1, Y_{t,2}^2, ..., Y_{t,2}^K, ..., Y_{t,D}^1, Y_{t,D}^2, ..., Y_{t,D}^K) \in \mathbb{R}^{D(K+1)}. \tag{62}$$

The random vector $\mathbf{Y}_t^{[K]}$ is a centered Gaussian process with covariance matrix

$$\mathbf{\Lambda}_t = diag(\mathbf{\Sigma}_t^{\mathbf{y}}, ..., \mathbf{\Sigma}_t^{\mathbf{y}}) \in \mathbb{R}^{D\cdot K, D\cdot K}, \quad \mathbf{\Sigma}_t^{\mathbf{y}} \in \mathbb{R}^{K,K}, \quad [\mathbf{\Sigma}_t^{\mathbf{y}}]_{k,l} = \mathbb{E}\left[Y_t^k Y_t^l\right] \tag{63}$$

where $\mathbf{\Sigma}_t^{\mathbf{y}}$ does not depend on the dimension $1 \le j \le D$ and we write q_t for the multivariate Gaussian density of $\mathbf{Y}_t^{[K]}$. Since we know the distribution of $\mathbf{Y}_0^{[K]}$, we can directly calculate the corresponding score-function by

$$\nabla_{\mathbf{y}^{[K]}} \log q_t\left(\mathbf{Y}_t^{[K]}\right) = -\mathbf{\Lambda}_t \mathbf{Y}_t^{[K]}. \tag{64}$$

A.3. Estimating the score via augmented score matching loss

Conditioning $\mathbf{Z}_t$ on $\mathbf{x}_0 \sim p_0$ *and* a realisation $\mathbf{y}_t^{[K]}$ of the stacked augmenting processes $\mathbf{Y}_t^{[K]}$ defined in eq. (62) at fixed time $t \in [0,T]$ results in the Gaussian vector $\tilde{\mathbf{X}}_t \sim \mathcal{N}(\tilde{\boldsymbol{m}}_t, \tilde{\mathbf{\Sigma}}_t)$ with mean

$$\tilde{\boldsymbol{m}}_t = c(t)\mathbf{x}_0 + \sum_{k=1}^K \eta_t^k \mathbf{y}_t^k, \quad \textit{where} \quad \eta_t^k = \sum_{l=1}^K \mathbb{E}\left[X_t Y_t^l\right]\left[\left(\mathbf{\Sigma}_t^{\mathbf{y}}\right)^{-1}\right]_{l,k} \tag{65}$$

and covariance

$$\tilde{\mathbf{\Sigma}}_t = \left(c^2(t)\sigma^2(t) - \tau_t^2\right)\mathbf{I}_d, \quad \textit{where} \quad \tau_t^2 = \sum_{k=1}^K \eta_t^k \mathbb{E}\left[X_t Y_t^k\right]. \tag{66}$$

We denote with $\nabla_{\mathbf{x}} \log p_{0t}$ the conditional score-function of $\tilde{\mathbf{X}}_t$ and calculate for the gradient w.r.t. $\mathbf{x} = (x_1, ..., x_D) \in \mathbb{R}^D$

$$\nabla_{\mathbf{x}} \log p_{0t}(\mathbf{x}|\mathbf{y}_t^{[K]}, \mathbf{x}_0) = -\tilde{\mathbf{\Sigma}}_t^{-1}(\mathbf{x} - \tilde{\boldsymbol{m}}_t) = -\frac{(\mathbf{x} - \tilde{\boldsymbol{m}}_t)}{(c^2(t)\sigma^2(t) - \tau_t^2)}. \tag{67}$$

and for the gradient w.r.t. $\mathbf{y}^k = (y_1^k, ..., y_D^k) \in \mathbb{R}^{\mathbb{D}}$

$$\nabla_{\mathbf{y}^k} \log p_{0t}(\mathbf{x}|\mathbf{y}_t^{[K]}, \mathbf{x}_0) = -\frac{1}{2}\nabla_{\mathbf{y}^k}\left[(\mathbf{x} - \tilde{\boldsymbol{m}}_t)^T \tilde{\mathbf{\Sigma}}_t^{-1}(\mathbf{x} - \tilde{\boldsymbol{m}}_t)\right] \tag{68}$$

$$= -\eta_t^k \nabla_{\mathbf{x}} \log p_{0t}(\mathbf{x}|\mathbf{y}_t^{[K]}, \mathbf{x}_0). \tag{69}$$

Deploying this relation of $\nabla_{\mathbf{x}} \log p_{0t}$ and $\nabla_{\mathbf{y}^k} \log p_{0t}$ we derive the *augmenting score matching loss* that reduces the dimensionality of the score-model we have to learn to the dimensionality of the data distribution and results in a score-model guided by the the known score-function $\nabla_{\mathbf{y}^{[K]}} \log q_t$.

Proposition A.4 (Optimal Score-Model)**.** *Assume that $s_{\boldsymbol{\theta}}$ is optimal w.r.t. the augmented score matching loss $\mathcal{L}$. The score-model*

$$S_\theta(\mathbf{Z}_t, t) := \left(s_{\boldsymbol{\theta}}(\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k, t), -\eta_t^1 s_{\boldsymbol{\theta}}(\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k, t), ..., -\eta_t^K s_{\boldsymbol{\theta}}(\mathbf{X}_t - \sum_k \eta \mathbf{Y}_t^k, t) \right)$$

yields the optimal $L^2(\mathbb{P})$ approximation of $\nabla_{\mathbf{z}} \log p_t(\mathbf{Z}_t)$ via

$$S_{\boldsymbol{\theta}}(\mathbf{Z}_t, t) + \nabla_{\mathbf{z}} \log q_t(\mathbf{Y}_t^{[K]}) \approx \nabla_{\mathbf{z}} \log p_t(\mathbf{Z}_t). \tag{70}$$

Proof. Fix $t \in [0, T]$. We write p_t^{aug} for the density of $\mathbf{Z}_t$, p_{0t}^{aug} for the conditional density of $\mathbf{Z}_t$ on $\mathbf{X}_0$, p_{0t} for the density of $\tilde{\mathbf{X}}_t$ and q_{0t} for the conditional density of $\mathbf{Y}_t^{[K]}$ on $\mathbf{X}_0$. First note that $\mathbf{Y}_t^{[K]}$ and $\mathbf{X}_0$ are independent by assumption and hence $q_t = q_{0t}$. By direct calculations we find

$$\begin{aligned}
\nabla_{\mathbf{x}} \log p_t^{aug}(\mathbf{Z}_t) &= \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{x}} \log p_{0t}^{aug}(\mathbf{Z}_t|\mathbf{X}_0) \right] && (71)\\
&= \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{x}} \log \left(p_{0t}(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0) q_{0t}(\mathbf{Y}_t^{[K]}|\mathbf{X}_0) \right) \right] && (72)\\
&= \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{x}} \log p_{0t}(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0) + \underbrace{\nabla_{\mathbf{x}} \log q_t(\mathbf{Y}_t^{[K]})}_{=\mathbf{0}_d} \right] && (73)\\
&= \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{x}} \log p_{0t}(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0) \right] && (74)\\
&\overset{(67)}{=} \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\frac{\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k - c(t)\mathbf{X}_0}{c^2(t)\sigma^2(t) - \tau_t^2} \right]. && (75)
\end{aligned}$$

Hence the best $L^2(\mathbb{P})$-approximation of $\nabla_{\mathbf{x}} \log p_t^{aug}(\mathbf{Z}_t)$ is a minimizer of the augmented score matching loss by

$$\begin{aligned}
\nabla_{\mathbf{x}} \log p_t^{aug}(\mathbf{Z}_t) &\overset{(75)}{=} \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\frac{\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k - c(t)\mathbf{X}_0}{c^2(t)\sigma^2(t) - \tau_t^2} \right] && (76)\\
&= \arg\min_{s_{\boldsymbol{\theta}}} \mathbb{E}_{(\mathbf{X}_0, \mathbf{Y}_t^{[K]})} \mathbb{E}_{(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0)} \left[\left\| s_{\boldsymbol{\theta}}(\mathbf{X}_t - \sum_{k=1}^K \eta_t^k \mathbf{Y}_t^k, t) - \frac{\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k - c(t)\mathbf{X}_0}{c^2(t)\sigma^2(t) - \tau_t^2} \right\|^2 \right] && (77)\\
&\overset{(67)}{=} \arg\min_{s_{\boldsymbol{\theta}}} \mathbb{E}_{(\mathbf{X}_0, \mathbf{Y}_t^{[K]})} \mathbb{E}_{(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0)} \left[\left\| s_{\boldsymbol{\theta}}(\mathbf{X}_t - \sum_{k=1}^K \eta_t^k \mathbf{Y}_t^k, t) - \nabla_{\mathbf{x}} \log p_{0t}(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0) \right\|^2 \right] && (78)
\end{aligned}$$

Assume now that $s_{\boldsymbol{\theta}}$ is a minimizer of the augmented score matching loss. Similar to the calculation above we have

$$\begin{aligned}
\nabla_{\mathbf{y}^k} \log p_t^{aug}(\mathbf{Z}_t) &= \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{y}^k} \log p_{0t}^{aug}(\mathbf{Z}_t|\mathbf{X}_0) \right] && (79)\\
&= \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{y}^k} \log \left(p_{0t}(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0) q_{0t}(\mathbf{Y}_t^{[K]}|\mathbf{X}_0) \right) \right] && (80)\\
&= \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{y}^k} \log p_{0t}(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0) + \nabla_{\mathbf{y}^k} \log q_t(\mathbf{Y}_t^{[K]}) \right] && (81)\\
&\overset{(68)}{=} -\eta_t^k \mathbb{E}_{(\mathbf{X}_0|\mathbf{X}_t, \mathbf{Y}_t^{[K]})} \left[\nabla_{\mathbf{x}} \log p_{0t}(\mathbf{X}_t|\mathbf{Y}_t^{[K]}, \mathbf{X}_0) \right] + \nabla_{\mathbf{y}^k} \log q_t(\mathbf{Y}_t^{[K]}) && (82)
\end{aligned}$$

and hence $-\eta_t^k s_{\boldsymbol{\theta}}(\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k) + \nabla_{\mathbf{y}^k} \log q_t(\mathbf{Y}_t^{[K]})$ is the best approximation of $\nabla_{\mathbf{y}^k} \log p_t^{aug}(\mathbf{Z}_t)$ in $L^2(\mathbb{P})$ and the score-model

$$S_\theta(\mathbf{Z}_t, t) := \left(s_\theta(\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k, t), -\eta_t^1 s_\theta(\mathbf{X}_t - \sum_k \eta_t^k \mathbf{Y}_t^k, t), ..., -\eta_t^K s_\theta(\mathbf{X}_t - \sum_k \eta \mathbf{Y}_t^k, t) \right)$$

yields the best $L^2(\mathbb{P})$-approximator of $\nabla_{\mathbf{z}} \log p_t$ via

$$S_{\boldsymbol{\theta}}(\mathbf{Z}_t, t) + \nabla_{\mathbf{z}} \log q_t(\mathbf{Y}_t^{[K]}) \approx \nabla_{\mathbf{z}} \log p_t(\mathbf{Z}_t). \tag{83}$$

□

B. Forward sampling

We assume throughout this section type II fBM. Given the marginal covariance matrix $\boldsymbol{\Sigma}_t$ of $\mathbf{Z}_t|\mathbf{x}_0$ we uniformly sample first a time point $t \in (0, T]$ and second $\mathbf{Z}_t \sim \mathcal{N}(\hat{\mathbf{z}}_t, \boldsymbol{\Sigma}_t)$ with

$$\hat{\mathbf{z}}_t = (c(t)\mathbf{x}_{0,1}, 0, ..., 0, c(t)\mathbf{x}_{0,2}, 0, ..., 0, ..., ..., ..., c(t)\mathbf{x}_{0,D}, 0, ...0) \in \mathbb{R}^{D(K+1)} \tag{84}$$

where we use $\mathbb{E}[\mathbf{X}_t|\mathbf{x}_0] = c(t)\mathbf{x}_0$ and $\mathbb{E}\left[\mathbf{Y}_t^k\right] = \mathbf{0}_D$. In the following we derive the entries of the marginal covariance matrix $\boldsymbol{\Sigma}_t$ using the dynamics

$$dY_t^k = -\gamma_k Y_t^k dt + d\mathbf{B}_t, \tag{85}$$

$$d\hat{\mathbf{B}}_t^H = d\left(\sum_{k=1}^K \omega_k Y_t^k\right) = -\sum_{i=1}^K \omega_k \gamma_k Y_t^k dt + \bar{\boldsymbol{\omega}} d\mathbf{B}_t, \quad \bar{\boldsymbol{\omega}} = \sum_k^K \omega_k, \tag{86}$$

$$dX_t = \mu(t)X_t dt + g(t)d\hat{\mathbf{B}}_t^H = \left[\mu(t)X_t - g(t)\sum_{k=1}^K \omega_k \gamma_k Y_t^k\right] dt + \bar{\omega} g(t) d\mathbf{B}_t, \tag{87}$$

and the continuous reparameterization

$$X_t = c(t)\left(\mathbf{x}_0 + \int_0^t \alpha(t,s) d\mathbf{B}_s\right), \quad c(t) = \exp\left(\int_0^t \mu(s)ds\right) \tag{88}$$

with

$$\alpha(t,s) = \bar{\omega}\frac{g(s)}{c(s)} - \sum_{k=1}^K \omega_k \gamma_k \int_s^t \frac{g(u)}{c(u)} e^{-\gamma_k(u-s)} du = \sum_{k=1}^K \omega_k \underbrace{\left(\frac{g(s)}{c(s)} - \gamma_k \int_s^t \frac{g(u)}{c(u)} e^{-\gamma_k(u-s)} du\right)}_{=\alpha_k(t,s)}. \tag{89}$$

With

$$f_k(u,s) = \frac{g(u)}{c(u)} e^{-\gamma_k(u-s)} \quad \textit{and} \quad I_k(t,s) = \int_s^t f_k(u,s) du \tag{90}$$

we have

$$\sigma_t^2 = c^2(t)\int_0^t \alpha^2(t,s) ds \tag{91}$$

$$= c^2(t)\int_0^t \left[\sum_{k=1}^K \omega_k \left(\frac{g(s)}{c(s)} - \gamma_k \int_s^t f_k(u,s) du\right)\right]^2 ds \tag{92}$$

$$= c^2(t)\int_0^t \left(\sum_{k=1}^K \omega_k \alpha_k(t,s)\right)^2 ds \tag{93}$$

$$= c^2(t)\int_0^t \sum_{i=1,j=1}^K \omega_i \omega_j \alpha_i(t,s)\alpha_j(t,s) ds \tag{94}$$

$$= \sum_{i=1,j=1}^K \omega_i \omega_j c^2(t)\int_0^t \alpha_i(t,s)\alpha_j(t,s) ds \tag{95}$$

$$= \sum_{i=1,j=1}^K \omega_i \omega_j c^2(t)\int_0^t \left(\frac{g(s)}{c(s)} - \gamma_i I_i(t,s)\right)\left(\frac{g(s)}{c(s)} - \gamma_j I_j(t,s)\right) ds \tag{96}$$

$$= \sum_{i,j=1}^K \omega_i \omega_j \left\{\text{var}_B(t) - c^2(t)\int_0^t \left[\frac{g(s)}{c(s)}\left(\gamma_i I_i(t,s) + \gamma_j I_j(t,s)\right) - \gamma_i \gamma_j I_i(t,s) I_j(t,s)\right] ds\right\}, \tag{97}$$

where

$$\text{var}_B(t) = c^2(t) \int_0^t \frac{g^2(s)}{c^2(s)} ds \tag{98}$$

corresponds to the purely Brownian marginal variance, explicitly calculated for VE and VP in (Song et al., 2021). Using the above derivation, we derive the closed-form variance schedule for FVE dynamics.

Fractional Variance Exploding Fix $\sigma_{\max} > \sigma_{\min} > 0$ and define $r := \frac{\sigma_{\max}}{\sigma_{\min}}$. Following (Song et al., 2021) we set

$$\mu(t) \equiv 0 \quad \textit{and} \quad g(t) = ar^t \quad \textit{with} \quad a = \sigma_{min}\sqrt{2\log(r)} \tag{99}$$

such that $c(t) = \exp(0) = 1$ and calculate

$$I_k(t,s) = \int_s^t f_k(u,s)du = \int_s^t ar^u e^{-\gamma_k(u-s)}du = F(t) - F(s) \tag{100}$$

$$= \underbrace{\frac{a}{\ln(r) - \gamma_k}}_{a_k} \left(e^{\ln(r)t - \gamma_k t + \gamma_k s} - e^{\ln(r)s}\right) = a_k \left(r^t e^{-\gamma_k(t-s)} - r^s\right), \tag{101}$$

since the derivative of $F(u) = a_k r^u e^{-\gamma_k(u-s)}$ is given by

$$\frac{d}{du}F(u) = \frac{d}{du}\left[a_k r^u e^{-\gamma_k(u-s)}\right] = a_k r^u \ln(r) e^{-\gamma_k(u-s)} + a_k r^u e^{-\gamma_k(u-s)}(-\gamma_k) \tag{102}$$

$$= \frac{a}{\ln(r) - \gamma_k}(\ln(r) - \gamma_k)(r^u e^{-\gamma_k(u-s)}) = ar^u e^{-\gamma_k(u-s)}. \tag{103}$$

Covariance for $\mathbf{Z}_t = (X_{t,1}, Y^1_{t,1}, ..., Y^K_{t,1}, X_{t,2}, Y^1_{t,2}, ..., Y^K_{t,2}, ..., X_{t,d}, Y^1_{t,d}, ..., Y^K_{t,d})$: We calculate

$$\langle X, X\rangle_t = \sigma_t^2 = \sum_{i,j=1}^K \omega_i\omega_j \left\{ var_B(t) - a\gamma_i \underbrace{\int_0^t r^s I_i(t,s)ds}_{J_i(t)} - a\gamma_j \underbrace{\int_0^t r^s I_j(t,s)ds}_{J_j(t)} + \gamma_i\gamma_j \underbrace{\int_0^t I_i(t,s)I_j(t,s)ds}_{=J_{i,j}(t)} \right\} \tag{104}$$

with

$$J_k(t) = a_k \int_0^t r^s \left(r^t e^{-\gamma_k(t-s)} - r^s\right) ds = a_k \int_0^t r^{t+s} e^{-\gamma_k(t-s)} ds - a_k \int_0^t r^{2s} ds \tag{105}$$

$$= a_k \left[F_1(t) - F_1(0)\right] - a_k \left[F_2(t) - F_2(0)\right] = a_k \left[\frac{r^{2t} - r^t \mathrm{e}^{-\gamma_k t}}{\ln(r) + \gamma_\mathrm{k}} - \frac{r^{2t} - 1}{2\ln(r)}\right], \tag{106}$$

since

$$\frac{d}{ds}F_1(s) = \frac{d}{ds}\left[\frac{1}{\ln(r) + \gamma_k} r^{t+s} e^{-\gamma_k(t-s)}\right] = \frac{\left(r^{t+s}\ln(r) e^{-\gamma_k(t-s)} + r^{t+s} e^{-\gamma_k(t-s)}(\gamma_k)\right)}{\ln(r) + \gamma_k} = r^{t+s} e^{-\gamma_k(t-s)} \tag{107}$$

$$\frac{d}{ds}F_2(s) = \frac{d}{ds}\left[\frac{r^{2s}}{2\ln(r)}\right] = \frac{r^{2s}\ln(r)2}{2\ln(r)} = r^{2s}. \tag{108}$$

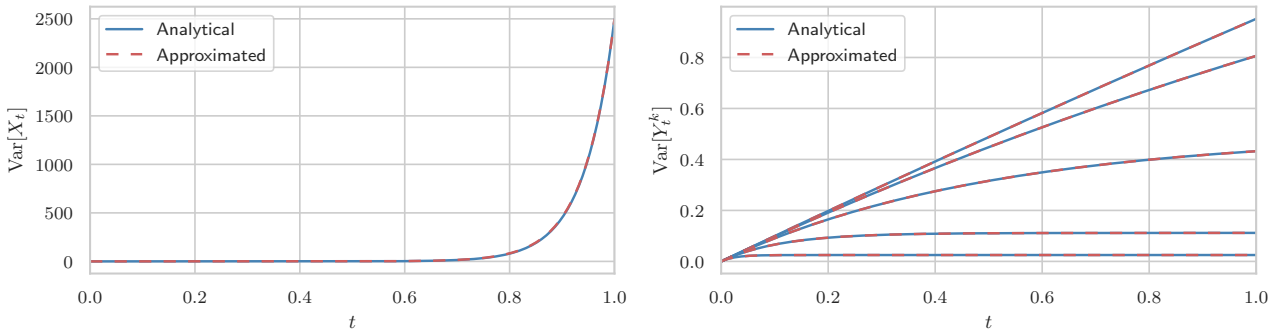
Finally

$$J_{i,j}(t) = a_i a_j \int_0^t \left(r^t e^{-\gamma_i(t-s)} - r^s\right)\left(r^t e^{-\gamma_j(t-s)} - r^s\right) ds \tag{109}$$

$$= a_i a_j \left(r^{2t} e^{-t(\gamma_i+\gamma_j)} \int_0^t e^{s(\gamma_i+\gamma_j)} ds - r^t e^{-\gamma_i t} \int_0^t e^{\gamma_i s} r^s ds - r^t e^{-\gamma_j t} \int_0^t e^{\gamma_j s} r^s ds + \int_0^t r^{2s} ds\right) \tag{110}$$

$$= a_i a_j \left[\left(\frac{r^{2t} e^{-t(\gamma_i+\gamma_j)}\left(e^{t(\gamma_i+\gamma_j)} - 1\right)}{\gamma_i + \gamma_j}\right) - r^t e^{-\gamma_i t}\frac{(e^{\gamma_i t} r^t - 1)}{\gamma_i + \ln(r)} - r^t e^{-\gamma_j t}\frac{(e^{\gamma_j t} r^t - 1)}{\gamma_j + \ln(r)} + \frac{r^{2t} - 1}{2\ln(r)}\right] \tag{111}$$

$$= a_i a_j \left[\left(\frac{r^{2t}\left(1 - e^{-t(\gamma_i+\gamma_j)}\right)}{\gamma_i + \gamma_j}\right) - \frac{r^{2t} - r^t e^{-\gamma_i t}}{\gamma_i + \ln(r)} - \frac{r^{2t} - r^t e^{-\gamma_j t}}{\gamma_j + \ln(r)} + \frac{r^{2t} - 1}{2\ln(r)}\right]. \tag{112}$$



(a) Variance schedule of the forward FVE process.

(b) Variance schedule of the augmenting processes.

Figure 3: Analytical solution (blue) used by our method for FVE dynamics with $K = 5$ and $H = 0.5$ compared to the approximated solution (dashed red) resulting from solving ODE (121).

We calculate the covariance of X and Y^l

$$\langle X, Y^l \rangle_t = c(t) \int_0^t \alpha(t,s) e^{-\gamma_l (t-s)} ds = \int_0^t \sum_{k=1}^K \omega_k \left[\frac{g(s)}{c(s)} - \gamma_k \int_s^t f_k(u,s) du \right] e^{-\gamma_l (t-s)} ds \tag{113}$$

$$= \sum_{k=1}^K \omega_k \left[a \int_0^t r^s e^{-\gamma_l (t-s)} ds - \gamma_k \int_0^t \int_s^t f_k(u,s) du e^{-\gamma_l (t-s)} ds \right] \tag{114}$$

$$= \sum_{k=1}^K \omega_k \left[a \int_0^t r^s e^{-\gamma_l (t-s)} ds - \gamma_k a_k \int_0^t \left(r^t e^{-\gamma_k (t-s)} - r^s \right) e^{-\gamma_l (t-s)} ds \right] \tag{115}$$

$$= \sum_{k=1}^K \omega_k \left[a e^{-\gamma_l t} \int_0^t r^s e^{\gamma_l s} ds - \gamma_k a_k \int_0^t \left(r^t e^{-\gamma_k (t-s)} - r^s \right) e^{-\gamma_l (t-s)} ds \right] \tag{116}$$

$$= \sum_{k=1}^K \omega_k \left[a e^{-\gamma_l t} \int_0^t r^s e^{\gamma_l s} ds - \gamma_k a_k \int_0^t r^t e^{-\gamma_k (t-s)} e^{-\gamma_l (t-s)} ds + \gamma_k a_k \int_0^t r^s e^{-\gamma_l (t-s)} ds \right] \tag{117}$$

$$= \sum_{k=1}^K \omega_k \left[\frac{a}{\ln(r) + \gamma_l} (r^t - e^{-\gamma_l t}) - \gamma_k a_k \frac{r^t (1 - e^{-t(\gamma_k + \gamma_l)})}{\gamma_k + \gamma_l} + \gamma_k a_k \frac{(r^t - e^{-\gamma_l t})}{\ln(r) + \gamma_l} \right] \tag{118}$$

$$= \sum_{k=1}^K \omega_k \left[(a + a_k \gamma_k) \frac{(r^t - e^{-\gamma_l t})}{\gamma_l + \ln(r)} - \gamma_k a_k \frac{r^t \left(1 - e^{-t(\gamma_k + \gamma_l)}\right)}{\gamma_k + \gamma_l} \right]. \tag{119}$$

Fractional Variance Preserving To the best of our knowledge, there is no closed form solution for $\int_s^t f_k(u,s) du$ for the dynamics of FVP. In this case, we numerically solve an ODE to determine the marginal covariance matrix of the conditional augmented forward process.

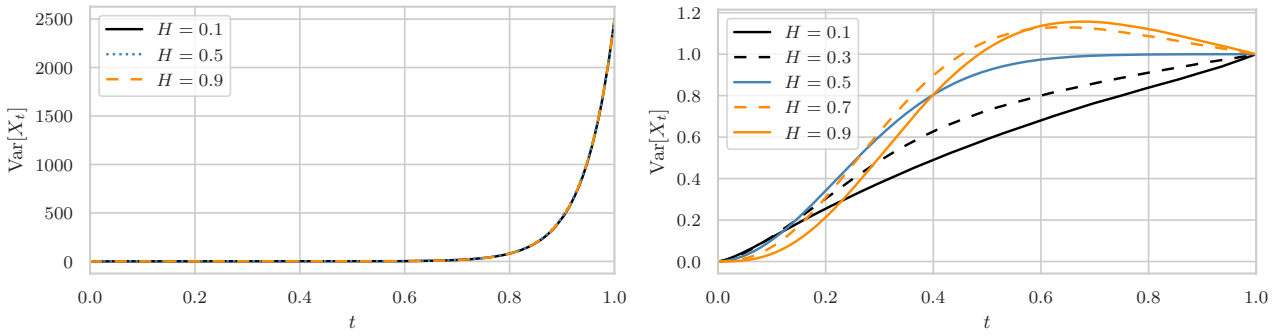
General Dynamics. The covariance matrix of the conditional augmented forward process with dynamics

$$d\mathbf{Z}_t = \mathbf{F}(t)\mathbf{Z}_t dt + \mathbf{G}(t) \mathrm{d}\mathbf{B}_t, \tag{120}$$

solves the ODE

$$\partial_t \mathbf{\Sigma}_t = \mathbf{F}(t)\mathbf{\Sigma}_t + \mathbf{\Sigma}_t \mathbf{F}(t)^T + \mathbf{G}(t)\mathbf{G}(t)^T, \tag{121}$$

lacking in general a closed form solution (Särkkä and Solin, 2019) in contrast to the setting of Song et al. (2021). This approach is applicable for any choice of μ and g in the forward dynamics, but depending on the choice of drift and diffusion function it might not yield a stable solution. We empirically observe in Figure 3 that the analytical solution for FVE and the numerical approximation of the variance schedule, determined by solving eq. (121) do not differ significantly.



(a) Variance schedules of the forward FVE process.

(b) Variance schedules of the forward FVP process.

Figure 4: Normalized variance schedules for $K = 5$ over time. (a) Variance schedules of FVE dynamics, calculated in closed form according to the derived formulas. The shape of the schedule is preserved throughout different values of H. (b) Variance schedules of FVP dynamics numerically approximated. The shape of the schedule is shifted for different values of H.

Variance schedules. We normalize the variance schedule of FVE and FVP dynamics such that the variance at $t = 0$ and at $t = T$ is equal to the variance used in the purely Brownian setting of VE and VP dynamics. For both FVE and FVP dynamics we calculate $\tilde{\boldsymbol{\omega}}$ according to Proposition 3.3 and determine $\tilde{\sigma}_T^2$ and define $\boldsymbol{\omega} = \tilde{\boldsymbol{\omega}}/\tilde{\sigma}_T^2$ to weight the OU-processes. By doing so, the terminal variance remains the same throughout different choices of H, as empirically confirmed in Figure 4. In Figure 4 we observe for FVE dynamics that not only the terminal variance is the same across different choices of H but also the shape of the variance schedule. For FVP dynamics, the shape of the variance schedule shifts with different values of H, approaching a nearly linear schedule for $H = 0.1$, while $H = 0.9$ offers a decreasing variance towards the end near $t = T$.

C. Experimental results

We used for all experiments a conditional U-Net (Ronneberger et al., 2015) architecture and the Adam optimizer (Kingma and Ba, 2014) with PyTorchs OneCylce learning rate scheduler (Smith and Topin, 2018). On CIFAR we trained first without exponential moving average (EMA) and second with EMA.

Set up on MNIST. We used an attention resolution of $[4, 2]$, 3 resnet blocks and a channel multiplication of $[1, 2, 2, 2, 2]$ and trained with a maximal learning rate of 10^{-4} for $50k$ iterations and a batch size of 1024. For all MNIST training runs we used one A100 GPU per run, taking approximately 17 hours.

Set up on CIFAR. We used an attention resolution of $[8]$, 4 resnet blocks and a channel multiplication of $[1, 2, 2, 2, 2]$. For the experiments without EMA, we used the same setup as with MNIST, but trained the models in parallel on two A100 GPUs for $300k$ iterations with an effective batch size of 1024. When training with EMA, we followed the set up of Song et al. (2021) using an EMA decay of 0.9999 for all FVP dynamics and an EMA decay of 0.999 for all FVE dynamics. In contrast to Song et al. (2021) we used PyTorchs OneCycleLR learning rate scheduler with a maximal learning rate of $2 \cdot 10^{-4}$ and trained only for $1mio$ iterations instead of the $1.3mio$ iterations in Song et al. (2021).

For the quantitative results on the MNIST dataset see Table 4 for FVP dynamics and Table 5 for FVE dynamics. For the results on CIFAR without EMA see Table 6 for training with FVE dynamics and Table 2a for training with FVP dynamics. For all quantitaitve results using EMA see Table 7 and Table 8.

Evaluation of different Hurst indices on CIFAR10. We trained for $1mio$ iterations with EMA and ema decay of 0.9999 for the FVP dynamics and an ema decay of 0.999 for VP dynamics.

Table 4: Quantitative results for FVP dynamics on the MNIST dataset.

FVP	$H = 0.9$		$H = 0.7$		$H = 0.5$		$H = 0.3$		$H = 0.1$	
	FID ↓	VS_p ↑	FID ↓	VS_p ↑	FID ↓	VS_p ↑	FID ↓	VS_p ↑	FID ↓	VS_p ↑
VP (retrained)	-	-	-	-	1.44	23.64	-	-	-	-
$K = 1$	-	-	-	-	2.81	23.69	2.94	23.55	-	-
$K = 2$	1.93	24.00	2.30	23.82	2.92	23.63	6.04	23.32	2.56	23.82
$K = 3$	**0.72**	24.18	2.67	23.96	3.51	23.78	54.48	26.32	4.87	23.60
$K = 4$	1.22	24.76	0.86	24.39	1.86	24.50	10.23	24.37	6.25	23.89
$K = 5$	2.17	**25.15**	1.36	24.63	4.89	24.56	7.44	24.71	9.57	23.70

Table 5: Quantitative results for FVE dynamics on the MNIST dataset.

FVE	$H = 0.9$		$H = 0.7$		$H = 0.5$		$H = 0.3$		$H = 0.1$	
	FID ↓	VS_p ↑	FID ↓	VS_p ↑	FID ↓	VS_p ↑	FID ↓	VS_p ↑	FID ↓	VS_p ↑
VE (retrained)	-	-	-	-	10.82	24.20	-	-	-	-
$K = 1$	10.06	24.05	9.95	24.24	10.30	24.22	9.91	24.19	9.98	24.20
$K = 2$	9.82	24.07	9.73	24.13	9.89	24.15	230.56	13.10	**9.42**	24.28
$K = 3$	11.02	24.53	9.96	24.37	9.74	24.42	31.86	20.29	10.12	24.44
$K = 4$	31.67	22.44	11.37	24.34	11.25	24.54	284.68	**25.77**	9.56	24.58
$K = 5$	50.42	23.74	22.03	22.09	25.51	23.08	260.38	14.56	10.39	24.33

Table 6: Quantitative results for FVE dynamics and varying Hurst index on CIFAR10. Original VE model compared to augmented FVE dynamics. We observe a slightly better performance with 1 or 2 augmenting processes in terms of quality and pixel-wise diversity and degrading performance for $K > 2$.

FVE	$H = 0.9$			$H = 0.5$			$H = 0.1$		
	FID ↓	IS ↑	VS_p ↑	FID ↓	IS ↑	VS_p ↑	FID ↓	IS ↑	VS_p ↑
VE (retrained)	-	-	-	9.38	**9.47**	3.21	-	-	-
$K = 1$	9.52	9.17	3.22	9.46	9.18	3.22	**8.93**	9.26	3.26
$K = 2$	8.99	8.95	3.26	9.62	8.82	3.22	10.23	8.7	3.09
$K = 3$	16.67	2.68	2.175	13.41	8.28	2.94	16.54	7.81	2.62
$K = 4$	40.03	7.25	1.41	17.74	7.94	2.26	14.49	8.14	2.46

Sampled from SDE dynamics	FID ↓	IS ↑	NLLs Test ↓	VS_p ↑	VS_p^{min} ↑
VE (retrained)	6.70	**9.71**	3.75	3.98	2.36
VP (retrained)	11.74	9.13	3.56	2.50	1.93
$FVE(K = 1, H = 0.7)$ (**our**)	7.93	8.91	3.78	4.06	2.4
$FVE(K = 1, H = 0.1)$ (**our**)	8.17	8.84	3.30	3.98	2.36
$FVE(K = 2, H = 0.9)$ (**our**)	8.13	8.86	3.61	3.33	2.42
$FVP(K = 2, H = 0.7)$ (**our**)	6.51	9.46	**3.37**	3.00	2.22
$FVP(K = 2, H = 0.9)$ (**our**)	**5.52**	9.50	3.43	3.07	2.21
$FVP(K = 3, H = 0.7)$ (**our**)	7.10	9.07	3.91	3.18	2.39

Table 7: Quantitative results on CIFAR10 for different Hurst indices compared to the purely Brownian dynamics VE and VP. Sampled from the SDE.

Euler PF ODE	FID ↓	IS ↑	VS_p ↑	VS_p^{min} ↑
VE (retrained)	**7.23**	9.39	2.97	2.32
VP (retrained)	11.10	8.98	2.7	2.09
$FVE(K = 1, H = 0.7)$ (**our**)	8.04	8.77	3.15	2.4
$FVE(K = 1, H = 0.1)$ (**our**)	8.39	8.65	3.17	2.37
$FVE(K = 2, H = 0.9)$ (**our**)	37.10	7.63	4.83	3.05
$FVP(K = 2, H = 0.7)$ (**our**)	17.74	9.51	3.89	2.60
$FVP(K = 2, H = 0.9)$ (**our**)	16.38	9.36	4.10	2.79
$FVP(K = 3, H = 0.7)$ (**our**)	26.76	**9.51**	**4.55**	**3.10**

Table 8: Quantitative results on CIFAR10 for different Hurst indices compared to the purely Brownian dynamics VE and VP. Sampled from the PF ODE

D. Illustration of generated data

Comparing visually the pixel-wise diversity of generated data.

(a) Purely Brownian VE sample

(b) Positively correlated fBM regime

Figure 5: (LHS) randomly choosen images generated with GFDMs PF ODE for smooth $H = 0.7$ and $K = 3$ augmenting processes, corresponding to the trained model with the highest pixel-wise diversity. (RHS) randomly choosen images generated with the original model and the PF ODE corresponding to the purely Brownian VE dynamics. The higher pixel-wise vendi score is visually noticeable.

E. Likelihood computation

Given the approximate probability flow ODE corresponding to the augmented forward process

$$\mathrm{d}\mathbf{z}_t = \underbrace{\left\{\mathbf{F}(t)\mathbf{z}_t - \frac{1}{2}\mathbf{G}(t)\mathbf{G}(t)^T\left[S_{\boldsymbol{\theta}}(\mathbf{z}_t, t) + \nabla_{\mathbf{z}} \log q_t(\mathbf{y}_t^{[K]})\right]\right\}}_{:=\tilde{\mathbf{f}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t)} dt, \quad t \in [0, T] \tag{122}$$

we estimate according to Song et al. (2021) the log-likelihoods of test data $\mathbf{z}_0$ under the learned density $\tilde{p}_0^{aug}$ via

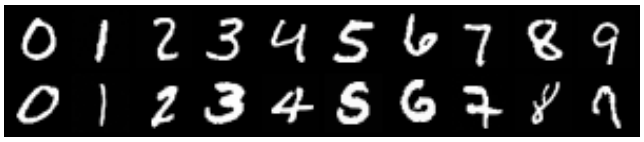
$$\log \tilde{p}_0^{aug}(\mathbf{z}_0) = \log \tilde{p}_T^{aug}(\mathbf{z}_T) + \int_0^T \nabla \tilde{\mathbf{f}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t)\mathrm{d}t. \tag{123}$$

According to Song et al. (2021), we integrate over $[\epsilon, T]$ rather than $[0, T]$, using the same value of $\epsilon = 10^{-3}$, which has been empirically shown to yield the best performance when simulating the SDE. For $\epsilon \neq 0$ and type II fBM we need to adjust the starting value of the augmenting processes from zero to a jointly sampled vector $\mathbf{y}_\epsilon = (y_\epsilon^1, ..., y_\epsilon^K) \sim \mathcal{N}(\mathbf{0}_K, \boldsymbol{\Lambda}_\epsilon)$ with

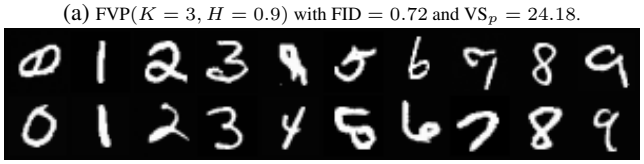
$$(\boldsymbol{\Lambda}_\epsilon)_{k,l} = \mathbb{E}\left[y_\epsilon^k y_\epsilon^l\right] = \int_0^\epsilon e^{-(\gamma_k+\gamma_l)(\epsilon-s)}\mathrm{d}s = \frac{1 - e^{-(\gamma_k+\gamma_l)\epsilon}}{\gamma_k + \gamma_l}. \tag{124}$$

Using the exact likelihood of $\mathbf{y}_\epsilon$ and the independence of $\mathbf{y}_\epsilon$ and $\mathbf{x}_0$ we have

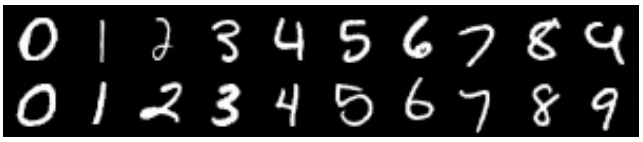
$$\log \tilde{p}_0^{aug}(\mathbf{z}_\epsilon) = \log \tilde{p}_0(\mathbf{x}_0) + \log q_\epsilon(\mathbf{y}_\epsilon) \tag{125}$$



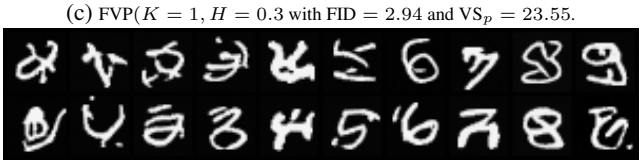
(a) FVP($K = 3$, $H = 0.9$) with FID $= 0.72$ and $\text{VS}_p = 24.18$.



(b) FVP($K = 5$, $H = 0.9$) with FID $= 2.17$ and $\text{VS}_p = 25.15$.



(c) FVP($K = 1$, $H = 0.3$ with FID $= 2.94$ and $\text{VS}_p = 23.55$.



(d) F($K = 3$, $H = 0.3$) with FID $= 54.48$ and $\text{VS}_p = 26.32$.

Figure 6: Diversifying effect of the augmenting processes with FVP dynamics on MNIST. (LHS) smooth regime with $H = 0.9$: For $K = 5$ instead of $K = 5$ augmenting processes the pixel VS increases from 24.18 to 25.15. (RHS) Rough regime with $H = 0.3$: Having $K = 3$ instead of $K = 1$ augmenting processes increases the pixel VS from 23.55 to 26.32, while degrading the FID from 2.94 to 54.48, a deterioration that is visually apparent.

where $\tilde{p}_0$ is the learned density of $\mathbf{x}_0$ corresponding to $\boldsymbol{\theta}$. Hence in total

$$\log \tilde{p}_0(\mathbf{x}_0) \overset{(123)}{=} \log \tilde{p}_T^{aug}(\mathbf{z}_T) + \int_0^T \nabla \tilde{\mathbf{f}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t)\mathrm{d}t - \log q_\epsilon(\mathbf{y}_\epsilon) \tag{126}$$

and we define the negative log-likelihoods $NLLs$ of test data $\mathbf{x}_0$ under the learned density by

$$NLLs(\mathbf{x}_0, \boldsymbol{\theta}) := -\log \tilde{p}_0^{aug}(\mathbf{z}_0) + \log q_\epsilon(\mathbf{y}_\epsilon). \tag{127}$$

F. Challenges in the attempt to generalize

In this work, we seek to determine the extent to which the continuous time framework of a SBGM can be generalized from an underlying BM to an underlying fBM. For a fBM W^H it is not straightforward to define the forward process

$$X_t = X_0 + \int_0^t f(X_s, s)ds + \int_0^t g(X_s, s)dW_s^H, \quad t \in [0, T] \tag{128}$$

driven by fBM, since fBM is neither a Markov process nor a semimartingale (Biagini et al., 2008), and hence Itô calculus may not be applied, to define the second integral. However, a definition of the integral w.r.t. fBM is established (Biagini et al., 2008; Hahn et al., 2011) such that the remaining problem is the derivation of the reverse time model. Following the second and more intuitive derivation of the reverse time model for BM from Anderson (1982), the conditional backward Kolmogorov equation and the unconditional forward Kolmogorov equation are applied. Starting point of the derivation is to rewrite $p(x_t, t, x_s, s) = p(x_s, s|x_t, t)p(x_t, t)$ with Bayes theorem to calculate with the product rule

$$\frac{\partial p(x_t, t, x_s, s)}{\partial t} = \frac{\partial p(x_s, s|x_t, t)}{\partial t} p(x_t, t) + \frac{\partial p(x_t, t)}{\partial t} p(x_s, s|x_t, t), \quad s \geq t. \tag{129}$$

Replacing $\frac{\partial p(x_t,t)}{\partial t}$ with the RHS of the unconditional forward Kolmogorov equation and $\frac{\partial p(x_s,s|x_t,t)}{\partial t}$ with the RHS of the conditional backward Kolmogorov equation one derives an equation that only depends on the joint density $p(x_t, t, x_s, s)$. Using Bayes theorem again leads to a conditional backward Kolmogorov equation for $p(x_t, t|x_s, s)$ that defines the dynamics of the reverse process by the one-to-one correspondence between the conditional backward Kolmogorov equation and the reverse time SDE (Anderson, 1982). Following these steps for fBM, starting from eq. (129) and deploying the one-to-one correspondence of fBM and the evolution of its density (Hahn et al., 2011), we could replace $\frac{\partial p(x_t,t)}{\partial t}$ in (129) by the RHS of

$$\frac{\partial p(x, t)}{\partial t} = \sum_{i=1}^{d} f_i(t, x) \frac{\partial p(t, x)}{\partial x_i} + Ht^{2H-1} \sum_{i,j=1}^{d} g_{ij}(x, t) \frac{\partial^2 p(t, x)}{\partial x_i \partial x_j}. \tag{130}$$

The missing part is however an analogous to the conditional backward Kolmogorov equation to replace $\frac{\partial p(x_s,s|x_t,t)}{\partial t}$ in eq. (129). The derivation of such an equation is to the best of our knowledge yet unsolved problem and hence the limiting factor in the generalization of continuous time SBGM from an underlying BM to an underlying fBM.

G. Notational conventions

$[0,T]$	Time horizon with terminal time $T>0$
$X=(X_t)_{t\in[0,T]}$	Stochastic forward process taking values in $\mathbb{R}$
$D\in\mathbb{N}$	Data dimension
$\mathbf{X}$	Vector valued stochastic forward process $\mathbf{X}=(\mathbf{X}_t)_{t\in[0,T]}$ with $\mathbf{X}_t=(X_{t,1},...,X_{t,D})$
$\overline{\mathbf{X}}$	Reverse time stochastic process with $\overline{\mathbf{X}}_t=\mathbf{X}_{T-t}$
$\mathbf{f}$	Function $\mathbf{f}:\mathbb{R}^D\times[0,T]\to\mathbb{R}^D$
μ,g	Functions $\mu,g:[0,T]\to\mathbb{R}$
$\overline{\mathbf{f}}$	Reverse time function with $\overline{\mathbf{f}}(\mathbf{x},t)=\mathbf{f}(\mathbf{x},T-t)$
$\bar{\mu},\bar{g}$	Reverse time functions with $\bar{\mu}(t)=\mu(T-t)$ and $\bar{g}(t)=g(T-t)$
p_0	Data distribution
p_t	Marginal density of (augmented) forward process at $t\in[0,T]$
B	Brownian motion (BM)
H	Hurst index $H\in(0,1)$
W^H	Type I fractional Brownian motion (fBM)
B^H	Type II fractional Brownian motion (fBM)
$Y^\gamma=(Y_t^\gamma)_{t\in[0,T]}$	Ornstein–Uhlenbeck (OU) process with speed of mean reversion $\gamma\in\mathbb{R}$
$K\in\mathbb{N}$	Number of approximating processes
$\gamma_1,...,\gamma_K$	Geometrically spaced grid
$\omega_1,...,\omega_K$	Approximation coefficients
$\boldsymbol{\omega}$	Optimal approximation coefficients $\boldsymbol{\omega}=(\omega_1,...,\omega_K)$
$\bar{\boldsymbol{\omega}}$	Sum of optimal approximation coefficients
$\hat{B}^H$	Markov-approximate fractional Brownian motion (MA-fBM)
$\mathbf{Y}^1,...,\mathbf{Y}^K$	Augmenting processes with $\mathbf{Y}^k=(Y^k,...,Y^k)$
$\mathbf{F},\mathbf{G}$	Vector valued functions $\mathbf{F},\mathbf{G}:[0,T]\to\mathbb{R}^{D\cdot(K+1)}$
$\mathbf{Z}$	By $\mathbf{Y}^1,...,\mathbf{Y}^K$ augmented forward process
$\mathbf{Y}^{[K]}$	Stacked vector of augmenting processes
q_t	Marginal density of $\mathbf{Y}^{[K]}$ at $t\in[0,T]$
$\boldsymbol{\theta}$	Weight vector of a neural network