Title	Busy periods of fractional Brownian storage: a large deviations approach
Author(s)	Norros, Ilkka
Citation	Advances in Performance Analysis vol. 2(1999):1, pp. 1-19
Date	1999
Rights	This article may be downloaded for personal use only

Busy periods of fractional Brownian storage: a large deviations approach

Ilkka Norros
VTT Information Technology
P.O.Box 1202, 02044 VTT, Finland
ilkka.norros@vtt.fi

Abstract

A storage with constant service rate and fractional Brownian input is considered. It is shown how large deviation asymptotics of the buffer occupancy and the ongoing busy period can be obtained applying a generalized Schilder's theorem. The crucial point of the method is to identify the "most probable path" satisfying some criterion. Whereas this turns out to be very simple in the case of achieving a given buffer occupancy, the case of achieving a given length of the ongoing busy period remains an open problem. Instead, tight bounds are obtained in this latter case.

1 Introduction

A teletraffic model based on fractional Brownian motion ("fractional Brownian traffic") and a corresponding storage system ("fractional Brownian storage") were introduced in [5]. The stationary storage level distribution of this system has been studied in several papers ([2, 7, 3, 4]) giving increasingly accurate estimates of its tail behaviour.

The present paper might be the first one where such an analysis is attempted on the distribution of a busy period of fractional Brownian storage. Our technical framework here is the theory of large deviations in path space of a Gaussian process. The inspiration for doing this came from reading the recent book by A. Shwartz and A. Weiss [10]. It deals, however, only with Markovian systems, and the general theory must in our case be taken from the world of general Gaussian processes.

The problem of describing long busy periods of the fractional Brownian storage was raised in [6] in the following context. There is lot of evidence that a self-similar correlation structure is a reasonable choice for a simplified model of data traffic. If the traffic to a particular link additionally comes from a large number of sources, traffic fluctuations can be assumed Gaussian, and the aggregated traffic can be modelled by fractional Brownian traffic. However, when a heavily loaded link is considered, the Gaussian variation is distorted, since sources using end-to-end control protocols like TCP or ABR can adjust their rates so that the aggregated rate roughly equals the link capacity over long periods. Huge queues, predicted by FBM-based theory, cannot be observed — the buffer sizes of real systems are limited. Instead, we can see very long busy periods on heavily loaded Internet links, international ones in particular.

It was suggested in [6] that, in the situation described above, one could still apply fractional Brownian traffic modelling. Indeed, one can think that the corresponding "free traffic", i.e., what the sources would transmit if the link speed would not be a restriction, would sum up to self-similar Gaussian traffic, and that there is indeed a *virtual* huge queue, which is distributed among the sources and thus not directly observable. The busy periods of the heavily loaded link, on the other hand, remain unchanged, in the ideal case at least. Thus, the fractional Brownian storage could be used to explain and to predict the behaviour of a heavily loaded link also when the buffer in front of it has only moderate size and the sources obey some rate control which keeps the losses low.

The formulation of the mathematical problem is as follows. Let $(Z_t)_{t\in\mathbb{R}}$ be a standard fractional Brownian motion (FBM) with self-similarity parameter $H \in (0,1)$, that is, a centered Gaussian process with stationary increments and variance $\mathrm{E}Z_t^2 = |t|^{2H}$. We always choose a version with continuous paths. (For the theory of FBM, see, e.g., Section 7.2 of [9].) Let us define the *normalised fractional Brownian storage* as the process

$$V_t = \sup_{s\leq t}(Z_t - Z_s - (t-s)).$$

V is a non-negative stationary process. Its path can be thought of as consisting of excursions from 0 to the domain of positive values and back. As in the thoroughly studied Brownian case (the special case $H = 1/2$), the random set of t such that $V_t = 0$ is an uncountable set of Lebesgue measure zero, and one cannot speak about the distribution of a "typical" excursion like one can speak about a typical busy period in an ordinary queue. (Although each path consists of a countable number of excursions, there are "too many" small excursions, and there is never (a.s.) such a thing as "the next" excursion.) Note also that the excursions are not independent when $H \neq 1/2$. It seems that little is known about them in the non-Brownian case.

Although there is no probability distribution for a "typical excursion" picked randomly from the set of all excursions, the excursion going on at a particular time t is a well defined random object. By stationarity, its distribution is the same for all t, and we can focus at the excursion or, in the storage terminology, busy period containing time 0.

In addition to more standard facts about FBM, we make use of Theorem 1.1 below which gives an explicite orthogonalization of FBM. This result has turned out to be very useful in many contexts. For a detailed presentation and references to earlier similar results, see [8]. Let $H \neq 1/2$, and define a process M as

$$M_t \doteq c_1 \int_0^t s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H} \, \mathrm{d}Z_s, \tag{1.1}$$

where

$$c_1 = \left[2H\Gamma(\frac{3}{2}-H)\Gamma(H+\frac{1}{2})\right]^{-1},$$

and $\Gamma(\cdot)$ denotes the gamma function.

Theorem 1.1 *The centered Gaussian process M has independent increments and is thus a martingale. Its variance function is*

$$\mathrm{E}M_t^2 = c_2^2 t^{2-2H},$$

where

$$c_2 = \left(H(2H-1)(2-2H)B(H-\frac{1}{2}, 2-2H)\right)^{-\frac{1}{2}},$$

and $B(\mu,\nu)$ denotes the beta function

$$B(\mu,\nu) = \int_0^1 x^{\mu-1}(1-x)^{\nu-1} \, \mathrm{d}x = \frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)}.$$

Moreover, $\mathrm{E}Z_s M_t = s$ *for* $0 \leq s \leq t$.

This result is easiest to understand in a Hilbert space framework. Let G denote the Gaussian space of Z, that is, the L^2 closure of linear combinations of random variables Z_t. The integral (1.1) can be defined as an L^2 limit. Doing this, one in fact constructs a new Hilbert space L_Γ by extending $Z_t \mapsto 1_{[0,t]}$ to an isometry from G to L_Γ. The Gaussian random variable M_t is then the counterpart of the function $w(t,\cdot) \in L_\Gamma$ defined by

$$w(t,s) = c_1 s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H} 1_{(0,t)}(s). \tag{1.2}$$

A central role is played below by a third Hilbert space, the reproducing kernel Hilbert space of Z. Its isometry with G and the last statement of Theorem 1.1

turn out to be surprisingly useful when considering busy periods of fractional Gaussian storage (Proposition 3.1 below).

The paper is organized as follows. The general large deviation results applied here are resumed in Section 2. Some basic "most probable paths" are identified in Section 3. The main result is derived in Section 4. Since we have only considered the normalised storage process, the result is also formulated for a system with arbitrary traffic and capacity parameters in Section 5. In Section 6, we indicate how the large deviations asymptotics of the storage occupancy distribution (that is, the stationary distribution of V) can be obtained with the path space approach, with the "typical path" to high occupancy as a by-product. We compare the large deviation estimate with simulation results in Section 7. Some final remarks are made in Section 8.

2 Large deviation principle for fractional Brownian motion

Let us first specify the framework of the generalized Schilder's theorem of [1] in our case. We shall use this framework throughout the paper. Denote by Ω the function space

$$\Omega = \left\{ \omega : \ \omega \text{ continuous } \mathbb{R} \to \mathbb{R}, \ \omega(0) = 0, \ \lim_{t\to\infty} \frac{\omega(t)}{1+|t|} = \lim_{t\to-\infty} \frac{\omega(t)}{1+|t|} = 0 \right\}.$$

Equipped with the norm

$$\|\omega\|_\Omega = \sup \left\{ \frac{|\omega(t)|}{1+|t|} : \ t \in \mathbb{R} \right\},$$

Ω is a separable Banach space.

We choose Ω also as our basic probability space by letting P be the unique probability measure on the Borel sets of Ω such that the random variables $Z_t(\omega) = \omega(t)$ form a normalised fractional Brownian motion with some fixed self-similarity parameter H. The covariance function of Z is denoted by

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

The *reproducing kernel Hilbert space* R related to Z is a space of functions $\mathbb{R} \to \mathbb{R}$ which is defined by letting the mapping

$$Z_t \mapsto \Gamma(t,\cdot)$$

span an isometry from the Gaussian space of Z, i.e. from the smallest closed linear subspace of $L^2(\Omega, \mathcal{B}_\Omega, P)$ containing all the Z_t's, onto R. The relation

$$\langle \Gamma(s,\cdot), \Gamma(t,\cdot)\rangle_R = \Gamma(s,t)$$

is generalized to the useful *reproducing kernel property*:

$$\langle f, \Gamma(t,\cdot)\rangle_R = f(t), \quad f \in R. \tag{2.1}$$

To see the basic relationships between R and Ω, let f be any linear combination of any finite choice of functions $\Gamma(s_i,\cdot)$. Then (2.1) holds for f and, by Cauchy-Schwarz,

$$\|f\|_\Omega = \sup_{t\in\mathbb{R}} \frac{|\langle f, \Gamma(t,\cdot)\rangle_R|}{1+|t|} \le \sup_{t\in\mathbb{R}} \frac{\|f\|_R \|\Gamma(t,\cdot)\|_R}{1+|t|} = \|f\|_R \sup_{t\in\mathbb{R}} \frac{|t|^H}{1+|t|}.$$

We see that all elements of R are continuous functions, R is a subset of Ω, and the topology of R is finer than that of Ω.

Next we define two families of linear transformations on R. First, set

$$\tau_s \Gamma(t,\cdot) \doteq \Gamma(t+s,\cdot) - \Gamma(s,\cdot)$$

for any $s \in \mathbb{R}$ and extend τ_s by linearity and continuity to all $f \in R$. Second, for any $\alpha > 0$, let

$$(\sigma_\alpha f)(t) \doteq \alpha^{-H} f(\alpha t).$$

Lemma 2.1 *All tranformations τ_s and σ_α are automorphisms of R. Moreover, they have the group properties $\tau_s\tau_t = \tau_{s+t}$ and $\sigma_\alpha\sigma_\beta = \sigma_{\alpha\beta}$.*

Proof The automorphism property of τ_s follows from the fact that Z has stationary increments. The automorhism property of σ_α follows from the self-similarity of Z. The group properties are easy to check. □

Let us now turn to the large deviations. It is straightforward to check that $(\Omega, R, \mathrm{Id}, P)$, where Id denotes the natural embedding of R into Ω, is a Wiener quadruple as defined on p. 88 of [1]. Then, theorem 3.4.12 of [1] gives the following generalized Schilder's theorem for fractional Brownian motion:

Theorem 2.2 *The function $I : \Omega \to [0,\infty]$,*

$$I(\omega) = \begin{cases} \frac{1}{2}\|\omega\|_R^2, & \text{if } \omega \in R, \\ \infty, & \text{otherwise,} \end{cases} \tag{2.2}$$

is a good rate function for the centered Gaussian measure P and satisfies the large deviations principle:

$$\text{for } F \text{ closed in } \Omega: \qquad \limsup_{n\to\infty} \frac{1}{n}\log P\left(\frac{Z}{\sqrt{n}} \in F\right) \le -\inf_{\omega\in F} I(\omega);$$

$$\text{for } G \text{ open in } \Omega: \qquad \liminf_{n\to\infty} \frac{1}{n}\log P\left(\frac{Z}{\sqrt{n}} \in G\right) \ge -\inf_{\omega\in G} I(\omega).$$

For any set $A \subseteq \Omega$, denote $I(A) = \inf_{\omega\in A} I(\omega)$. We call a function $f \in A$ such that $I(f) = I(A) < \infty$ a *most probable path* of A. The following explanatory passage shows that a most probable path can be intuitively (remember that there is no counterpart to Lebesgue measure on Ω!) understood as a point of maximum likelihood.

Motivation from the finite-dimensional case

Since the above framework is somewhat difficult to understand at the first sight, a reader unfamiliar with it can obtain some insight from the finite-dimensional case. Consider a finite-dimensional centered Gaussian vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with non-singular covariance matrix Γ. The density function of $\boldsymbol{X}$ is

$$g(\boldsymbol{x}) = const \cdot e^{-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\Gamma^{-1}\boldsymbol{x}}. \tag{2.3}$$

Now, the expression $\boldsymbol{x}^{\mathrm{T}}\Gamma^{-1}\boldsymbol{x}$ is in fact the square of a norm in $\mathbb{R}^d$. Indeed, write $\Gamma = [\Gamma_1, \ldots, \Gamma_d]$. Since Γ in non-singular, the vectors Γ_i are linearly independent and thus form a (non-orthogonal) basis of $\mathbb{R}^d$. Define an inner product $\langle\cdot,\cdot\rangle_R$ in $\mathbb{R}^d$ by setting

$$\langle\Gamma_i, \Gamma_j\rangle_R = \Gamma_{ij} = \mathrm{E}X_iX_j$$

and extending it by linearity to all $\boldsymbol{x} = \sum_1^d a_i\Gamma_i$. Now,

$$\|\boldsymbol{x}\|_R^2 = \left\langle \sum_{i=1}^d a_i\Gamma_i, \sum_{i=1}^d a_i\Gamma_i \right\rangle_R = \boldsymbol{a}^{\mathrm{T}}\Gamma\boldsymbol{a} = \boldsymbol{x}^{\mathrm{T}}\Gamma^{-1}\boldsymbol{x}, \tag{2.4}$$

so that the density function can be written as

$$g(\boldsymbol{x}) = const \cdot e^{-\frac{1}{2}\|\boldsymbol{x}\|_R^2}. \tag{2.5}$$

The "kernel" Γ has the reproducing property

$$\langle\boldsymbol{x}, \Gamma_j\rangle_R = \sum_{i=1}^d a_i\Gamma_{ij} = x_j. \tag{2.6}$$

This implies that the conditional expectation of $\boldsymbol{X}$, when some of its components are known, is closely related to the R-norm. Indeed, consider the conditional distribution of $\boldsymbol{X}$ on the condition $\{X_i = y\}$. Since the conditional distribution is Gaussian, its expectation is at the point of highest probability density, and by (2.5), the task is to minimize the norm $\|\boldsymbol{x}\|_R$ in the set $\{x_i = y\}$. But by the reproducing kernel property (2.6), the condition can be written with the inner product as

$$\langle \boldsymbol{x}, \Gamma_i \rangle_R = y.$$

Thus, the task is to minimize the norm when the inner product with a fixed vector is given. The solution is of course found in the subspace $\mathrm{Sp}\{\Gamma_i\}$, i.e.

$$\boldsymbol{x} = \mathrm{E}\left[\boldsymbol{X} \mid X_i = y\right] = \frac{y}{\|\Gamma_i\|_R^2}\Gamma_i.$$

More generally, the conditional expectation of $\boldsymbol{X}$ given that $X_{i_1} = x_{i_1}, \ldots, X_{i_k} = x_{i_k}$ is the linear combination of $\Gamma_{i_1}, \ldots, \Gamma_{i_k}$ which satisfies $\langle \boldsymbol{x}, \Gamma_{i_j} \rangle_R = x_{i_j}$ for $j = 1, \ldots, k$. Finding the conditional expectation requires only solving a linear equation of order k.

3 Most probable paths of fractional Brownian motion

In this section, we consider the simplest optimization problems in the space R. These facts turn out to be very useful in the study of busy periods.

The first question is about the "most probable path" to reach a value $x > 0$ at a time $t > 0$, i.e., we want to minimize $\|f\|_R$ in the set

$$D(t, x) = \{f \in R : \ f(t) = x\}.$$

Now, by the reproducing kernel property (2.1), the condition $f(t) = x$ can be written in the form $\langle f, \Gamma(t, \cdot) \rangle_R = x$. Obviously, the solution f^* with smallest R-norm is

$$f^* = \frac{x}{\Gamma(t, t)}\Gamma(t, \cdot).$$

One can hardly visually distinguish f^* from a straight line from the origin to (t, x). For the busy period problem it is, however, important to note that $f^*(t/2) = x/2$ and f^* has an S-form for $H > 1/2$ and a mirrored S-form for $H < 1/2$. Figure 3.1 shows the difference $f^*(t) - t$ in the case $t = x = 1$, $H = 0.75$.

The next step is to find the "most typical path" with fixed values a and b at two points $s < t$, respectively. As in the previous case, the solution is a linear

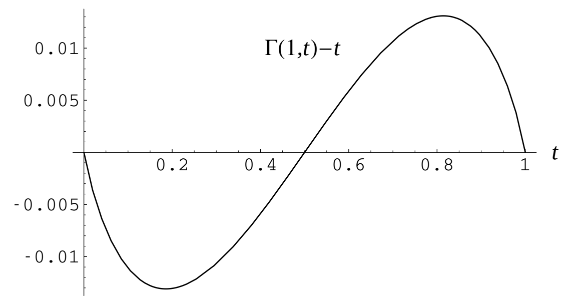

Figure 3.1: The difference $\Gamma(1,t)-t$ in the case $H=0.75$.

combination of the basic functions: $f^* = u\Gamma(s,\cdot)+v\Gamma(t,\cdot)$. Denoting

$$\Gamma_{[s,t]} = \begin{bmatrix} \Gamma(s,s) & \Gamma(s,t) \\ \Gamma(s,t) & \Gamma(t,t) \end{bmatrix},$$

the condition for u and v reads

$$[u\ v]\Gamma_{[s,t]} = [a\ b],$$

so that

$$f = [a\ b]\Gamma^{-1}_{[s,t]} \begin{bmatrix} \Gamma(s,\cdot) \\ \Gamma(t,\cdot) \end{bmatrix}. \tag{3.1}$$

Further, we have

$$\|f\|^2_R = \langle f,f\rangle_R = [a\ b]\Gamma^{-1}_{[s,t]}\Gamma_{[s,t]}\Gamma^{-1}_{[s,t]} \begin{bmatrix} a \\ b \end{bmatrix} = [a\ b]\Gamma^{-1}_{[s,t]} \begin{bmatrix} a \\ b \end{bmatrix}. \tag{3.2}$$

In the “excursion” case $b=0$, $t=1$, we have the particularly simple formula

$$\|f^*\|^2_R = \frac{a^2}{\Gamma(s,s)-\Gamma(1,s)^2}.$$

Since Z has stationary increments, we have

$$\Gamma(t,t)-\Gamma(t,1)^2 = \frac{1}{2}\left(v(t)+v(1-t)-\frac{1}{2}-\frac{1}{2}(v(t)-v(1-t))^2\right),$$

where $v(t)=\Gamma(t,t)$. This is symmetric with respect to the point $t=\frac{1}{2}$. Figure 3.2 shows the “typical path” for $H=0.5$, 0.9 and 0.2. We see three clearly different temperaments!

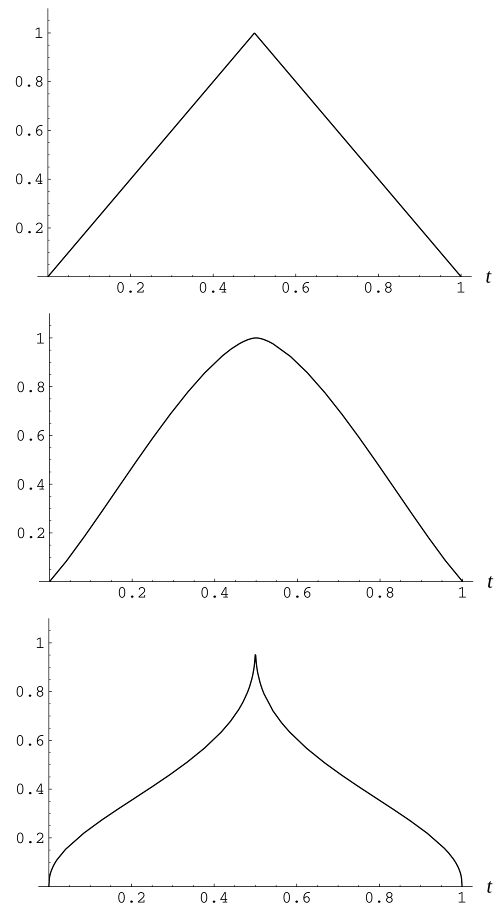

Figure 3.2: Typical forms of excursions of fractional Brownian motions to the value 1 and back, with H values 0.5, 0.9 and 0.2 (from top).

Later we shall also need the solution to the following optimization problem: find the path $f \in R$ with smallest R-norm satisfying $f(t) = t$ for $t \in [0, 1]$. Since its solution is less trivial than the previous examples, we formulate it as a proposition. The result is a by-product of Theorem 1.1 which was briefly discussed in Section 1.

Proposition 3.1 *The element of*

$$J \doteq \{f \in R : \ f(t) = t \ \forall t \in [0, 1]\}$$

with the smallest R-norm, say χ, is the counterpart of the random variable M_1, defined in (1.1), in the isometry $Z_t \mapsto \Gamma(t, \cdot)$ from the Gaussian space of Z onto R. In particular, $\|\chi\|_R = c_2$, where c_2 is the constant defined in Theorem 1.1.

Proof First note that $\chi \in J$ since, by the reproducing kernel property and Theorem 1.1,

$$\chi(t) = \langle \chi, \Gamma(t, \cdot) \rangle = \mathrm{E} M_1 Z_t = t \quad \forall t \in [0, 1].$$

Second, χ belongs to the closed linear subspace $R_{[0,1]}$ generated by the functions $\Gamma(t, \cdot)$, $t \in [0, 1]$, since the corresponding fact holds for M_1 by (1.1). For any function $g \in J$, the reproducing kernel property implies that the orthogonal projection of g on $R_{[0,1]}$ is χ and, consequently, $\|g\|_R \geq \|\chi\|_R$. □

Remark 3.2 The function χ has the explicite expression

$$\chi(s) = w(1, \frac{1}{2})\Gamma(1, s) - \int_0^{\frac{1}{2}} \Gamma(s, t) w'(1, t)\, \mathrm{d}t - \int_{\frac{1}{2}}^1 (\Gamma(s, t) - \Gamma(s, 1)) w'(1, t)\, \mathrm{d}t,$$

where $w(1, t)$ was defined in (1.2). This can be found by applying integration by parts to (1.1) and changing to the space R.

Remark 3.3 The following table summarizes some isometry counterparts in the Gaussian space G and in the function spaces L_Γ and R. (The spaces G and L_Γ and the function $w(1, \cdot)$ were defined in Section 1.)

L_Γ	G	R
$1_{[0,t]}$	Z_t	$\Gamma(t, \cdot)$
$w(1, \cdot)$	M_1	χ

4 Path approach to asymptotics of storage occupancy distribution

For any path $\omega \in \Omega$, we define the corresponding storage level path

$$v_t(\omega) = \sup_{s \le t}(\omega(t) - \omega(s) - (t-s)).$$

Since $|\omega|$ grows slower than linearly at infinities, $v_t(\omega)$ is finite for every $\omega \in \Omega$. By *busy periods* of $v_t(\omega)$ we mean its positive excursions, that is, intervals $[a,b]$ such that $v_t(\omega) > 0$ for $t \in (a,b)$ and $v_a(\omega) = v_b(\omega) = 0$. The busy period containing 0 is defined as

$$[a(\omega), b(\omega)] \doteq [\sup\{t \le 0 : \ v_t(\omega) = 0\}, \ \inf\{t \ge 0 : \ v_t(\omega) = 0\}],$$

if $a(\omega) < 0 < b(\omega)$, otherwise the system is considered non-busy at time 0 (which happens in our case only with probability 0). Note that we have

$$\omega(t) - \omega(a(\omega)) > t - a(\omega) \quad \text{for } t \in (a(\omega), b(\omega)). \tag{4.1}$$

Denote by

$$K_T = \{\omega \in \Omega : \ a(\omega) < 0 < b(\omega), \ b(\omega) - a(\omega) > T\} \tag{4.2}$$

the set of paths for which the busy period containing 0 is strictly longer than T. It is easy to see that K_T is an open subset of Ω. The following proposition, a direct consequence of the self-similarity of Z, indicates how Schilder's theorem can give estimates of $P(Z \in K_T)$ for large T.

Proposition 4.1 *For $T > 0$ and the set $K_T \subset \Omega$ as defined in (4.2), we have*

$$P(Z \in K_T) = P\left(\frac{Z}{T^{1-H}} \in K_1\right).$$

Proof

$$\begin{aligned} P(Z \in K_T) &= P\left(\exists a < 0, \, b > (a+T)^+ : \ Z_t - Z_a > t - a \ \forall t \in (a,b)\right) \\ &= P\left(\exists a < 0, \, b > (a+1)^+ : \ Z_{Tt} - Z_{Ta} > Tt - Ta \ \forall t \in (a,b)\right) \\ &= P\left(\exists a < 0, \, b > (a+1)^+ : \ T^{H-1}(Z_t - Z_a) > t - a \ \forall t \in (a,b)\right) \\ &= P\left(\frac{Z}{T^{1-H}} \in K_1\right). \end{aligned}$$

□

Thus, we obtain a large deviations lower bound by minimizing $I(f) = \frac{1}{2}\|f\|_R^2$ for $f \in K_1 \cap R$. Denote for $s < t$

$$Q(s,t) = \{\omega \in \Omega : \ \omega(t) - \omega(s) > t - s\}, \quad Q^*(s,t) = \bigcap_{u \in (s,t)} Q(s,u).$$

Then

$$K_1 = \bigcup \{Q^*(s,t) : \ s < 0 < t, \ t - s > 1\}.$$

To get the large deviations upper bound result, we also have to consider the closed set $\overline{K_1}$.

Proposition 4.2

$$\overline{K_1} = \bigcup \left\{\overline{Q^*(s,t)} : \ s \le 0 \le t, \ t - s \ge 1\right\}, \tag{4.3}$$

and

$$\overline{Q^*(s,t)} = \{\omega : \ \omega(u) - \omega(s) \ge u - s \ \forall u \in (s,t)\}. \tag{4.4}$$

Proof We first prove (4.4). The inclusion "$\subseteq$" is obvious. To show "$\supseteq$", let ω be an arbitrary element of the set at the right hand side. Then $\omega_n(u) = \omega(u) + ((u \wedge t) - s)^+/n$ defines a sequence in $Q^*(s,t)$ such that $\omega_n \to \omega$.

Let us then prove the inclusion "$\supseteq$" of (4.3). Taking into account (4.4) and its proof, it remains to show that any path $\omega \in \overline{Q^*(0,t)}$ can be approximated by a sequence ω_n such that $\omega_n \in \overline{Q^*(-1/n,t)}$ and $\|\omega_n - \omega\|_\Omega \to 0$. Such a sequence can be defined simply by

$$\omega_n(u) = \begin{cases} \omega(u) - \omega(-\frac{1}{n}) - \frac{1}{n}, & u \le -\frac{1}{n}, \\ u, & u \in (-\frac{1}{n}, 0], \\ \omega(u), & u > 0. \end{cases}$$

Finally, we show "$\subseteq$" of (4.3). Let $\omega \in \overline{K_1}$ be arbitrary. There exists a sequence $\omega_n \in K_1$ such that $\|\omega_n - \omega\| \to 0$. For each n, choose s_n and t_n such that $\omega_n \in Q^*(s_n, t_n)$ and $t_n - s_n > 1$. First note that the sequences s_n and t_n must be bounded. Indeed, using the condition $\omega_n(t_n) - \omega_n(s_n) \ge t_n - s_n$, and the definition of $\|\cdot\|_\Omega$, we can deduce that

$$\frac{\omega(t_n)}{1+t_n} \ge \frac{t_n}{1+t_n} + \frac{1+|s_n|}{1+t_n}\left(\frac{\omega(s_n)}{1+|s_n|} + \frac{|s_n|}{1+|s_n|} - \|\omega - \omega_n\|_\Omega\right) - \|\omega - \omega_n\|_\Omega.$$

Since $\omega(u)/(1+|u|) \to 0$ when $u \to \infty$, the right hand side would obtain larger values than the left hand side if either of the sequences were unbounded. Thus, we can pick a subsequence such that there exist finite limits $s_{n_k} \to s_\infty$ and $t_{n_k} \to t_\infty$. Obviously $t_\infty - s_\infty \geq 1$.

Now, the fact that $\omega \in \overline{Q^*(s_\infty, t_\infty)}$ follows by taking limits in the inequality

$$\omega_{n_k}(u) - \omega_{n_k}(s_{n_k}) \geq u - s_{n_k},$$

which holds for every $u \in (s_\infty, t_\infty)$ for k sufficiently large. □

It is no big surprise that the closure does not matter in $I(\overline{Q^*(0,1)})$, as shown in the next proposition.

Proposition 4.3 $I(\overline{Q^*(0,1)}) = I(Q^*(0,1))$.

Proof It is sufficient to show that $I(\overline{Q^*(0,1)} \cap R) \geq I(Q^*(0,1) \cap R)$. Take any $f \in \overline{Q^*(0,1)} \cap R$ and let $\chi \in R$ be as in Proposition 3.1. Then $f + \epsilon\chi \in Q^*(0,1)$ for every $\epsilon > 0$, and

$$\|f + \epsilon\chi\|_R^2 \to \|f\|_R^2 \quad \text{as } \epsilon \to 0,$$

which proves the claim. □

Next we show that it is sufficient to consider paths on the interval $[0,1]$. (Note that this means that the huge difference between an on-going and an arbitrary busy period, discussed in the Introduction, vanishes in the large deviations limit.)

Proposition 4.4

$$I(K_1) = \frac{1}{2} \inf \left\{ \|f\|_R^2 : \; f \in Q^*(0,1) \cap R \right\}$$

and

$$I(\overline{K_1}) = \frac{1}{2} \inf \left\{ \|f\|_R^2 : \; f \in \overline{Q^*(0,1)} \cap R \right\}.$$

Proof We have to show that

$$I(K_1) = \inf_{s<0<t, \; t-s>1} \; \inf_{f \in Q^*(s,t)} I(f) = I(Q^*(0,1)).$$

By the reproducing kernel property (2.1) and Lemma 2.1,

$$\begin{aligned} I(Q^*(s,t)) &= \frac{1}{2} \inf \{ \|f\|_R^2 : \; f \in R, \; \langle f, \Gamma(u,\cdot)\rangle_R - \langle f, \Gamma(s,\cdot)\rangle_R > u - s, \\ &\qquad \forall u \in (s,t) \} \\ &= \frac{1}{2} \inf \left\{ \|f\|_R^2 : \; f \in R, \; \langle f, \tau_s \Gamma(u,\cdot)\rangle_R > u, \; \forall u \in (0, t-s) \right\} \\ &= \frac{1}{2} \inf \left\{ \|\tau_{-s} f\|_R^2 : \; f \in R, \; \langle \tau_{-s} f, \Gamma(u,\cdot)\rangle_R > u, \; \forall u \in (0, t-s) \right\} \\ &= I(Q^*(0, t-s)). \end{aligned}$$

Thus, $I(K_1) = \inf_{u>1} I(Q^*(0,u))$. Since $u \mapsto I(Q^*(0,u))$ is increasing, it suffices to note that for any $f \in Q^*(0,1)$ and $n > 0$ there is a number $u_n > 1$ such that $f + \frac{1}{n}\chi \in Q^*(0,u_n)$ (cf. the proof of Proposition 4.3).

Thanks to Proposition 4.2, the proof for $\overline{K_1}$ is similar, except for replacing certain strict inequalities by non-strict ones. □

We can now state our main result.

Theorem 4.5

$$\lim_{T\to\infty} \frac{1}{T^{2-2H}} \log P\left(Z \in K_T\right) = -I(Q^*(0,1)).$$

The number $I(Q^*(0,1))$ is $1/2$ for $H = 1/2$ and lies in the interval $(\frac{1}{2}, \frac{c_2^2}{2})$ for $H \neq 1/2$, where $c_2 = c_2(H)$ is the number given in Theorem 1.1.

Proof We have shown everything except the bounds for $I(Q^*(0,1))$. First, $I(Q^*(0,1)) \le c_2^2/2$ because $(1+\epsilon)\chi \in Q^*(0,1)$ for every ϵ. On the other hand,

$$I(Q^*(0,1)) \ge I(Q(0,1)) = I(\Gamma(1,\cdot)) = \frac{1}{2}. \tag{4.5}$$

For $H = 1/2$, we have $c_2(1/2) = 1$. It remains to assume $H \neq 1/2$ and show that the interval in the assertion is open.

Consider first the left endpoint. Here the essential observation is that the set $Q^*(0,1)$ is convex (by the way, note that K_1 is not convex!) and the level sets of I are strictly convex and compact. Since $\Gamma(1,\cdot) \notin \overline{Q^*(0,1)}$ (by (4.4) and the fact illustrated in Figure 3.1), the lower bound in (4.5) must be strict.

Finally, we show that the bound $c_2^2/2$ can be improved by finding a function $\psi \in R$ such that $\psi(t) \ge 0$ for $t \in [0,1]$ and $\|\chi + \psi\|_R < \|\chi\|_R$. Let $t \in (0,1)$ and denote by $\psi(t;\cdot)$ the function with minimal R-norm satisfying $\psi(t;t) = 1$, $\psi(t;1) = 0$, identified in Section 3. As shown there, it has the expression

$$\psi(t;\cdot) = \frac{1}{\Gamma(t,t) - \Gamma(1,t)^2}\left(\Gamma(t,\cdot) - \Gamma(1,t)\Gamma(1,\cdot)\right).$$

Now,

$$\langle \chi, \psi(t;\cdot)\rangle_R = \frac{t - \Gamma(1,t)}{\Gamma(t,t) - \Gamma(1,t)^2},$$

which is negative when $t > 1/2$ for $H > 1/2$ and negative when $t < 1/2$ for $H < 1/2$. Fix any $t \in (1/2,1)$ if $H > 1/2$ and any $t \in (0,1/2)$ if $H < 1/2$. Since the minimum of

$$\|\chi + a\psi(t;\cdot)\|_R^2 = \|\chi\|_R^2 + 2a\langle \chi, \psi(t;\cdot)\rangle_R + a^2\|\psi(t;\cdot)\|_R^2, \tag{4.6}$$

with respect to a is obtained with

$$a = -\frac{\langle \chi, \psi(t;\cdot)\rangle_R}{\|\psi(t;\cdot)\|_R^2} = \Gamma(1,t) - t > 0, \tag{4.7}$$

$\|\chi + a\psi(t;\cdot)\|_R^2 < \|\chi\|_R^2$. □

Substituting (4.7) into (4.6), we can further minimize with respect to t and find that the minimal norm is obtained with

$$t = \text{argmin}_t \left\{ -\frac{(t-\Gamma(1,t))^2}{\Gamma(t,t)-\Gamma(1,t)^2} \right\}. \tag{4.8}$$

Further, this approach can be extended to any finite number of points t_i and multipliers a_i. Thus, one can obtain arbitrarily good approximations to $I(Q^*(0,1))$ and to the most probable busy period path by solving minimization problems numerically.

For $H > 1/2$, c_2 is very close to 1 (see Figure 4.1) so in practice one can use the approximation $I(Q^*(0,1)) \approx 1/2$. For $H < 1/2$, this estimate is less good.

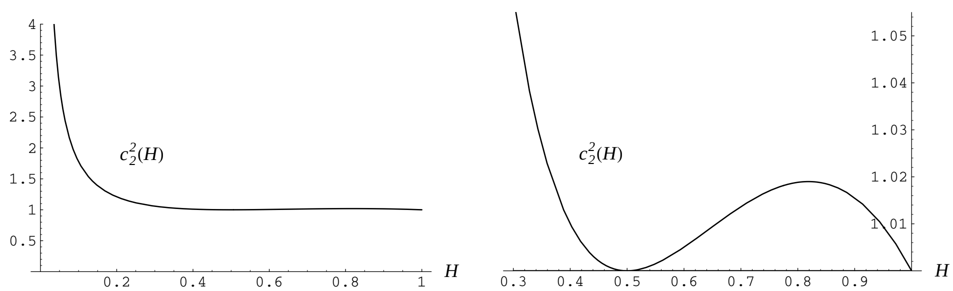

Figure 4.1: The constant c_2^2 as a function of H. The right plot shows the values for $H \in (0.3, 1)$ in higher resolution.

The exact value of $I(Q^*(0,1))$ remains unknown in the present paper. We have shown, however, that there exists a non-trivial most probable busy period path, whose identification is an interesting problem for further study.

5 The effect of traffic parameters on the busy period distribution

It is now straightforward to obtain the large deviations asymptotics of busy periods of a fractional Brownian storage with parameters m, a and c as defined in [5]. That is, the input in an interval $(s,t]$ is $m(t-s)+\sqrt{ma}(Z_t-Z_s)$ and the leak rate is c. The following proposition is easily proved by a scaling argument (cf. Theorem 3.1 of [5]).

Proposition 5.1 *Denote by $\widetilde{V}_t$ the storage occupancy process of a fractional Brownian storage with parameters m, a, H and c. Then*

$$\left(\widetilde{V}_t\right)_{t\in\mathbb{R}} \stackrel{\mathcal{D}}{=} \left(\frac{c-m}{\alpha}V_{\alpha t}\right)_{t\in\mathbb{R}}, \quad \text{where} \quad \alpha = \left(\frac{c-m}{\sqrt{ma}}\right)^{1/(1-H)}.$$

In particular,

$$\lim_{t\to\infty}\frac{1}{T^{2-2H}}\log P\left([\text{busy period of } \widetilde{V} \text{ containing } 0]\geq T\right)$$
$$= -\frac{(c-m)^2}{ma}I(Q^*(0,1)).$$

6 The typical path to high buffer occupancy

In a similar way, we can derive the large deviation asymptotics of the stationary storage occupancy distribution. As a by-product, we obtain the "typical path" to reach a high buffer level x. We restrict to a heuristic presentation, since it is straightforward to write the analog of the detailed reasoning of the previous section in this considerably simpler case.

So, reaching level x in the buffer at time 0 means that for some $T>0$ we have $(0-Z_{-T})-(0-(-T))\geq x$. Using stationarity and self-similarity, this is transformed to an equiprobable event on time interval $[0,1]$: $Z_1\geq T^{1-H}+xT^{-H}$. The most probable path to realize this is $(T^{1-H}+xT^{-H})\Gamma(1,\cdot)$, and its R-norm is $(T^{1-H}+xT^{-H})$. Minimizing this with respect to T gives

$$T^* = \frac{H}{1-H}x.$$

Substituting this, re-scaling and translating the big queue back to the origin, yields that the most probable way to reach x is that Z_t follows the path $T^{*H}f(1+t/T^*)$ defined by

$$f(s) = \frac{x^{1-H}}{\kappa(H)}\Gamma(1,s), \quad s\in\mathbb{R},$$

where $\kappa(H) = H^H(1-H)^{1-H}$, and the probability of doing this is, in the exponential large deviation asymptotics, roughly

$$P\left(V_0>x\right)\approx \exp(-\frac{1}{2}\|f\|_R^2) = \exp\left(-\frac{x^{2-2H}}{2\kappa(H)^2}\right).$$

This is (as it should!) the same Weibull distribution which was obtained as asymptotics of a lower bound in [5] and with full one-dimensional large deviations approach in [2].

7 Comparison with simulation results

We have made some comparisons of the large deviations estimates with results obtained by simulation. For $H = 0.7$, 0.9 and 0.2, ten simulation runs of length 2^{20} time points were made in each case. More precisely, discrete time queues with normalised fractional Gaussian noise X_n as input sequence and unit service capacity were simulated, starting with empty queue and determining queue length V_{n+1} at time $n+1$ by $V_{n+1} = (V_n + X_{n+1})^+$. From these realizations, the conditional probabilities

$$\mathrm{P}[\text{we are in busy period with length} \geq n \mid \text{we are in busy period}]$$

were estimated. The results are shown in Figure 7.1. For $H = 0.7$ and $H = 0.2$, the large deviations estimate is surprisingly good indeed. Note, however, that the discretization makes part of the busy periods shrink or disappear (there are long idle periods also, unlike the continuous time system), so the true distribution in the continuous time case may deviate from the large deviations estimate more than the simulations.

As regards $H = 0.9$, it is also somewhat surprising to note that in this case a simulation with one million time points is clearly too little for estimating the distribution of busy period length.

8 Concluding remarks

The results of this paper show that the fractional Brownian storage can quite well be analysed with large deviation techniques in path space. Simple and useful approximate results are obtained relatively easily, and simulations indicate that they are usable not only for extremely rare events but for "moderate rarity" as well. Determining the most probable busy period path remains a nice open problem. However, we indicated a method for approximating the asymptotics numerically.

Finally, it is worth of noting that this approach can also be applied to the case of a general Gaussian input process. Without self-similarity, it is then not sufficient to find the typical busy period of length 1, but some other technical ideas of this paper, like the important role of the counterpart of the function χ, can be expected to be useful in the more general setting as well.

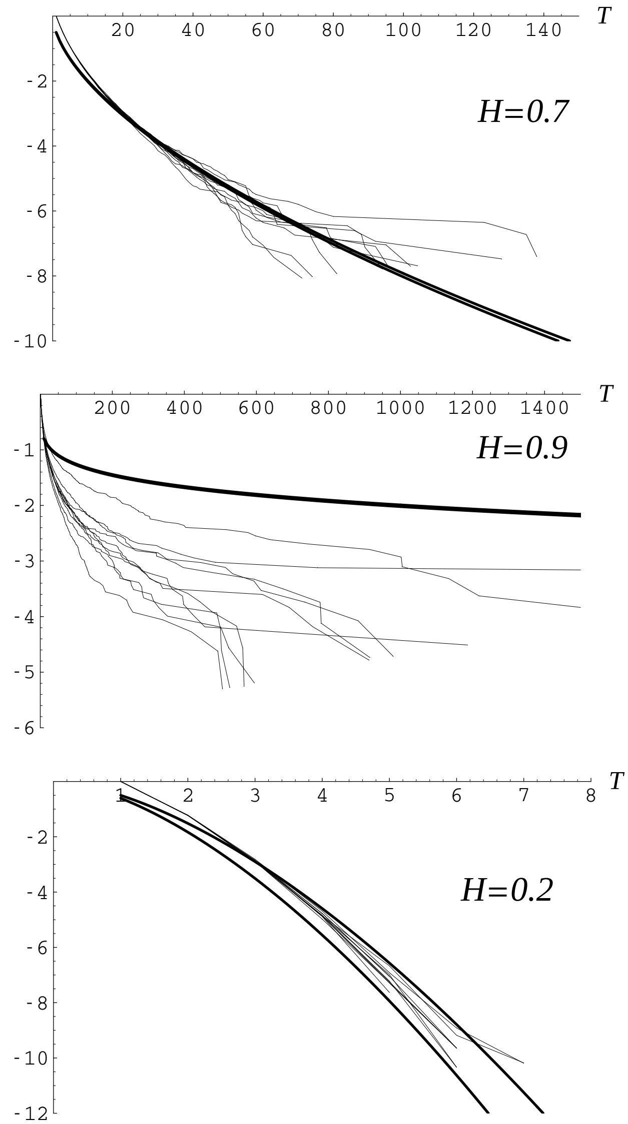

Figure 7.1: Plots of $\log P$ (busy period containing 0 longer than T) for a discrete time version of fractional Brownian storage. For $H = 0.7$ (top), $H = 0.9$ (middle) and $H = 0.2$ (bottom), estimates from 10 simulated realizations of fractional Gaussian noise (sequence length 2^{20}) and the large deviations estimates (thick lines) are presented. Both approximations $I(Q^*(0,1)) \approx c_2^2/2$ and $I(Q^*(0,1)) \approx 1/2$ are shown.

Acknowledgement I thank the referees for several improvements of the details of the paper.

References

[1] J.-D. Deuschel and D.W. Stroock. *Large Deviations.* Academic Press, Boston, 1989.

[2] N. G. Duffield and N. O'Connell. Large deviations and overflow probabilities for the general single-server queue, with applications. *Mathematical Proceedings of the Cambridge Philosophical Society*, 1996.

[3] L. Massoulie and A. Simonian. Large buffer asymptotics for the queue with FBM input. To appear.

[4] O. Narayan. Exact asymptotic queue length distribution for fractional Brownian traffic. *Advances in Performance Analysis*, 1:39–63, 1998.

[5] I. Norros. A storage model with self-similar input. *Queueing Systems*, 16:387–396, 1994.

[6] I. Norros. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE Journal on Selected Areas in Communications*, 13(6), August 1995.

[7] I. Norros. Four approaches to the fractional Brownian storage. In J. Lévy Véhel, E. Lutton, and C. Tricot, editors, *Fractals in Engineering.* Springer, 1997.

[8] I. Norros, E. Valkeila, and J. Virtamo. An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. To appear in Bernoulli.

[9] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman & Hall, 1994.

[10] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis.* Chapman & Hall, 1995.