Sampling first-passage times of fractional Brownian motion using adaptive bisections

Benjamin Walter[1] and Kay Jörg Wiese[2]

[1]*Department of Mathematics, Imperial College London, London SW7 2AZ, England, United Kingdom*

[2]*Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL, Centre National de la Recherche Scientifique, Sorbonne Université, Université Paris-Diderot, Sorbonne Paris Cité, 24 rue Lhomond, 75005 Paris, France*

(Received 8 October 2019; accepted 17 March 2020; published 29 April 2020)

We present an algorithm to efficiently sample first-passage times for fractional Brownian motion. To increase the resolution, an initial coarse lattice is successively refined close to the target, by adding exactly sampled midpoints, where the probability that they reach the target is non-negligible. Compared to a path of N equally spaced points, the algorithm achieves the same numerical accuracy N_{eff}, while sampling only a small fraction of all points. Though this induces a statistical error, the latter is bounded for each bridge, allowing us to bound the total error rate by a number of our choice, say $P^{\text{tot}}_{\text{error}} = 10^{-6}$. This leads to significant improvements in both memory and speed. For $H = 0.33$ and $N_{\text{eff}} = 2^{32}$, we need 5 000 times less CPU time and 10 000 times less memory than the classical Davies-Harte algorithm. The gain grows for $H = 0.25$ and $N_{\text{eff}} = 2^{42}$ to 3×10^5 for CPU and 10^6 for memory. We estimate our algorithmic complexity as $\mathcal{C}^{\text{ABSec}}(N_{\text{eff}}) = O[(\ln N_{\text{eff}})^3]$, to be compared to Davies-Harte, which has complexity $\mathcal{C}^{\text{DH}}(N) = O(N \ln N)$. Decreasing $P^{\text{tot}}_{\text{error}}$ results in a small increase in complexity, proportional to $\ln(1/P^{\text{tot}}_{\text{error}})$. Our current implementation is limited to the values of N_{eff} given above, due to a loss of floating-point precision. Our algorithm can be adapted to other extreme events and arbitrary Gaussian processes. It enables one to numerically validate theoretical predictions that were hitherto inaccessible.

DOI: 10.1103/PhysRevE.101.043312

I. INTRODUCTION

Estimating the distribution of first-passage times (FPTs) is a key problem in understanding systems as different as financial markets or biological systems [1,2], and the dynamics of local reactions [3,4]. Typically, research focuses on non-Markovian processes and bounded geometries, where first-passage time distributions (FPTDs) are difficult to obtain analytically [5–9]. Within the class of non-Markovian processes, fractional Brownian motion (FBM) is of particular interest as it naturally extends standard diffusion to sub- and superdiffusive self-similar processes [10]. Fractional Brownian motion is a one-parameter family of Gaussian processes, indexed by the Hurst parameter $H \in (0, 1]$. The latter parametrizes the mean-square displacement via

$$\left\langle X_t^2 \right\rangle = 2t^{2H}, \tag{1}$$

recovering standard Brownian motion at $H = \frac{1}{2}$. It retains from Brownian motion scale and translational invariance, both in space and time. Due to its correlations, it has peculiar characteristics, as, e.g., the recently observed behavior near a reflecting boundary [11,12]. FBM has long defied an analytic study of its extreme events, except for results in the mathematical literature concerning the tail of these distributions [13–16].

In order to render the extreme events of this process accessible to an analytical treatment, an ε expansion around Brownian motion in $\varepsilon = H - \frac{1}{2}$ was proposed [17]. This field theoretic approach was applied to a variety of extreme events, yielding the first-order corrections of several probability distributions [8,17–19]. The scaling functions predicted by this perturbative field theory are computationally expensive to verify, since they require a high numerical resolution of the path. Typically this is done by simulating a discretized version of the path over a grid of N equidistant points. Measuring a first-passage time then amounts to finding the first passage of a linear interpolation of these grid points. This approximation, however, can lead to a systematic overestimation of the first-passage time. As can be seen in Fig. 1, a high resolution of the path is necessary in order to find the first-passage event at $t = 0.36$ instead of the one at $t = 0.45$ or even $t = 0.47$ for the coarser grids. To account for this, usually the number of grid points is increased. As the size of fluctuations between grid points diminishes as

$$\delta X = N^{-H}, \tag{2}$$

the subdiffusive regime ($H < \frac{1}{2}$) necessitates an enormous computational effort.

This poses challenges to the numerical validation of high-precision analytical predictions as can be seen, for instance, in [8]. There, in order to validate the analytically obtained scaling functions, and to minimize discretization errors in the subdiffusive regime, system sizes up to $N = 2^{24}$ are sampled using the standard Davies-Harte algorithm. The implementation used there required a CPU time of 6 s per sample. This illustrates that if theories of such high precision are to be tested against simulations new numerical techniques need to be developed. The present paper addresses this problem by designing, implementing, and benchmarking an algorithm sampling first-passage times of fractional Brownian motion using several orders of magnitude less CPU time and memory than traditional methods. The general idea is to start from a

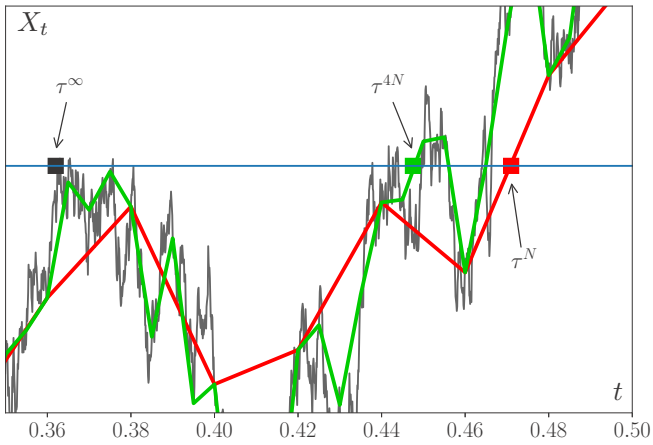

FIG. 1. The continuous stochastic path (gray rough line) crosses the barrier (blue horizontal line) for the first time at τ^∞ (black leftmost square mark). The discretization with N points (red line passing through rightmost square) overestimates this time as τ^N (red rightmost square mark). The numerical estimate is improved to τ^{4N} (green middle square mark) when refining the discretization (green line passing through middle square mark). This systematic error worsens for diminishing values of Hurst parameter H.

rather coarse grid (as the red one in Fig. 1), and to refine the grid where necessary. As a testing ground, we simulate and compare to theory the first-passage time of an FBM with drift [9].

The algorithm proposed here is an adaptive bisection routine (ABSec) that draws on several numerical methods already established in this field, notably the Davies-Harte algorithm [20], bisection methods [21,22], and the random midpoint displacement method [23,24]. The central, and quite simple, observation is that in order to resolve a first-passage event it is necessary to have a high grid resolution only near the target. This translates into an algorithm that generates a successively refined grid, where refinement takes place only at points close to the target, with the criterion of *closeness* scaling down by 2^{-H} for each bisection. This refinement is stopped after the desired resolution is reached. The sampling method is *exact*, i.e., the collection of points is drawn from the ensemble of FBM, a continuous process, with no bias. The only error one can make is that one misses an intermediate point. We have been able to control this error with a failure rate smaller than 10^{-6} per realization.

While there is a relatively large overhead for the nonhomogenous refinement, this is compensated by the use of far less points, leading to a significant increase both in speed and in memory efficiency over sampling methods that produce points for the full grid. For $H = 0.33$ and system size $N = 2^{32}$, our algorithm is 5000 times faster than the Davies-Harte (DH) algorithm, the fastest exact sampler (see [25,26]) if all points are needed. It has computational complexity $O[N \ln(N)]$, which makes it the standard algorithm in most current works (see, e.g., [6,7,27]), with system size N ranging from 2^{21} to 2^{24}. Our maximal grid size is limited by the precision of the floating-point unit to $N_{\max} \approx 2^{11/H}$.

This paper is organized as follows. In Sec. II, we introduce our adaptive bisection algorithm. First, its higher-level structure is outlined and then each subroutine is detailed. Possible generalizations to other extreme events or other Gaussian processes are discussed at the end of this section. In Sec. III, we present our implementation of the adaptive bisection in C, which is freely available [28]. We benchmark it against an implementation of the Davies-Harte algorithm. We compare error rates, average number of bisections, CPU time, and memory. Section IV contains a summary of our findings.

II. ALGORITHM

In this section, we introduce the ABSec. The central aim is to translate the idea of refining the grid "where it matters" into a rigorous routine.

A. Fractional Brownian motion and first-passage times

Gaussian processes X_t are stochastic processes for which X_t evaluated at a finite number of points $\mathcal{T}$ in time has a multivariate Gaussian distribution [29]. They are simple to handle, since their path probability measure can be obtained from their correlation function. The best known Gaussian process is Brownian motion, which is the only translational invariant Gaussian process with stationary and independent increments.

Fractional Brownian motion generalizes Brownian motion by relaxing the requirement of independent increments, while keeping self-similarity. The latter property means that its path probability measure is invariant under a space-time transformation $t \to ct$, $x \to c^{-H}x$ for $c > 0$. The parameter H is referred to as the *Hurst* exponent. As a Gaussian process, FBM is entirely characterized by its mean $X_0 = \langle X_t \rangle = 0$ and correlation function

$$C(s,t) = \langle X_s X_t \rangle = |s|^{2H} + |t|^{2H} - |t-s|^{2H}, \tag{3}$$

where $H \in (0, 1]$. As a consequence, $\langle (X_t - X_s)^2 \rangle = 2|t - s|^{2H}$, and in particular $\langle X_t^2 \rangle = 2|t|^{2H}$. From the correlation function it follows that on all time scales nonoverlapping increments are positively correlated for $H > \frac{1}{2}$ and negatively correlated for $H < \frac{1}{2}$. For $H = \frac{1}{2}$ one recovers Brownian motion with uncorrelated increments.

The FPT of a stochastic process is the first time the process crosses a threshold m. Since we use $X_0 = 0$, it is defined for $m > 0$ as

$$\tau_m = \inf_{t>0} \{t | X_t \geqslant m\}. \tag{4}$$

B. Notation

In simulating a FBM on a computer, one is forced to represent the continuous path by a *discretized path* that takes values on a finite set of points in time, the *grid*. We denote the grid by *ordered* times $\mathcal{T} = \{t_1, t_2, \ldots, t_N\}$, and the corresponding values of the process by $\mathcal{X} = \{X_{t_1}, X_{t_2}, \ldots, X_{t_N}\}$. Together, $(\mathcal{X}, \mathcal{T})$ form the discretized path. Due to self-similarity of the process, we can restrict ourselves to $\mathcal{T} \subset [0, 1]$ with no loss of generality. The intervals between any two successive times $t_i, t_{i+1} \in \mathcal{T}$ are referred to as *bridges* (t_i, t_{i+1}). Each connected component of $[0, 1] \backslash \mathcal{T}$ is a bridge.

We denote the dyadic lattice on the unit interval by $\Lambda^k = \{i2^{-k}; 0 \leqslant i \leqslant 2^k\}$. Our adaptive bisection algorithm sets out from a dyadic lattice Λ^g of relatively low resolution (typically $g \lesssim 8$ or 10). A FBM path is sampled for every point of the coarse grid Λ^g. If the linear interpolation of this coarse path already surpasses the threshold m at a time $\tau^{(0)}$, or in other words if there is a smallest K such that $X_{K2^{-g}} > m$, then the grid is truncated at $\tau^{(0)}$. (That this truncation does not introduce a bias is shown below.) If this is not the case, i.e., if all points remain below the threshold m, then the full grid is kept. We define the truncation of the grid $\mathcal{T}$ to a certain time $\tau \in [0, 1]$ as

$$\mathcal{T}|_\tau := \{t_i \in \mathcal{T} | t_{i-1} < \tau\}, \tag{5}$$

i.e., the truncation contains all points in time up to time τ *plus the next grid point of the initial grid* Λ^g (see Sec. II C 3). This procedure results in an initial grid $\mathcal{T}^{(0)} \subset \Lambda^g$ containing $|\mathcal{T}^{(0)}| = K$ points, with $K \leqslant 2^g$. Next, the algorithm performs bisections of this grid in successive iterations $\mathcal{T}^{(0)}, \mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(M)}$, where M is the total number of bisections before the routine terminates. Since each new bisection adds exactly one point to the grid, M also denotes the total number of points *added* to the initial grid. The final grid $\mathcal{T}^{(M)}$ contains $K + M$ points.

To each bridge (t_l, t_r) between left and right end points t_l and t_r and contained in a grid $\mathcal{T}^{(m)}$, we associate a level ℓ defined by $\ell = -\log_2(t_r - t_l)$. A bridge is bisected by introducing its midpoint $t_m = \frac{1}{2}(t_l + t_r) = t_l + 2^{-\ell-1}$ and inserting t_m into the grid $\mathcal{T}^{(m+1)} = \{t_1, \ldots, t_l, t_m, t_r, \ldots, t_N\}$. A bridge can be bisected until its level reaches a *maximum bisection level* L (typically $L \lesssim 30$ for $H = 0.33$). Since each iteration only halves an existing interval, all grids are subsets of the maximal dyadic lattice Λ^L:

$$\Lambda^g \supseteq \mathcal{T}^{(0)} \subset \mathcal{T}^{(1)} \subset \cdots \subset \mathcal{T}^{(M)} \subseteq \Lambda^L. \tag{6}$$

Note that for each bridge (t_l, t_r) there is always one dyadic lattice Λ^n, such that t_i and t_{i+1} are neighboring points in Λ^n; they are members, but not neighbors in $\Lambda^{n'}$ for $n' > n$; at least one of them does not exist in $\Lambda^{n'}$ for $n' < n$.

C. Definition of the algorithm

The algorithm consists of two phases. In the *first phase*, the initialization, a coarse grid is generated. In the *second phase*, the adaptive bisection, this grid is successively bisected where necessary. Once the second phase terminates, the first-passage time is calculated using the final grid.

The first phase starts by sampling an initial discretized path $\mathcal{X}^{(0)}$ over a dyadic lattice $\mathcal{T}^{(0)} = \Lambda^g$ with $N = 2^g$ equidistant points, using the *Davies-Harte* algorithm. The latter is the fastest known algorithm to sample an exact FBM path on an equidistant grid in time [25]; its execution time scales as $N \ln(N)$, thus only slightly slower than what is needed to generate an uncorrelated sample of the same length N. From this relatively coarse grid, $(\mathcal{X}^{(0)}, \mathcal{T}^{(0)})$, the first-passage time is estimated via linear interpolation as $\tau^{(0)}$.

Subsequently, the grid is truncated by discarding all points behind the first point surpassing m [see Eq. (5)]. That this does not change the measure is explained in Sec. II C 3. If no such point exists, the full grid is kept. The correlations between the different points X_t at times t stored in the grid are given by the correlation matrix

$$C_{ij}(\mathcal{T}) = C(t_i, t_j), \quad t_i, t_j \in \mathcal{T}, 1 \leqslant i, j \leqslant |\mathcal{T}|. \tag{7}$$

It is a symmetric matrix computed from the correlation function (3). It is then inverted to obtain the inverse correlation matrix $C^{-1}_{ij}(\mathcal{T})$. The inversion is optimized by using precalculated tabularized matrices. This concludes the *first phase*.

In the *second phase*, bridges are checked successively until the maximum level is reached. The order in which the bridges of the growing grid are checked is determined by a subroutine the aim of which is to find the first-passage event with the least amount of bisections. The check consists in testing whether the *midpoint* X_{t_m} of the bridge (t_l, t_r) could surpass the threshold m with a probability larger than ε, taken small. If this is the case the bridge is deemed *critical* and bisected. The bisection consists in generating a midpoint X_{t_m} at time t_m conditional to the preexisting grid. This computation requires the inverse correlation matrix and is detailed in Sec. II C 6. Once the midpoint is generated, it is added to the path $(\mathcal{X}, \mathcal{T})$. In a last step the inverse correlation matrix of the new grid $C^{-1}(\mathcal{T} \cup t_m)$ is stored. In Algorithm 1, the algorithm is given in pseudocode.

The routines in the pseudocode are described in Secs. II C 1–II C 7.

1. Davies-Harte algorithm

The DH algorithm is a widely used method to generate FBM samples. It was introduced in [20], is pedagogically described in [25], and has been extended to other Gaussian processes in [26], allowing us to omit an introduction. It generates a sample of fractional Gaussian noise (FGN) $\xi_1, \xi_2, \ldots, \xi_N$, the incremental process of FBM $\xi_j = X_{j+1} - X_j$, $j \in \mathbb{N}$, and then sums the increments to a FBM sample with values $X_{i\delta t} = (\delta t)^H \sum_{j=1}^{i} \xi_j$. Simulating the increments is more efficient since FGN is a stationary Gaussian process which, for equally sized increments, has a circulant correlation matrix, which can be diagonalized using a fast Fourier transform (FFT). Therefore a FGN sample of N increments can be simulated with computational complexity $O[N \ln(N)]$. The FFT algorithm works optimally when the number of points is a natural power of 2, i.e., if the grid is a dyadic lattice.

2. Estimating the first-passage time

Given a discretized path $(\mathcal{X}, \mathcal{T})$, we use its linear interpolation to give the first-passage time as its first intersection with the threshold (see Fig. 1).

3. Truncating the grid

A further optimization is to discard grid points beyond the first point crossing the threshold [see Eq. (5)]. It is necessary to show that the density of first-passage times conditioned on the full grid equals the distribution conditioned on the truncated grid, i.e., that truncating does not change the measure.

The FPTD $P(\tau)$ can be decomposed into a sum of conditional probabilities for disjoint events. Each term of the sum is the probability that the ith point of a grid surpasses m, the threshold, *for the first time* ["$P^{\text{grid}}(X_{t_i} > m$ first)"], times the FPTD of a FBM conditioned on the event that its

discretization on grid $\mathcal{T}$ surpasses m at t_i for the first time, i.e.,

$$P_{\mathcal{T}}(\tau|X_{t_i} > m \text{ first}) = P\big(\tau|X_t : X_{t_i} > m \text{ and } X_{t_j} < m\, \forall t_j < t_i\big) \quad (8)$$

for $t_j, t_i \in \mathcal{T}$. The decomposition thus reads

$$P(\tau) = \sum_{t_i \in \mathcal{T}} P_{\mathcal{T}}\big(\tau|X_{t_i} > m \text{ first}\big) P^{\text{grid}}\big(X_{t_i} > m \text{ first}\big). \quad (9)$$

By continuity of the process,

$$P_{\mathcal{T}}\big(\tau > t_i|X_{t_i} > m \text{ first}\big) = 0, \quad (10)$$

such that the sum in Eq. (9) can be truncated to

$$P(\tau) = \sum_{t_{i-1} < \tau} P_{\mathcal{T}}\big(\tau|X_{t_i} > m \text{ first}\big) P^{\text{grid}}\big(X_{t_i} > m \text{ first}\big). \quad (11)$$

In order to sample $P_{\mathcal{T}}(\tau|X_{t_i} > m \text{ first})$, one would naively sample the entire grid $\mathcal{X}$ over all of $\mathcal{T}$, but since

$$P_{\mathcal{T}}\big(\tau|X_{t_i} > m \text{ first}\big) = P_{\mathcal{T}|_\tau}\big(\tau|X_{t_i} > m \text{ first}\big), \quad (12)$$

where the restriction is defined in Eq. (5), it is sufficient to only regard the smaller grid $\mathcal{T}|_\tau$, i.e.,

$$P(\tau) = \sum_{t_{i-1} < \tau} P_{\mathcal{T}|_\tau}\big(\tau|X_{t_i} > m \text{ first}\big) P^{\text{grid}}\big(X_{t_i} > m \text{ first}\big). \quad (13)$$

Discarding points in the initial stage leads to a smaller correlation matrix to be inverted, which increases performance and decreases memory.

4. Tabulating inverse correlation matrices

The inverse of the correlation matrix (7) is necessary to compute the conditional probability of any further midpoint (see Appendix B). Its computation is costly and typically scales with $O(N^3)$ where $N = 2^g$ is the number of points in $\mathcal{T}^{(0)}$. If the algorithm is run multiple times, this computation slows it down. The initial grid, however, is always a dyadic lattice truncated at some point, i.e., $\mathcal{T}^{(0)} = \{k2^{-g}; 0 \leqslant k \leqslant K\}$, where $X_{K2^{-g}}$ is the first point to surpass m. Therefore, the initial inverse correlation matrix $C^{-1}(\mathcal{T}^{(0)})$ can take $2^g - 1$ possible values, one for each possible value of K. It is more efficient to precalculate all possible inverse correlation matrices in the beginning, and store them in a vector "CMatrixTable":

$$\text{CMatrixTable}[K] = \big([C(i2^{-g}, j2^{-g})]_{i,j=1}^{K}\big)^{-1}. \quad (14)$$

After generating the initial grid and measuring $\tau^{(0)}$, one reads out the appropriate entry of the table at $K = \min\{n \in \mathbb{Z}; n2^{-g} \geqslant \tau^{(0)}\}$.

5. Deciding whether a bridge is critical

Once entering the bisection phase, the algorithm needs to decide whether a particular bridge is *critical*, i.e., whether it is suspicious of hiding a "dangerous" excursion crossing the threshold at m (see Fig. 1). Rather than determining whether *any* point in (t_l, t_r) surpasses the threshold, we focus on the midpoint t_m conditioned on all other points $\mathcal{X}$, and ask how likely $X_{t_\text{m}} > m$. Such an event needs to be avoided with a very low probability ε, the *error tolerance*. The relevant probability,

$$P\big(X_{t_\text{m}} > m|\mathcal{X}\big) < \epsilon, \quad (15)$$

is too costly to be computed for every bridge in every step of the iteration, as the midpoint is a Gaussian random variable, with its mean and variance determined by every other point in the grid. If we *ignore* all points of the path apart from (t_i, X_{t_i}) and $(t_{i+1}, X_{t_{i+1}})$, a calculation given in Appendix A shows that mean and variance would be given by

$$\mu = \tfrac{1}{2}\big(X_{t_i} + X_{t_{i+1}}\big) \quad (16)$$

and

$$\sigma^2 = \big(2^{1-2H} - \tfrac{1}{2}\big)2^{-2\ell H}. \quad (17)$$

Here ℓ is the level of the bridge of width $\delta t = 2^{-\ell}$. Interestingly, adding to the bridge's end points further *lowers* the variance [see Eq. (30)], which means that neglecting all but nearest neighbors gives an upper bound on the variance of the midpoint. Further, we conservatively bound the mean by the maximum of both end points, $\mu \lesssim \max(X_{t_i}, X_{t_{i+1}})$. This is *a priori* not a precise approximation, since far-away grid points are able to "push" the expected midpoint above the bridge's end points for values of $H \neq \frac{1}{2}$. As is shown in Sec. III C, this systematic error can be absorbed by introducing an even smaller error tolerance ε'. Furthermore, it is less relevant in the subdiffusive regime, where the process is negatively correlated. By giving conservative bounds on mean and variance with quantities that are *local* (i.e., do not depend on the remaining grid), we can replace the original criterion (15) by a computationally cheaper alternative, namely, the local condition

$$P\big[X_{t_\text{m}} > m\big|\big(X_{t_\text{l}}, X_{t_\text{r}}\big)\big] < \epsilon'. \quad (18)$$

This implies that Eq. (15) holds for an appropriate choice of ε', on *average*. This is to be understood as follows. In a simulation, there are n decisions of type (15) to be taken. The total error is approximately $P_\text{error}^\text{tot} \approx n\epsilon$. The parameter ϵ' is chosen such that the total error rate remains smaller than 10^{-6}, and is thus negligible as compared to Monte Carlo fluctuations. The dependence between ϵ' and $P_\text{error}^\text{tot}$ is investigated in Sec. III C (see Fig. 3).

Criterion (18) is rephrased, using again ℓ as the level of the bridge, to

$$\Phi\left(\frac{m - \max\big(X_{t_i}, X_{t_{i+1}}\big)}{\big(\sqrt{2^{1-2H} - \frac{1}{2}}\big)2^{-\ell H}}\right) > 1 - \varepsilon', \quad (19)$$

which implies

$$\max\big(X_{t_i}, X_{t_{i+1}}\big) < m - \big(\sqrt{2^{1-2H} - \tfrac{1}{2}}\big)2^{-\ell H}\Phi^{-1}(1-\varepsilon'), \quad (20)$$

where we introduced Φ, the cumulative distribution function of the standard normal distribution, and its inverse Φ^{-1}. This is further simplified by defining the *critical strip*

$$c_0 = \big(\sqrt{2^{1-2H} - \tfrac{1}{2}}\big)\Phi^{-1}(1 - \varepsilon') \quad (21)$$

and the level-corresponding critical strips

$$c_\ell = 2^{-\ell H} c_0. \quad (22)$$

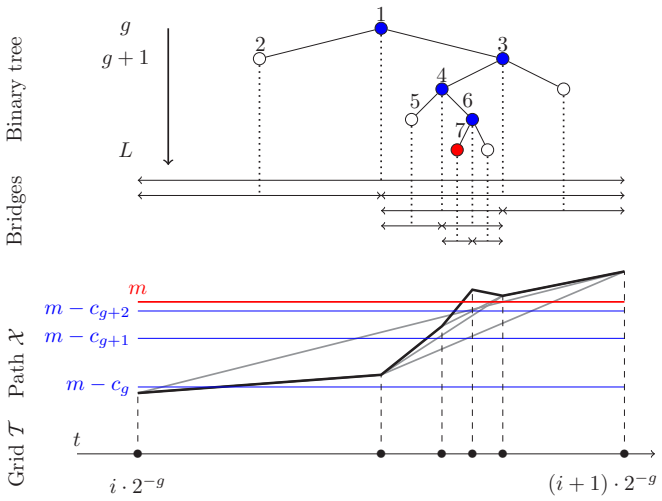

FIG. 2. Illustration of the adaptive bisection routine. The grid $\mathcal{T}$ (bottom) contains points in time; here details are shown of the initial bridge $t_\mathrm{l} = i2^{-g}, t_\mathrm{r} = (i+1)2^{-g}$ (labeled bullets) and successively introduced midpoints (bullets on the time axis). The path $\mathcal{X}$ (above) samples values at times (dashed lines) which approximate the path by linear interpolations (gray and black thick lines). The threshold m (red uppermost horizontal line) is crossed by the path and bisections are generated for every bridge the end points of which lie in the critical strip corresponding to its level (blue vertical lines underneath). The horizontal arrows on top of the path indicate the bridges in between the grid points. The mapping from bridges to the binary tree (top) is indicated with dotted lines. The top node (1) corresponds to the widest bridge $(i2^{-g}, (i+1)2^{-g})$, and children correspond to subintervals generated by the midpoint. The bridges are explored in order as given by numbers above nodes and chosen by the bridge-selection routine (see text for details). Bridges that are critical (blue filled nodes) are bisected, and their children are checked from left to right, until a first-passage event has been identified at maximum bisection level L (red filled node 7). This event terminates the algorithm. In contrast to node 1, which belongs to the initial grid $\mathcal{T}^{(0)}$, nodes 2 to 7 stem from adaptive bisections and contribute to the total count of bisections M. The maximum number of nodes which could theoretically be spawned off this particular subinterval is $2^{L-g} - 1$.

A bridge $(X_{t_\mathrm{l}}, X_{t_\mathrm{r}})$ of level ℓ is deemed critical if either of its end points lies above the critical strip corresponding to ℓ, i.e.,

$$\max\left(X_{t_i}, X_{t_{i+1}}\right) > m - c_\ell. \tag{23}$$

This makes for a computationally fast decision process, since the critical strip width has to be computed only once. The procedure then checks for a given level of the bridge whether it reaches into the critical strip, in which case it is bisected (see Fig. 2 for illustration).

6. Generating the new midpoint efficiently

If a bridge triggers a bisection, the midpoint is drawn according to its probability distribution, given all points that have been determined previously. If this occurs at, say, the mth iteration, the discretized path is $((X_1, t_1), \ldots, (X_N, t_N))$ with $|\mathcal{T}^{(m)}| = |\mathcal{X}^{(m)}| = N = K + m$ where $K \leqslant 2^g$ is the number of points in the truncated initial grid. Denoting the midpoint

ALGORITHM 1: Adaptive bisection

procedure ABSEC(g, L, m, ε)
  $\mathcal{T} \leftarrow \Lambda^g$
  $\mathcal{X} \leftarrow$DAVIES_HARTE(Λ^g) ▷ II C 1
  $\tau \leftarrow$ FPT_FROM_GRID($\mathcal{X}, \mathcal{T}$) ▷ II C 2
  $(\mathcal{X}, \mathcal{T}) \leftarrow (\mathcal{X}, \mathcal{T})|_{\tau^{(0)}}$ ▷ II C 3
  $C^{-1} \leftarrow$ `CMatrixTable`$[\tau^{(0)}]$ ▷ II C 4
  $(t_\mathrm{l}, t_\mathrm{r}) \leftarrow$NEXT BRIDGE($\mathcal{T}, 0, \tau^{(0)}$)
  while $(t_\mathrm{l}, t_\mathrm{r})$ defined **do**
    if Bridge $(t_\mathrm{l}, t_\mathrm{r})$ critical and not yet bisected **then** ▷ II C 5
      $C^{-1} \leftarrow$ AUGMENT C^{-1}-MATRIX(C^{-1}, t_m)
      $X^* \leftarrow$ GENERATE MIDPOINT(C^{-1}, t_m) ▷ II C 6
      $\mathcal{X} \leftarrow \mathcal{X} \cup X^*$
      $\mathcal{T} \leftarrow \mathcal{T} \cup t_\mathrm{m}$
      if $X^* > m$ **then**
        $\tau \leftarrow$ FPT_FROM_GRID($\mathcal{X}, \mathcal{T}$)
    $(t_i, t_{i+1}) \leftarrow$NEXT BRIDGE($\mathcal{T}, (t_i, t_{i+1}), \tau$) ▷ II C 7
  output(τ)

to be inserted by (X_{t^*}, t^*), one needs to find

$$P(X_{t^*}|X_1, \ldots, X_N). \tag{24}$$

The midpoint is again normal distributed with mean $\mu_*(N)$ and variance $\sigma_*(N)$. Let $\mathcal{T}^{(m+1)} = (\mathcal{T}^{(m)}, t^*)$ be the augmented grid, and $C^{-1}(N+1) = C^{-1}(\mathcal{T}^{(m+1)})$ the associated inverse correlation matrix [see Eq. (7)]. Then, as detailed in Appendix B, the inverse of the variance is given by

$$\sigma_*^{-2}(N) = [C^{-1}(N+1)]_{N+1,N+1} \tag{25}$$

and the mean by

$$\mu_*(N) = \sigma_*^2(N) \sum_{i=1}^{N} [C^{-1}(N+1)]_{N+1,i} X_{t_i}. \tag{26}$$

Computing the inverse correlation matrix from scratch at every iteration would require a matrix inversion which typically uses $O(N^3)$ steps. We do this in $O(N^2)$ steps, by starting from the already calculated inverse correlation matrix of the previous grid $C^{-1}(N) = C^{-1}(\mathcal{T}^{(m)})$. As detailed in Appendix C, the inverse correlation matrix $C^{-1}(N+1) = C^{-1}(\mathcal{T}^{(m+1)})$ can be constructed as follows. First, generate a vector containing all correlations of the new point with the grid, using Eq. (3):

$$\vec{\gamma}(N) = (C(t^*, t_1), C(t^*, t_2), \ldots, C(t^*, t_N))^T. \tag{27}$$

Second, multiply it with the (already constructed) inverse correlation matrix to obtain

$$\vec{g}(N) = C^{-1}(N) \cdot \vec{\gamma}(N). \tag{28}$$

In terms of $\vec{\gamma}$ and $\vec{g}$, the mean and variance can be expressed as

$$\mu_*(N) = \vec{X} \cdot \vec{g}, \tag{29}$$

where we use $\vec{X} = (X_{t_1}, \ldots, X_{t_N})$ for short, and

$$\sigma_*^2(N) = 2(t^*)^{2H} - \vec{\gamma} \cdot \vec{g}. \tag{30}$$

Since $\vec{\gamma}^T \vec{g} = \vec{\gamma}^T C^{-1}(N) \vec{\gamma} > 0$, conditioning on more points diminishes the variance of a midpoint. The outer product of $\vec{g}$

defines the matrix

$$G(N) := \vec{g} \otimes \vec{g}^T. \tag{31}$$

It is used to build the enlarged inverse correlation matrix

$$C^{-1}(N+1) = \left(\begin{array}{c|c} C^{-1}(N) + \sigma^{-2} G(N) & -\sigma^{-2}\vec{g}(N) \\ \hline -\sigma^{-2}\vec{g}^T(N) & \sigma^{-2} \end{array}\right), \tag{32}$$

where $\sigma^2 = \sigma_*^2(N)$. In our implementation entries in $\vec{\mu}, \vec{g}, C^{-1}$, etc., are generally not in order of time but in order of their addition to the grid.

7. Bridge selection

The task of the bridge-selection routine (see Algorithm 2) is to choose the order in which bridges of the successively refined grid are tested, and possibly inserted. Its aim is to find the first-passage event with the least number of bisections. To this aim, it zooms in into areas where a first-passage time is likely, and zooms out when the possibility of a crossing becomes negligible. In this subsection, we phrase this intuition in more rigorous terms.

Prior to the first call of the routine, the initial grid consists of 2^g bridges of uniform width 2^{-g}. The routine selects the earliest bridge, i.e., $(t_\mathrm{l} = 0, t_\mathrm{r} = 2^{-g})$, and scans all bridges of the initial grid in ascending order in time until a critical bridge is found [by applying the criticality criterion (18)]. Once such a bridge is found, the algorithm explores this bridge by successive bisections. After a finite number of bisections the algorithm either has identified a first-passage event to the desired precision or no crossing was found. In the latter case, the routine then moves on to the next bridge of the initial grid.

In order to illustrate the workings of the bridge-selection routine, it is helpful to consider a bijection between the adaptively bisected grid and a rooted binary tree (see Fig. 2). Every bridge $(t_\mathrm{l}, t_\mathrm{r})$ that is bisected by introducing a point at t_m contains two sub-bridges $(t_\mathrm{l}, t_\mathrm{m})$ and $(t_\mathrm{m}, t_\mathrm{r})$. We refer to these bridges as the *left* and *right* children of $(t_\mathrm{l}, t_\mathrm{r})$. Vice versa, every bridge that is not part of the initial bridge (i.e., with level $\ell > g$) is the child of another bridge which is referred to as the parent of the bridge. The set of all bridges that are contained in an initial bridge of width 2^{-g} is mapped to a rooted binary tree by identifying every node with a bridge, where a node can either have zero or two children depending on whether the bridge has been bisected or not. The root of the tree corresponds to the bridge contained in the initial bridge from where the bisections were spawned off. The generation of a node in the tree corresponds to its level by generation $= \ell - g + 1$. Therefore, the depth of the tree is limited to $\mathrm{generation}_\mathrm{max} = L - g + 1$.

The routine stores a representation of this tree internally, together with the information as to whether a node or bridge has previously been checked for criticality or not. If a bridge is bisected, but its two children have not yet been checked for criticality, the left child is selected. This is because earlier crossings of the threshold render later crossings irrelevant. If a bridge has two children, but the left has already been checked (implying that neither it nor any of its further descendants contains a first-passage event), the right child is selected. If both children of the bridge have already been checked, none of the descendants contains a first-passage event. In that case the parent of the bridge is returned (zooming out). If the routine returns to the root, the bridge of type $(i2^{-g}, (i+1)2^{-g})$ has no parent, and the next such bridge $((i+1)2^{-g}, (i+2)2^{-g})$ is returned. If $i = 2^g - 1$, the routine terminates by returning an empty bridge since the entire grid has been checked. This selection routine implies that in the "worst case," when every point of the initial grid $\mathcal{T}^{(0)}$ lies in the critical strip without ever crossing the threshold m, every single subinterval will be analyzed. This means that in a worst-case scenario up to 2^g search trees (see Fig. 2) would be generated, each tree containing up to $2^{L-g} - 1$ nodes. This scenario is extremely unlikely.

To summarize, the routine is either descending (zooming in) or ascending (zooming out) within the tree, depending on whether the children of a node, if existent, have been visited or not.

The routine takes into account two additional constraints. First, concerning the maximum bisection level L, if a bridge of maximum level L contains a first-passage event, the routine terminates since this estimate has reached the desired resolution. If it contains no crossing, the parent is returned. Second, it takes into account whether a bridge is early enough in time to improve the first-passage estimate. If a bridge at level ℓ records a first-passage event, only its descendants can improve this result.

We give the pseudocode of the routine below. In the implementation we present later (Sec. III A), the algorithm is implemented slightly differently for performance reasons. The logical steps, however, are the same and we decided to present them here for pedagogical reasons.

D. Adding deterministic functions

The adaptive bisection routine can be adapted to further generate first-passage times of stochastic processes of the form

$$Z_t = X_t + f(t), \tag{33}$$

where $f(t)$ is a deterministic smooth function, e.g., a linear or fractional drift term, and X_t is again a fractional Brownian motion. In its first phase, X_t is generated on a subgrid, and $f(t)$ is added accordingly. The resulting process $Z_t, t \in \mathcal{T}^{(0)}$ is then passed to the bridge-selection routine, where the bridges are checked for criticality using the values of Z_t in the criticality criterion (20). Once a bisection is required, the midpoint is generated using the subtracted process $X_t = Z_t - f(t)$, i.e., the vector used to generate the midpoint's mean [see Eq. (26)] is $\vec{X}$, not $\vec{Z}$. Then, the generated midpoint X_m is transformed back using $Z_\mathrm{m} = X_\mathrm{m} + f(t_\mathrm{m})$, and inserted into the path of Z. Note that even if $f(t) = \mu t$ (linear drift), and contrary to Brownian motion, the iteration cannot be performed directly on Z_t.

E. Further generalizations

The underlying idea of the algorithm—to generate a grid that is fine only where it matters—lends itself to various other nonlocal observables, in particular extreme events, such as running maxima (minima), positive time (time spent in the

region $X_t > 0$), last returns, or the range or span ($\max X_t - \min X_t$) of a process.

In each of these examples, one needs to adapt two logical steps in the procedure: (i) the order in which bridges of the grid are iterated and (ii) the criterion for triggering a bisection. For first-passage times, the order of the bridges is given by the subroutine described above in Sec. II C 7. The criterion for bisection is determined by the bridge's distance to the threshold. These two choices are particular to first-passage events.

For running maxima, the bridges should be tested in descending order of height, and the bisection criterion adapted to decide whether the midpoint could surpass the current maximum with a probability larger than ε. If the current maximum changes, the criterion for triggering a bisection also changes. As the maximum can only increase, bridges which were uncritical before do not become critical by a change of the estimate of the maximum.

To find the last return to zero ($t_0 = \sup_{t'<t}\{t'|X_{t'} = 0\}$), the bisection criterion is the same as for first-passage times (with m set to zero), but bridges should be iterated over from latest to earliest, choosing the right subinterval first after bisection (see Fig. 2).

The span of a process at time t is defined as the running maximum minus the running minimum [8,30–33]. To find the first time the span reaches one is more delicate. There are two cases, given a discretization. Either span 1 is reached first when the maximum increases, or the minimum decreases. Suppose that the maximum increases. Then there is a minimum for a smaller time. By refining the grid close to this minimum, the latter may decrease. This in turn shifts down the critical strip for the maximum, and one has to redo all checks for bridges close to the maximum.

The algorithm can be generalized to other Gaussian processes, since the derivations given in Sec. II C 6 and Appendix B for the insertion of a conditional midpoint apply to *any* Gaussian process. The only point at which we made explicit use of properties for FBM was at the initialization step, where the Davies-Harte method was employed to generate a path on a coarse dyadic lattice. If one were to study another Gaussian process, one would need to replace the correlation function (3), and adapt the routine generating the initial grid.

Once these modifications are made for the new problem, we expect the algorithm to deliver similar improvements in performance and memory.

III. RESULTS AND BENCHMARKING

In this section, we compare an implementation of our adaptive bisection method (ABSec) with an implementation of the Davies-Harte (DH) method. Our focus lies on comparing both CPU time and memory usage for a simulation of equal discretization error. We find that for large system sizes, $N_{\text{eff}} \gtrsim 10^{2/H}$, the adaptive bisection routine outperforms the Davies-Harte method both in CPU time and memory. This advantage grows markedly for lower values of H. At $H = 0.33$, for instance, and a final grid size of $N_{\text{eff}} = 2^{32}$ we need 5000 times less CPU time and 10 000 less memory. At $H = 0.25$ we find ABSec to be 300 000 times faster and 10^6 less memory intensive than DH at an effective system size of $N_{\text{eff}} = 2^{42}$.

ALGORITHM 2: Finding the next bridge to be checked.

```
procedure NEXT BRIDGE(T, (t_l, t_r), τ_m)
    if (t_l, t_r) = 0 then
        return (0, 2^{-g})                    ▷ Initialise with first bridge
    if (t_l, t_r) has no children then
        return parent bridge
    if (t_l, t_r) early enough for τ_m AND level < L then
        if left child not checked then
            return left child                 ▷ Move down left
        if left child checked AND right child not checked then
            return right child                ▷ Move down right
        if both children checked then
            return parent bridge              ▷ Move up to parent
    if level of (t_l, t_r) = L then
        if Bridge crosses threshold then
            return NULL
        else
            return parent bridge
```

We then discuss systematic errors and analyze how they depend on the parameters, in order to clarify the payoff between computational cost and numerical accuracy. We conclude with a discussion of our findings.

A. Implementation in C

We implemented the adaptive bisection algorithm in C, using external libraries LAPACK [34], GSL [35], FFTW3 [36], and CBLAS [37]. The code is published [28] and available under a BSD license. It was compiled using the Clang/LLVM compiler using the −03 flag as the only compiler optimization. The code was executed on an Intel(R) Core(TM) i5-7267U CPU 3.10-GHz processor.

As reference, we use an implementation of the Davies-Harte method in C [38]. Compiler settings and hardware are identical to those used for the adaptive bisection algorithm.

In order to compare performance, we used user time and maximum resident set size as measured by the POSIX command GETRUSAGE; user time indicates the time the process was executed in user space, and maximum resident set size indicates the amount of RAM held by the process.

B. Numerical errors and fluctuation resolution

The adaptive bisection algorithm suffers from three errors.

(i) The resolution of the grid itself is determined by the maximum grid size if all bridges were triggered, which we refer to as *horizontal error*. Any discretization of a continuous path suffers from errors that are made when replacing the rough continuous path by the linear interpolation of a grid. Even if the true first-passage time is optimally approximated, the error still depends on the system size N. In that respect, our algorithm does not differ from DH or other exact sampling methods.

(ii) The adaptive bisection routine suffers from a probabilistic error, namely, false negative results of the criticality check, i.e., bridges which do contain an excursion crossing the threshold m, but the end points of which do not lie in

the critical strip (see Sec. II C 5). We refer to these errors as *vertical errors*.

(iii) The algorithm suffers from rounding errors of the floating-point unit.

Horizontal errors correspond to the resolution of the process's fluctuations. To contain fluctuations of a FBM between two grid points at distance N^{-1} to the order of δX, one needs to choose $N \sim (\delta X)^{-\frac{1}{H}}$. Horizontal errors are therefore characterized by the effective discretization resolution $N^H \sim (\delta X)^{-1}$ which corresponds to the inverse fluctuation resolution. In order to compare two discretizations of a FBM path for two different values of the Hurst parameter H, comparing N is misleading. Rather, we compare their effective discretization resolutions N^H. Horizontal errors are impossible to measure numerically, since there exists no way to simulate a continuous path. They are, however, independent of the sampling method used; this implies that the horizontal error of a path generated by DH with system size 2^L and an adaptive bisection routine of maximum bisection level L are *exactly* the same, given no vertical error occurred. For a deeper discussion of discretization errors of the DH algorithm, see Sec. V E of [33].

Vertical errors are controlled by the error tolerance ε', of Eqs. (20) and (21). To study vertical errors systematically, one needs to compare the results with a fully sampled grid using (for instance) DH. This is discussed in the next section.

In the remainder of the section, we run benchmarking experiments that repeat the adaptive bisection routine a large number of times, typically $I = 10^4$. Following the insights of Sec. III C, we choose an error tolerance that is small enough to neglect errors of the vertical kind (whenever the vertical error rate is much smaller than I^{-1}). In doing so, we can ignore the vertical error such that the numerical discretization error becomes a good common error for both adaptive bisections and DH. This allows us to compare grids sampled with both methods systematically across various values of H and L.

Finally, errors due to the finite precision of the floating-point unit are considered. These arise in the matrix inversion (32), where inspection reveals terms of opposite sign. They can be detected by plotting $\sigma_*^2(N)$ as a function of grid resolution. For small grids, $\sigma_*^2(N)$ almost follows a power law, with little spread. Numerical errors are visible as a net increase of this spread (see Fig. 10). To be on the safe side, we choose the maximal L to be four less than the point where we first see numerical errors appear.

C. Error rate depending on ε'

This section addresses the question of vertical errors, i.e., bridges that were deemed uncritical by the adaptive bisection routine (see Sec. II C 5), yet contained an excursion that crossed the threshold for the first time. This probability, $P(X_{t_m} > m)$, where X_{t_m} marks the midpoint of a bridge, was bounded using an error tolerance ε'. Therefore, we need to know how ε' controls the error rate. Since we can only measure the error rate when compared to another numerically generated grid, we compare our algorithm to a path generated using the Davies-Harte algorithm of equal precision. The procedure is as follows. In a first step, the Davies-Harte method is used to generate a path on the dyadic lattice Λ^L. For this path, and a threshold m, the first-passage time is calculated using its linear interpolation as detailed in Sec. II C 2. Then, only times in the subgrid $\Lambda^g \subset \Lambda^L$ are copied into a second path. This path is handed over to a modified adaptive bisection routine (see Algorithm 1). The bridges of the grid are successively checked, at each step deciding whether to bisect as discussed in Sec. II C 5. Once a midpoint needs to be drawn, it is not randomly generated, but taken from the full grid at the same time. The full grid thus serves as a *phone book* for the adaptive bisection algorithm, where points are *looked up* if they lie at points the algorithm would have otherwise generated randomly. The algorithm then outputs its own estimate of the first-passage time. If the first-passage times disagree, this is considered an error. We refer to this check as the *phone book test*. This test is iterated 10^6 times, and the error rate $P_{\text{error}}^{\text{tot}}$ is defined as the ratio between errors and the number of iterations.

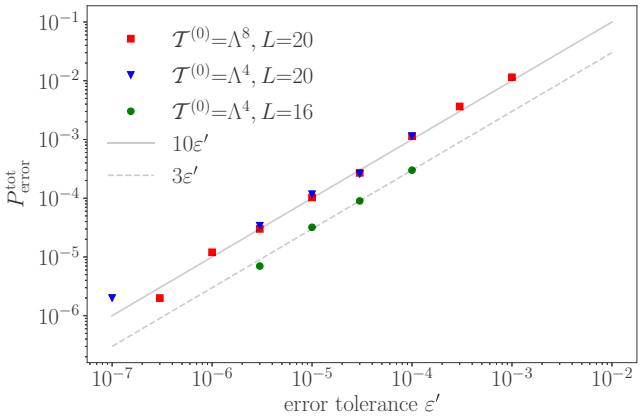

FIG. 3. Error rate from the phone book test for various values of ε' for $H = 0.33$. The error rate is almost identical when changing the initial grid size from 2^8 (red square marks) to 2^4 (blue triangle marks) at the same maximum bisection level $L = 20$. When lowering the maximum bisection level to $L = 16$, the error rate improves. The relation between error rate and error tolerance decreases approximatively linearly over several orders of magnitude (compare with solid gray line). The total error rate is approximately $10\varepsilon'$ for $L = 20$ (solid gray line) and about $3\varepsilon'$ for $L = 16$ (dashed gray line). Note that the prefactor is much smaller than the number of points, which can be read off from Fig. 5. Error rates were averaged over 10^5 to 10^6 iterations.

The results are shown in Fig. 3, where we compare the error rate for different values of ε' and for three different grids of varying initial grid size, and maximum bisection levels. The plot shows that the total error rate and error tolerance ε' depend on each other linearly, indicating that ε' is a suitable replacement for ϵ introduced in Eq. (15). The plot further shows that the error rate remains almost identical when replacing the initial grid Λ^8 by Λ^4 (which contains 16 points only). Further, the error rate improves if the maximum bisection level is lowered. When lowering the effective system size from 2^{20} to 2^{16}, the error rate lowers approximately by a factor of 3.

In summary, this plot confirms that the computationally cheap variant (18) allows us to control the vertical errors (false negative results of the criticality test).

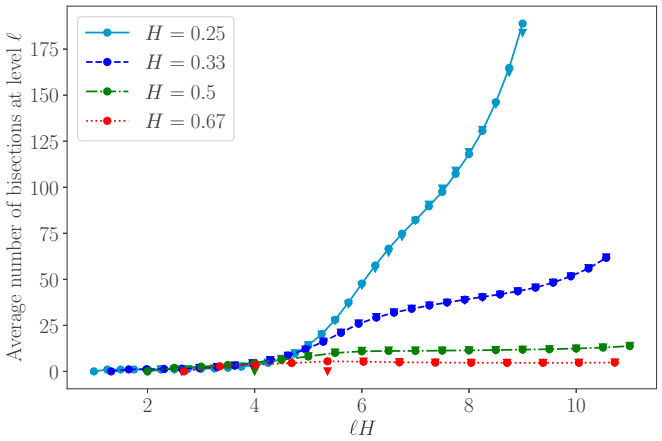

FIG. 4. Average number of new midpoints generated at bridge level ℓ, for various values of H (solid, dashed, dash-dotted, and dotted lines) as a function of ℓH. For equal values of ℓH, a lower Hurst parameter implies a larger number of average bisections. These numbers are virtually independent of the initial grid size, as is shown for Λ^4 (circle marks) and Λ^8 (triangle marks).

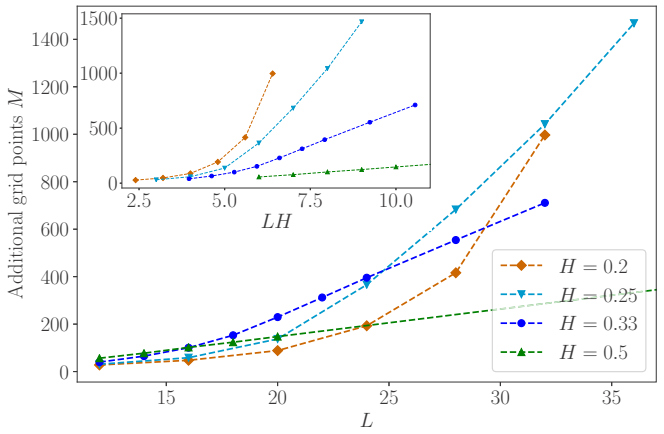

FIG. 5. Average number of bisections M as a function of the maximum bisection level L (i.e., $N_{\text{eff}} = 2^L$) for different values of H (diamond, circle, upright, and upside down triangle marks). Inset shows M vs LH. As long as $H \geqslant 0.33$ growth is asymptotically approximately linear in L, corroborating $M \sim \ln(N_{\text{eff}})$. For smaller values of H, either the linear regime is not yet reached, or the growth is stronger (5000 iterations with initial grid Λ^8 and error tolerance $\varepsilon' = 10^{-9}$). For $H = 0.5$, extrapolation was used.

D. Average number of bisections

In this section, we investigate how many points are added to the initial grid, and how the additionally inserted midpoints are distributed over the different generations. The number of midpoints generated, M, is the main expense of computational resources, since each point requires promoting an inverse correlation matrix from size n to $n+1$ requiring $O(n^2)$ steps.

Each midpoint that is generated bisects a bridge at level ℓ and creates two sub-bridges at level $\ell + 1$. In order to know how the algorithm spends most of its time, we simulated the adaptive bisection routine 10^4 times over an initial grid of size Λ^4 or Λ^8 and measured the average distribution of the M newly generated midpoints over the different levels. The results are shown in Fig. 4.

While the distribution remains virtually unchanged when replacing the initial grid by Λ^8, its shape changes for different values of Hurst parameter H. For $H > \frac{1}{2}$, the distribution remains flat and even descends for $\ell > 5/H$. For $H = \frac{1}{2}$ it remains constant for $\ell > 8$ (at around 11 midpoints per generation), while for $H < \frac{1}{2}$ (see figure for $H = \frac{1}{3}$ and $\frac{1}{4}$) the number of inserted midpoints increases, and tends to be at higher bridges.

Since the number of additionally inserted points M is crucial to the performance of ABSec, the routine is designed to minimize this number, with a hypothetical minimum of $L - g$ points (when finding the first-passage event with no fault). The hypothetical maximum corresponds to a full bisection of the grid which would require $2^L - 2^g \approx 2^L$ additional points (this occurs when the path does not cross the threshold at all and $\varepsilon' \to 0$). In Fig. 5, we show the total number of bisections M for various system sizes L, averaged over 10^4 realizations. The number of additional points ranges from 40 to 1500, where larger system sizes lead to an increase of M. For $H = 0.33$ and $L = 32$, the average of additional points is $M = 710$, which corresponds to 1.6×10^{-7} of the full grid. This means that, with that fraction of the full grid only, the algorithm identifies the first-passage time to the same accuracy as DH (up to vertical errors controlled by $\varepsilon' = 10^{-9}$ in this case).

We observe that for values of $H \gtrsim \frac{1}{3}$ the number of bisections grows first sublinearly and then linearly in L. This behavior changes for values $H \lesssim \frac{1}{4}$, where growth is stronger, and we may not yet be in the asymptotic regime. This is also indicated by the profiles shown in Fig. 4, where for lower values of H the distribution ceases to tend to a plateau, but grows for higher levels of bisection ℓ.

E. Computing time and complexity estimate

In this section, we analyze how the performance of our algorithm varies with different parameters, and how it compares to DH. In loose terms, we expect the initial grid, generated by DH, to cost $O[2^g \ln(2^g)]$, and each of the M bisections to cost k^2 with k, the number of grid points, i.e., costs, or more precisely the algorithmic complexity, should behave as

$$\mathcal{C}^{\text{ABSec}}(g, M) \sim \sum_{k=2^g}^{2^g+M} k^2 \approx \frac{1}{3}(2^g + M)^3. \tag{34}$$

It is therefore evident that the majority of the computational cost lies in the bisection phase, and the overall complexity is of order $O[(2^g + M)^3]$. When comparing this to the complexity of generating 2^L grid points with DH, which is $O[2^L \ln(2^L)]$, one estimates that ABSec outperforms DH whenever $M^3 \ll 2^L \ln(2^L)$. As is shown below, ABSec outperforms DH for $L \gtrsim 12$ to 16.

We define the performance of the algorithm via its user time, i.e., the share of the CPU time the process spends in user space. This means that, depending on the implementation, the total of CPU time ("wall time") might differ. User time is a more robust observable, so we use it as the best approximation to the performance of the implementation.

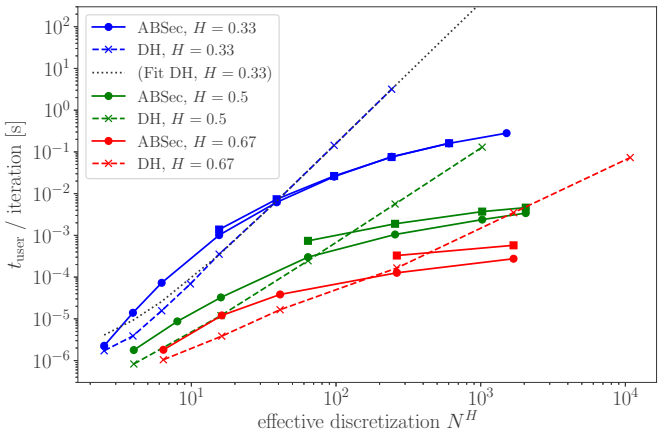

FIG. 6. Average user time required to find the first-passage time in a grid of effective discretization precision 2^{-LH}. The dashed lines indicate user time for Davies-Harte method, and solid lines are for the adaptive bisection method. The three different colors indicate $H = 0.33, 0.5, 0.67$ (bottom, center, top pairs of lines). Simulations were run 10^4 times for $\varepsilon' = 10^{-9}$ and for two different initial subgrid sizes (Λ^4, circle marks; Λ^8, square marks). For $H = 0.33$ (top solid blue lines), the effective system sizes range from $L = 4$ to 32 for Λ^4, and $L = 12$ to 28 for Λ^8. For $H = 0.5$ (center solid green lines), L ranges from 4 to 22 for Λ^4 and from $L = 12$ to 22 for Λ^8. For $H = 0.67$ (bottom solid red lines), L ranges from 4 to 16 for Λ^4 and 12 to 16 for Λ^8.

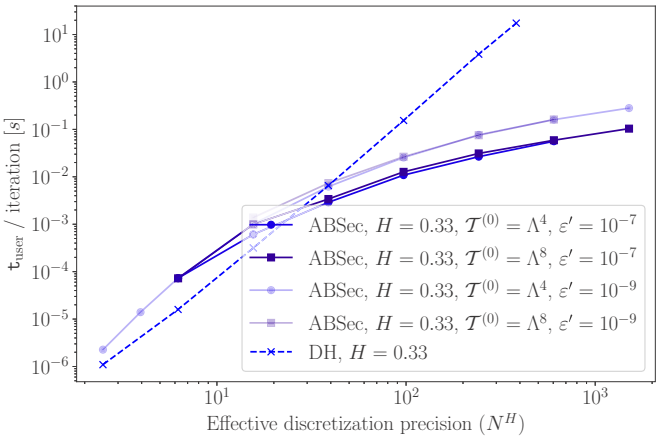

FIG. 7. $(t_{\text{user}}/\text{iteration})^{1/3}$ plotted vs effective discretization N^H for various values of H (blue circle marks $H = 0.33$, green square marks $H = 0.5$, red diamond marks $H = 0.67$; see Fig. 6). They corroborate the estimate of $\mathcal{C}^{\text{ABSec}} \sim (\ln N_{\text{eff}})^3$. Straight lines indicate fits of the form $a \ln(N) + b$ vs $t_{\text{user}}^{1/3}$ implying a scaling of $t_{\text{user}} \sim a^3[\ln(N_{\text{eff}})^3] + O[(\ln N_{\text{eff}})^2]$. This is in agreement with the complexity estimate in Eq. (35). The inset shows the ratio between data points and the fit.

We measure the average user time per generated first-passage time, using either DH or ABSec. To render different Hurst values and algorithms comparable, we plot the user time versus the inverse of the effective discretization error, which scales as N^H for DH and 2^{LH} for ABSec. It describes how well the FBM path is resolved numerically, taking into account the fluctuation scaling for different Hurst parameters.

Since at the beginning of the ABSec procedure inverse correlation matrices are tabulated (see Sec. II C 4), we measured the run time for 10^4 iterations, in order to render the initial overhead irrelevant.

Figure 6 shows the result of the benchmarking. For small effective system sizes, ABSec performs slower than DH, which is due to the relatively complex overhead of bisections. For (effective) system sizes of $N \gtrsim 10^{\frac{2}{H}}$ the ABSec algorithm gains an increasing and significant advantage since its run time only grows sublinearly.

To estimate the performance time, we observe that for values of $H \geqslant 0.33$ the number of additional grid points M grows linearly in the logarithm of the effective system size (see Fig. 5) throughout the entire observed range. Based on our empirical findings, we propose a linear relation $M \sim L = \log_2(N_{\text{eff}})$, which implies [see Eq. (34)] an overall computational complexity of

$$\mathcal{C}^{\text{ABSec}}(N_{\text{eff}}) = O[(\ln N_{\text{eff}})^3], \quad H \gtrsim \tfrac{1}{3}, \tag{35}$$

since eventually $M \gg 2^g$ for N_{eff} large enough (see Fig. 5 for $H \gtrsim \frac{1}{3}$). This estimate is corroborated by Fig. 7, where the scaling of user time with system size agrees with our estimate of $(\ln N_{\text{eff}})^3$ for sufficiently large system sizes. The linear relation between the number of bisections M and the logarithmic system size L, however, does not extend to smaller values of H, where Fig. 5 indicates superlinear growth. Still, testing the ABSec routine at $H = 0.25$ for an effective system size of $N_{\text{eff}} = 2^{42}$ gave an average user time of 6.2 s and was about 300 000 faster than an extrapolation of the user time for DH at the same system size.[1] This shows that for all practical purposes ABSec remains a much faster algorithm even at parameters where estimate (35) seems to no longer hold.

For $H = 0.33$, due to memory limitations, DH is unable to generate paths larger than $N = 2^{24}$, where ABSec is already about 40 times faster. Since ABSec is also more memory efficient (see next section), we can generate grids of size up to 2^{32} for which, if we interpolate the growth of DH,[2] we find that ABSec is 5500 times faster than DH for these parameters. For $H > \frac{1}{2}$, the advantage is less pronounced, and at a comparable discretization precision the algorithm is "only" 40–50 times faster at $H = 0.67$.

Performance also depends on the initial grid size. In Figs. 6 and 8, we compare run times for two different initial grids, Λ^4 and Λ^8. For larger initial grid sizes, the algorithm is slower since more points need to be generated initially. An increase in initial grid size leads to a decrease of 15% (for $H = 0.33$) in the average number of bisections. This is approximately outweighed by the time DH takes to generate a path on Λ^8 (see Fig. 6).

The run time increases only slowly for a smaller error tolerance. In Fig. 8, we show how user time decreases when changing ϵ' from $\varepsilon' = 10^{-9}$ to 10^{-7}. For an effective precision of $2^{\frac{32}{3}}$, user time increases by roughly 60%. Since error rates grow linearly with ε' (see Fig. 3), we conclude that for an error

[1] This experiment was run with an initial grid Λ^4 and $\varepsilon' = 10^{-9}$.

[2] Since DH scales with $N \ln(N)$, we fit with $f(N; a, b, c) = N[a \ln(N) + b] + c$.

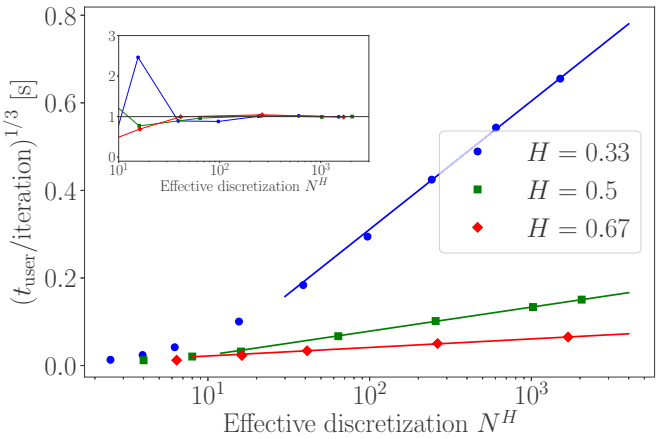

FIG. 8. User time for ABSec (solid lines) compared to DH (dashed line) for two different initial grid sizes and two different values of error tolerance ε'. For a 100 times higher error tolerance (top semitransparent pair of lines), user time increases by up to 60%.

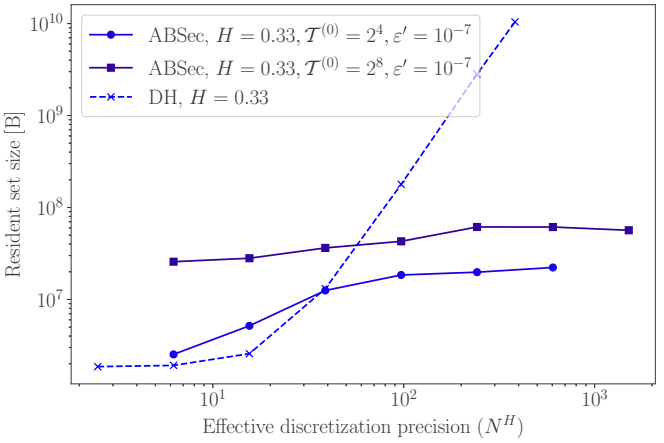

FIG. 9. Memory usage for DH (dashed line) and ABSec (solid line) for two different initial subgrid sizes. DH scales linearly in N, while ABSec grows only slowly (see text for estimate). For system of size $N_{\text{eff}} = 2^{28}$, ABSec needs only 10^{-2} to 10^{-3} of the memory for DH. For larger systems or smaller H, the advantage of ABSec is even bigger. Measurements were taken after 10^4 iterations.

rate 100 times lower one only needs to invest 60% more user time.

All together, these observations show that the algorithm behaves in a controlled manner for varying error tolerances and initial grid sizes. Depending on the number of iterations, and the quality of the data desired, choosing g (initial grid size), L (desired precision), and ε' (error tolerance level) accordingly leads to an algorithm that performs up to 5000 times faster than DH at $H = 0.33$, that was hitherto very hard to access with high precision. The algorithm should be tested more for $H = 0.25$, where it allows one to reach a precision unimaginable by DH.

F. Memory requirements

As a final benchmark of our algorithm, we consider memory usage. The latter is defined by the resident set size of the process, as measured by GETRUSAGE. When using DH, the full grid needs to be saved, and in doing so memory usage scales like N. Figure 9 shows memory usage for both DH and ABSec when performed for different effective discretization precisions and initial grid sizes. It shows that for large system sizes ABSec gains a growing and significant advantage. To generate a path of 2^{28} lattice points in double precision via DH, one requires 10 GB working memory, whereas ABSec uses between 20 and 80 MB, depending on the initial grid size. This represents an improvement by a factor of 125 to 500. This is due to the fact that only the initial grid which scales as $O(2^g)$, the additional grid points of order $O(M)$, and a correlation matrix scaling as $O(2^g + M)^2$ need to be stored. As implemented, additional memory is needed for the catalog of inverse correlation matrices [see Eq. (14)] which occupies memory of order $O(2^{3g})$, so including the catalog overall memory space grows like $2^{3g} + (2^g + M)^2$. For N_{eff} large enough, we assume that $2^g \ll M$, such that asymptotically for large effective system sizes the necessary memory grows as order M^2. For values of $H \gtrsim \frac{1}{3}$, we empirically found that $M \sim \ln(N_{\text{eff}})$, such that in that parameter range we estimate memory to grow as

$$\mathcal{M}^{\text{ABSec}}(N_{\text{eff}}) = O((\ln N_{\text{eff}})^2), \quad H \gtrsim \tfrac{1}{3}. \tag{36}$$

This advantage is again due to $M \ll 2^L$, i.e., using the fact that the first-passage time can be found to equal precision with many less grid points.

G. Floating-point precision

Currently, our implementation uses the 64-bit DOUBLE type. Since the variance of a bridge point is calculated from the subtraction of quantities of $O(1)$ [see Eq. (30)] the difference of which can be as small as $O(2^{-LH})$, the subtraction suffers from the finite floating-point precision when L is too large, as is demonstrated in Fig. 10 (see caption for details). This leads to $L_{\max} \simeq 10.5/H$, or $N \simeq 2 \times 10^{\frac{3}{H}}$.

H. Discussion

In this section we illuminated several aspects of our algorithm that show how it is capable of generating first-passage times with high numerical precision using several orders of magnitude less CPU time and memory as compared to DH. We chose to compare ABSec to DH because the latter is widely spread in simulating first-passage times of FBM (see, e.g., [6,7]), and since it is the fastest known exact generator of FBM. Since our method is also exact (the statistics of the grid generated is bias free), we think of DH as the natural benchmark. There are related approximative algorithms like the random midpoint displacement algorithm $R_{\ell,r}$ that also inserts midpoints, only taking into account the ℓ left and r right nearest neighbors [24]. This neglects long-range correlations between small increments at t_1, t_2 which even for $t_1 \ll t_2$ are correlated algebraically via $(t_1 - t_2)^{-1} + O[(H - \frac{1}{2})^2]$ (for $H \neq \frac{1}{2}$). The ABSec algorithm uses the full inverse correlation matrix of all points generated and is therefore closely related to exact procedures like DH.

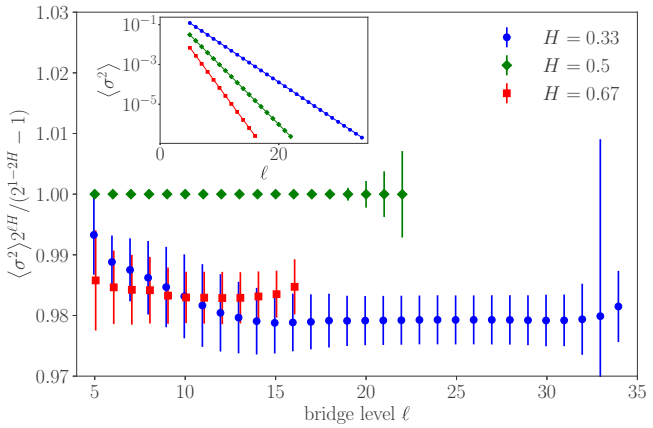

FIG. 10. Ratio between the sampled variance and no-neighbor-estimate of variance [see Eq. (17)] of an inserted midpoint X_m vs the level of the bisected bridge. For $H = 0.5$ (green diamond marks), the ratio equals 1, as BM is Markovian. For $H \neq 0.5$ (red square marks $H = 0.67$, blue circle lines $H = 0.33$), the variance fluctuates, as shown by the error bars for one standard deviation. Numerical errors due to a loss of floating-point precision become relevant around $L_{\max} \simeq 11/H$. ABSec was used with an initial grid Λ^8 and $\varepsilon' = 10^{-9}$.

Supported by our experiments, we are able to control both vertical and horizontal errors at the scale of inherent errors of a Monte Carlo simulation. In practice, the limiting factors are not systematical errors of the algorithm but floating-point imprecisions stemming from the matrix inversion.

The phone-book test used to asses the error rate does not take into account issues of precision when drawing new midpoints, which are copied from a pregenerated grid. Since this is an implementation-dependent grid, we decided to only use the phone-book test since the errors caused in that procedure are the ones inherent to the algorithm itself. An implementation with a higher-precision floating-point unit seems highly desirable.

IV. SUMMARY

When simulating first-passage times, or any other nonlocal observable, of fractional Brownian motion, the large fluctuations for $H < \frac{1}{2}$ require the grid to have a very high resolution for the same quality of data as for $H \geqslant \frac{1}{2}$. Generating a fine grid is particularly expensive, both in memory and in time. The algorithm proposed here refines the grid only where it is likely to impact the first-passage event. To give rigorous notion to that idea, we developed a precise criterion for when and where the grid should be refined. The new midpoints are then sampled exactly. Comparing it to the fastest known exact sampler, the Davies-Harte algorithm, we find that our implementation of the algorithm is 5000 times faster and uses 1000 times less memory when applied to $H = 0.33$ at $N_{\text{eff}} = 2^{32}$, due to the fact that only roughly 0.1% of the full grid is needed to determine the first-passage event. Our algorithm works with a probabilistic approximation, and the error rate can be bounded by 10^{-6} or even 10^{-8}. This should be sufficient for most Monte Carlo experiments and on the order of numerical (algorithm-independent) errors.

We have successfully used the algorithm to validate the analytic results for the first-passage time in [9]. There we used 2.5 CPU years at precision $N = 2^{28}$. With DH we would have had to reduce the precision to $N = 2^{24}$, which still would have taken 75 CPU years.

Finally, the concepts presented here can be used for other observables and other Gaussian processes. We hope that our algorithm contributes to confirming theoretical predictions on extreme events in Gaussian processes that were hitherto numerically inaccessible at the required precision.

ACKNOWLEDGMENTS

The authors thank Marc-Thierry Jaekel and Andy Thomas for computing support and resources. B.W. is grateful to Gunnar Pruessner for insightful discussions and support, and thanks Laboratoire de Physique Théorique de l'École Normale Supérieure and Laboratoire de Physique de l'École Normale Supérieure for hospitality. We thank Matteo D'Achille for a careful reading of the paper.

APPENDIX A: DERIVATION OF THE CRITICAL STRIP LENGTH

In this section we derive the width of the critical strip which was introduced in Sec. II C 5. The critical strip refers to the distance between a FBM bridge of size $\delta t = 2^{-\ell}$ and the threshold m, below which the midpoint of the bridge may surpass the threshold with probability larger than ε. We ignore any other grid points beyond the two fixed bridge points. By translational invariance, we set $X_0 = 0$, and $X_{\delta t} = a$ ($a \in \mathbb{R}$). The problem is then equivalently stated as

$$P\left[X^{\mathrm{B}}_{\delta t/2} > c(\varepsilon)\right] = \varepsilon, \tag{A1}$$

where X^{B}_t is the FBM-bridge process conditioned on $X_0, X_{\delta t}$. Following the derivation in [39], the law of the FBM bridge is itself a Gaussian process with first and second moment,

$$\left\langle X^{\mathrm{B}}_t \right\rangle = \frac{\langle X_t \delta(X_{\delta t} - a)\rangle}{\langle \delta(X_{\delta t} - a)\rangle}, \tag{A2}$$

$$\left\langle X^{\mathrm{B}}_s X^{\mathrm{B}}_t \right\rangle = \frac{\langle X_s X_t \delta(X_{\delta t} - a)\rangle}{\langle \delta(X_{\delta t} - a)\rangle}, \tag{A3}$$

where on the right-hand-side the averages are over free FBM paths. As shown in [39], Eqs. (8) and (9), the averages are

$$\left\langle X^{\mathrm{B}}_t \right\rangle = a\frac{C(t,\delta t)}{C(\delta t,\delta t)}, \tag{A4}$$

$$\left\langle X^{\mathrm{B}}_s X^{\mathrm{B}}_t \right\rangle^{\mathrm{c}} = C(s,t) - \frac{C(s,\delta t)C(t,\delta t)}{C(\delta t,\delta t)}, \tag{A5}$$

where $C(s,t)$ is the correlation function of Eq. (3). Since we are only interested in the midpoint with $s = t = \delta t/2$, this yields

$$\mu = \left\langle X^{\mathrm{B}}_{\delta t/2} \right\rangle = \frac{a}{2}, \tag{A6}$$

$$\sigma = \left\langle \left(X^{\mathrm{B}}_{\delta t/2}\right)^2 \right\rangle^{\mathrm{c}} = \left(2^{1-2H} - \tfrac{1}{2}\right)(\delta t)^{2H}. \tag{A7}$$

This determines the normal distribution of the midpoint and by translational invariance proves the values used in Sec. II C 5.

APPENDIX B: HOW TO GENERATE AN ADDITIONAL RANDOM MIDPOINT

We derive the conditional law of an additional randomly generated midpoint for an arbitrary Gaussian process as given in Eqs. (25) and (26). Let $\mathcal{T}^N = t_1, \ldots, t_N$ and $\mathcal{X}^N = X_{t_1}, \ldots, X_{t_N}$ be given, and denote the point to be inserted by t_{N+1} and $X_{t_{N+1}}$ (the times are not ordered). For ease of notation, we write $X_i = X_{t_i}$. As a Gaussian process, the vector $\vec{X} = (X_1, \ldots, X_N, X_{N+1})^T$ is a normal random variable with mean zero and covariance matrix

$$\langle \vec{X} \otimes \vec{X} \rangle = C(t_i, t_j) =: C(N+1), \quad 1 \leqslant i, j \leqslant N. \tag{B1}$$

It has a symmetric inverse correlation matrix $C^{-1}_{i,j}$. Its probability law is therefore given by

$$P(\vec{X}) = \frac{\exp\left(-\frac{1}{2}\sum_{i,j=1}^{N+1} X_i C^{-1}_{i,j} X_j\right)}{\sqrt{(2\pi)^{N+1}\det(C)}}. \tag{B2}$$

Since $X_1, \ldots, X_N$ are fixed, X_{N+1} conditioned on $\mathcal{X}^N$ follows the marginal distribution

$$P(X_{N+1}|\mathcal{X}^N) = \frac{\exp\left(-\frac{1}{2}X^2_{N+1}C^{-1}_{N+1,N+1} - \sum_{j=1}^{N} X_j C^{-1}_{N+1,j} X_{N+1}\right)}{\sqrt{2\pi / C^{-1}_{N+1,N+1}}}. \tag{B3}$$

Note that the normalizing factor in Eq. (B2) has canceled, since Eq. (B3) is a conditional average. This is a Gaussian distribution

$$P(X_{N+1}|\mathcal{X}^N) = \frac{\exp\left[-\frac{\sigma^2}{2}(X_{N+1}-\mu)^2\right]}{\sqrt{2\pi}\sigma}, \tag{B4}$$

with variance

$$\sigma^2 = \frac{1}{C^{-1}_{N+1,N+1}} \tag{B5}$$

and mean

$$\mu = -\sum_{j=1}^{N} X_j \frac{C^{-1}_{N+1,j}}{C^{-1}_{N+1,N+1}}. \tag{B6}$$

The mean can be seen as an average of the X_j with weight $C^{-1}_{N+1,j}/C^{-1}_{N+1,N+1}$.

APPENDIX C: DERIVATION OF THE ENLARGED CORRELATION MATRIX

In this section, we derive the algorithm to promote inverse correlation matrices as given in Eqs. (27)–(32). Assuming that $C(N)$ and $C^{-1}(N)$ are known, the aim is to find $C(N+1)$ and $C^{-1}(N+1)$ in as few as possible computational steps. The starting point is the observation that $C(N+1)$ contains $C(N)$ as the block matrix and is only augmented by a row and identical column:

$$C(N+1) = \left(\begin{array}{c|c} C(N) & \vec{\gamma} \\ \hline \vec{\gamma}^T & \langle X^2_{N+1}\rangle \end{array}\right), \tag{C1}$$

where $\vec{\gamma}$ is defined in Eq. (27) and $\langle X^2_{N+1}\rangle = 2t^{2H}_{N+1}$ in the case of FBM, but is intentionally left general. For the more difficult part, the inversion, we assume that the inverse correlation matrix is of the form

$$C^{-1}(N+1) = \left(\begin{array}{c|c} A(N) & \vec{b} \\ \hline \vec{b}^T & c \end{array}\right) \tag{C2}$$

for some arbitrary (symmetric) matrix A, vector $\vec{b}$, and number c. Multiplying matrices (C1) and (C2) results in

$$CC^{-1} = \left(\begin{array}{c|c} C(N)A(N) + \gamma \otimes \vec{b}^T & C(N)\vec{b} + c\vec{\gamma} \\ \hline (C(N)\vec{b} + c\vec{\gamma})^T & \vec{b}^T\vec{\gamma} + c\langle X^2_{N+1}\rangle \end{array}\right) \stackrel{!}{=} \mathbf{1}_{N+1}, \tag{C3}$$

such that one obtains the system of equations

$$C(N)\cdot A(N) + \vec{\gamma} \otimes \vec{b}^T = \mathbf{1}_N, \tag{C4}$$

$$C(N)\cdot\vec{b} + c\vec{\gamma} = \vec{0}, \tag{C5}$$

$$\vec{b}\cdot\gamma + c\langle X^2_{N+1}\rangle = 1. \tag{C6}$$

This is solved by

$$A(N) = C^{-1}(N) + \frac{C^{-1}(N)\cdot\vec{\gamma}\otimes\vec{\gamma}^T\cdot C^{-1}(N)}{\langle X^2_{N+1}\rangle - \vec{\gamma}\cdot C^{-1}(N)\cdot\vec{\gamma}}, \tag{C7}$$

$$\vec{b} = -\frac{C^{-1}(N)\cdot\vec{\gamma}}{\langle X^2_{N+1}\rangle - \vec{\gamma}\cdot C^{-1}(N)\cdot\vec{\gamma}}, \tag{C8}$$

$$c = \frac{1}{\langle X^2_{N+1}\rangle - \vec{\gamma}\cdot C^{-1}(N)\cdot\vec{\gamma}}. \tag{C9}$$

Defining $\vec{g}$ as in Eq. (28) and σ^2 as in Eq. (30), one arrives at the inverse matrix (32).

[1] S. Redner, *A Guide to First-Passage Processes* (Cambridge University, Cambridge, England, 2001).

[2] R. Metzler, G. Oshanin, and S. Redner, *First-Passage Phenomena and Their Applications* (World Scientific, Singapore, 2014).

[3] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: Fifty years after Kramers, Rev. Mod. Phys. **62**, 251 (1990).

[4] A. Godec and R. Metzler, First passage time distribution in heterogeneity controlled kinetics: Going beyond the mean first passage time, Sci. Rep. **6**, 20349 (2016).

[5] L. P. Sanders and T. Ambjörnsson, First passage times for a tracer particle in single file diffusion and fractional Brownian motion, J. Chem. Phys. **136**, 175103 (2012).

[6] T. Guérin, N. Levernier, O. Bénichou, and R. Voituriez, Mean first-passage times of non-Markovian random walkers in confinement, Nature (London) **534**, 356 (2016).

[7] N. Levernier, O. Bénichou, T. Guérin, and R. Voituriez, Universal first-passage statistics in aging media, Phys. Rev. E **98**, 022125 (2018).

[8] K. J. Wiese, First passage in an interval for fractional Brownian motion, Phys. Rev. E **99**, 032106 (2019).

[9] M. Arutkin, B. Walter, and K. J. Wiese, Extreme Events for Fractional Brownian motion with drift: Theory and numerical validation, arXiv:1908.10801.

[10] B. B. Mandelbrot and J. W. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. **10**, 422 (1968).

[11] A. H. O. Wada and T. Vojta, Fractional Brownian motion with a reflecting wall, Phys. Rev. E **97**, 020102(R) (2018).

[12] T. Guggenberger, G. Pagnini, T. Vojta, and R. Metzler, Fractional Brownian motion in a finite interval: correlations effect depletion or accretion zones of particles near boundaries, New J. Phys. **21**, 022002 (2019).

[13] J. Pickands III, Asymptotic properties of the maximum in a stationary Gaussian process, Trans. Am. Math. Soc. **145**, 75 (1969).

[14] G. M. Molchan, Maximum of a fractional Brownian motion: Probabilities of small values, Commun. Math. Phys. **205**, 97 (1999).

[15] V. I. Piterbarg, *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, Translations of Mathematical Monographs Vol. 148 (American Mathematical Society, Providence, 1995).

[16] V. I. Piterbarg, *Twenty Lectures About Gaussian Processes* (Atlantic Financial, London, 2015).

[17] K. J. Wiese, S. N. Majumdar, and A. Rosso, Perturbation theory for fractional Brownian motion in presence of absorbing boundaries, Phys. Rev. E **83**, 061141 (2011).

[18] M. Delorme and K. J. Wiese, Maximum of A Fractional Brownian Motion: Analytic Results from Perturbation Theory, Phys. Rev. Lett. **115**, 210601 (2015).

[19] T. Sadhu, M. Delorme, and K. J. Wiese, Generalized Arcsine Laws for Fractional Brownian Motion, Phys. Rev. Lett. **120**, 040603 (2018).

[20] R. B. Davies and D. S. Harte, Tests for Hurst effect, Biometrika **74**, 95 (1987).

[21] D. M. Ceperley, Path integrals in the theory of condensed helium, Rev. Mod. Phys. **67**, 279 (1995).

[22] M. Sprik, M. L. Klein, and D. Chandler, Staging: A sampling technique for the Monte Carlo evaluation of path integrals, Phys. Rev. B **31**, 4234 (1985).

[23] A. Fournier, D. Fussell, and L. Carpenter, Computer rendering of stochastic models, Commun. ACM **25**, 371 (1982).

[24] I. Norros, P. Mannersalo, and J. L. Wang, Simulation of fractional Brownian motion with conditionalized random midpoint displacement, Adv. Perfor. Anal. **2**, 77 (1999).

[25] A. B. Dieker, Simulation of fractional Brownian motion, Ph.D. thesis, University of Twente, 2004.

[26] P. F. Craigmile, Simulating a class of stationary gaussian processes using the Davies-Harte algorithm, with application to long memory processes, J. Time Ser. Anal. **24**, 505 (2003).

[27] D. Krapf, N. Lukat, E. Marinari, R. Metzler, G. Oshanin, C. Selhuber-Unkel, A. Squarcini, L. Stadler, M. Weiss, and X. Xu, Spectral Content of A Single Non-Brownian Trajectory, Phys. Rev. X **9**, 011019 (2019).

[28] B. Walter and K. J. Wiese, Monte Carlo sampler of first-passage times for fractional Brownian motion using adaptive bisections: Source code, https://hal.archives-ouvertes.fr/hal-02270046.

[29] V. I. Piterbarg, *Twenty Lectures about Gaussian Processes* (Atlantic Financial, London, 2015).

[30] G. H. Weiss and R. J. Rubin, The theory of ordered spans of unrestricted random walks, J. Stat. Phys. **14**, 333 (1976).

[31] G. H. Weiss, E. A. DiMarzio, and R. J. Gaylord, First passage time densities for random walk spans, J. Stat. Phys. **42**, 567 (1986).

[32] V. Palleschi and M. R. Torquati, Mean first-passage time for random-walk span: Comparison between theory and numerical experiment, Phys. Rev. A **40**, 4685 (1989).

[33] K. J. Wiese, Span observables: "When is a foraging rabbit no longer hungry?", J. Stat. Phys. **178**, 625 (2019).

[34] E. Anderson Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, 1999).

[35] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi, *GNU Scientific Library Reference Manual* (Network Theory, London 2009), 3rd ed.

[36] M. Frigo, A fast Fourier transform compiler, in *Proceedings of the ACM SIGPLAN 1999 Conference on Programming Language Design and Implementation, PLDI '99* (ACM, New York, 1999), pp. 169–180.

[37] L. S. Blackford, A. Petitet, R. Pozo, K. Remington, R. C. Whaley, J. Demmel, J. Dongarra, I. Duff, S. Hammarling, G. Henry *et al.*, An updated set of basic linear algebra subprograms (BLAS), ACM Trans. Math. Softw. **28**, 135 (2002).

[38] B. Walter and K. J. Wiese, https://github.com/benjamin-w/davies-harte-fpt.git.

[39] M. Delorme and K. J. Wiese, Extreme-value statistics of fractional Brownian motion bridges, Phys. Rev. E **94**, 052105 (2016).