The Calculus of Hurstiness*

Damon Wischik[†] and Ayalvadi Ganesh[‡]

28 September 2004

Abstract

Traffic flows with long-range dependence satisfy a certain type of scaling relationship, and it is this scaling relationship which governs the scaling behaviour of a queueing network fed by such flows. In this paper we present a mathematical technique for reasoning about scaling behaviour in queueing systems, based on large deviations theory. Our technique makes it straightforward to calculate the impact of long-range dependent traffic on tandem queues, priority queues, and many other systems.

1 Introduction

The term 'long-range dependence' is loosely used to cover a range of phenomena relating to scaling behaviour. These phenomena have attracted a great deal of attention in the teletraffic community in the past ten years. We feel there is still space for more attention, and in this paper we present a new approach, based on a novel conception of long-range dependence which we call 'Hurstiness'. Our approach makes it possible to reason about scaling behaviour in networks (for example, the scaling behaviour of a departure process from a processor-sharing queue) using a bare minimum of calculation. The underlying mathematical technology is large deviations theory; our approach was inspired by the remarks of O'Connell [8, Section 2.8], formalized in [6, Theorem 6.16], about the generality of exponential tails.

One standard way to define long-range dependence is this. Let $X(t)$ be the amount of work that arrives at a queue in a time-interval of duration t. Say that the arrival process X is long-range dependent if

$$\operatorname{Var} X(t) \sim \sigma^2 t^{2H} \quad \text{as } t \to \infty \tag{1}$$

*Keywords: long range dependence; Hurst parameter; fractional Brownian motion; queues; queueing networks; heavy tails; large deviations

†Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK. Email D.J.Wischik@statslab.cam.ac.uk. Tel +44 1223 339094.

‡7 J.J.Thomson Avenue, Cambridge CB3 0FB, UK.

for some *Hurst parameter* $H \in (\frac{1}{2}, 1)$ and some $\sigma^2 > 0$. We will not use this definition. For our purposes, it is more useful to suppose instead that

$$\frac{1}{t^{2(1-H)}} \log \mathbb{P}\big(t^{-1}X(t) \in B\big) \to -\inf_{x \in B} I(x) \quad \text{as } t \to \infty \tag{2}$$

for suitable sets B, for some $H \in (0, 1)$ and for some function I. When this is true, we will say that X has *Hurstiness* H. (There are actually some further subtleties given in the formal definition in Section 2 below.) The term 'Hurstiness' was coined by John Lewis to refer loosely to the range of phenomena associated with long-range dependence; in this paper we have co-opted his term for our precise technical use.

To see that the two definitions are related, suppose that $X(t)$ is distributed like a normal random variable $\mathrm{N}(0, \sigma^2 t^{2H})$. Then clearly X has Hurst parameter H. Also, taking for example $B = [x, \infty)$,

$$\mathbb{P}\big(t^{-1}X(t) \in B\big) = \mathbb{P}\big(\mathrm{N}(0, \sigma^2) \geq t^{1-H}x\big) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[-\frac{x^2 t^{2(1-H)}}{2\sigma^2}\Big].$$

It can further be shown that (2) is satisfied, with $I(x) = x^2/2\sigma^2$. Thus X has Hurstiness H.

The benefit of our Hurstiness formulation is this: the driving term in (2) is $t^{2(1-H)}$, and it is straightforward to show that this driving term also applies to derived quantities like the departure process from a queue or the queue size in a downstream queue. By contrast, it is not so straightforward to determine whether (1) holds for derived quantities.

The organization of the paper is as follows. In Section 2 we give an abstract definition of Hurstiness. In Section 3 we detail our notation for traffic processes and queueing systems. In Sections 4 and 5 we explain how to use Hurstiness to analyse large-buffer asymptotics for queueing networks fed by long-range dependent processes such as fractional Brownian motion; we cover aggregates, priority queues, processor sharing queues, and departure processes. In Section 6 we explain how Hurstiness helps us understand many-flows asymptotics, including the causes of long-range dependence. In Section 7 we conclude with a summary of the 'calculus of Hurstiness'.

2 Abstract definition of Hurstiness

This paper is based on large deviations theory. For a comprehensive reference on all the large deviations techniques used here, see [2]. The only thing we will reproduce here is the definition of a large deviations principle:

Let X^N be a sequence of random variables, indexed by $N \in \mathbb{N}$ or $N \in \mathbb{R}^+$, taking values in some regular Hausdorff space $\mathcal{X}$. Suppose that all open and closed sets are measureable. Say that X^N satisfies a *large deviations principle* at speed N^α (for some real number $\alpha > 0$) and with some rate function I if for all measurable sets B

$$
\begin{aligned}
-\inf_{x\in B^\circ} I(x) &\le \liminf_{N\to\infty} \frac{1}{N^\alpha}\log \mathbb{P}(X^N\in B)\\
&\le \limsup_{N\to\infty} \frac{1}{N^\alpha}\log \mathbb{P}(X^N\in B) \le -\inf_{x\in \bar{B}} I(x).
\end{aligned}
$$

We say I is a *rate function* if $I(x)\ge 0$ for all x, and if all level sets $\{x : I(x)\le\alpha\}$ for $\alpha\in\mathbb{R}$ are closed. We say I is a *good rate function* if in addition all level sets are compact. It is a standard result that if I is a good rate function and C is a closed set and $\inf_{x\in C} I(x)<\infty$ then the infimum is attained at some $x\in C$.

(Here B° denotes the interior of B, and $\bar{B}$ the closure. The space is regular if for every closed set F and point $x\notin F$ there exist disjoint open neighbourhoods of F and x; this will be needed for reasoning about uniqueness of the rate function. All the spaces we work with in this paper are regular Hausdorff spaces.)

Definition 1 (Hurstiness) *Say that the sequence X^N has* Hurstiness $H\in(0,1)$ *if it satisfies a large deviations principle with speed $N^{2(1-H)}$ and some good rate function I, and furthermore*
i. there is some $\hat{x}\in\mathcal{X}$ for which $0<I(\hat{x})<\infty$;
ii. there is some $\mu\in\mathcal{X}$ such that $I(x)=0$ if and only if $x=\mu$.
A rate function which satisfies these two properties we call non-trivial. *Call μ the* zero-value *of the rate function.*

Note that the 'if' part of (ii) is a consequence of the LDP: by the upper bound $\inf_{x\in\mathcal{X}} I(x)=0$, and by goodness $I(x)=0$ for some $x\in\mathcal{X}$.

This is a useful definition, because Hurstiness is preserved by continuous functions, as indicated by the following lemma, a simple consequence of the contraction principle.

Lemma 1 (Contraction) *Let X^N have Hurstiness H, and let the corresponding LDP hold in the space $\mathcal{X}$ and have rate function I and zero-value μ. Let $f:\mathcal{X}\to\mathcal{Y}$ be a continuous function from $\mathcal{X}$ to another regular Hausdorff space $\mathcal{Y}$.*
i. $f(X^N)$ satisfies an LDP in $\mathcal{Y}$ at speed $N^{2(1-H)}$ with good rate function

$$J(y)=\inf_{x:f(x)=y} I(x),$$

and $J(y)=0$ if and only if $y=f(\mu)$.
ii. If there is some $\hat{x}$ such that $I(\hat{x})<\infty$ and $f(\hat{x})\ne f(\mu)$ then $f(X^N)$ has Hurstiness H.

The condition in (ii) is to guard against trivial functions like $f(x)=0$. The proof of this lemma follows the statement of two more general lemmas: the first is a key result for comparing sequences with different Hurstinesses, and the second tells us that Hurstiness is well-defined.

Lemma 2 (Trivial LDP at higher speed) *Suppose that X^N has Hurstiness H with zero-value μ and good rate function I. Let $G > H$, $G \in (0,1)$. Then X^N satisfies an LDP at speed $N^{2(1-G)}$ with good rate function*

$$I'(x) = \begin{cases} 0 & \textit{if } x = \mu \\ \infty & \textit{otherwise.} \end{cases}$$

Lemma 3 (Uniqueness) *Suppose that X^N has Hurstiness H with zero-value μ, and it also has Hurstiness G with zero-value ν. Then $H = G$ and $\mu = \nu$.*

Proof of Lemma 1. Proof of (i). The LDP for $f(X^N)$ and the goodness of J follow immediately from the contraction principle.

If $y = f(\mu)$ then clearly $J(y) = 0$. Suppose conversely that $J(y) = 0$. Since I is good, and the set $f^{-1}(\{y\})$ is closed, there exists some x such that $f(x) = y$ and $I(x) = J(y) = 0$. Since I is non-trivial, $x = \mu$, and so $y = f(\mu)$. This fulfils (ii) of the definition of non-triviality.

Proof of (ii). Let $\hat{y} = f(\hat{x})$. If $J(\hat{y}) = 0$ then by the above $f(\hat{x}) = f(\mu)$, which we have assumed is not the case, hence $J(\hat{y}) \neq 0$. Since J is a rate function it is non-negative, and so $J(\hat{y}) > 0$. From the definition of J, $J(\hat{y}) \leq I(\hat{x})$; by our assumption that $I(\hat{x}) < \infty$ we see $J(\hat{y}) < \infty$. This fulfils (i) of the definition of non-triviality.

We have shown that J satisfies the two conditions of non-triviality. Thus $f(X^N)$ has Hurstiness H. □

Proof of Lemma 2. I' is a good rate function, since it is non-negative and its only level set is $\{\mu\}$ which is certainly compact.

LD upper bound. Let $B \subset \mathcal{X}$ be some closed set. We need to prove that

$$\limsup_{N\to\infty} \frac{1}{N^{2(1-G)}} \log \mathbb{P}(X^N \in B) \leq -\inf_{x\in B} I'(x). \tag{3}$$

Suppose first that $\inf_{x\in B} I(x) = 0$. Since I is good, $I(x) = 0$ for some $x \in B$. Since I is non-trivial, $x = \mu$, and so $\mu \in B$. Thus $\inf_{x\in B} I'(x) = 0$. So (3) is trivially satisfied.

Now consider the case where $\inf_{x\in B} I(x) > 0$. From the LDP at speed $N^{2(1-H)}$,

$$\limsup_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}(X^N \in B) \leq -\inf_{x\in B} I(x) < 0.$$

Thus there exists an $\varepsilon > 0$ and an N_0 such that for $N > N_0$

$$\frac{1}{N^{2(1-H)}} \log \mathbb{P}(X^N \in B) < -\varepsilon,$$

which implies that

$$\frac{1}{N^{2(1-G)}} \log \mathbb{P}(X^N \in B) = N^{2(G-H)} \frac{1}{N^{2(1-H)}} \log \mathbb{P}(X^N \in B) < -\varepsilon N^{2(G-H)}.$$

Therefore

$$\limsup_{N\to\infty} \frac{1}{N^{2(1-G)}} \log \mathbb{P}(X^N \in B) = -\infty$$

and so (3) is trivially satisfied. This completes the proof of the LD upper bound.

LD lower bound. Let $B \subset \mathcal{X}$ be some open set. We need to prove that

$$\liminf_{N\to\infty} \frac{1}{N^{2(1-G)}} \log \mathbb{P}(X^N \in B) \geq -\inf_{x\in B} I'(x). \tag{4}$$

Suppose first that $\inf_{x\in B} I'(x) = \infty$. Then (4) is trivially satisfied.

Now consider the case where $\inf_{x\in B} I'(x) < \infty$. This means that $\mu \in B$, hence $\inf_{x\in B} I'(x) = 0$ and $\inf_{x\in B^c} I'(x) = \infty$. Since B is open, B^c is closed. So we can use the LD upper bound established above to find that

$$\limsup_{N\to\infty} \frac{1}{N^{2(1-G)}} \log \mathbb{P}(X^N \in B^c) \leq -\inf_{x\in B^c} I'(x) = -\infty.$$

In particular, $\mathbb{P}(X^N \in B^c) \to 0$. Therefore $\mathbb{P}(X^N \in B) = 1 - \mathbb{P}(X^N \in B^c) \to 1$ and so

$$\liminf_{N\to\infty} \frac{1}{N^{2(1-G)}} \log \mathbb{P}(X^N \in B) = 0.$$

But this is exactly equal to $-\inf_{x\in B} I'(x)$, and so the LD lower bound is proved. □

Proof of Lemma 3. Let I be the rate function for the LDP at speed $N^{2(1-H)}$ and the zero-value be μ, and let J be the rate function for the LDP at speed $N^{2(1-G)}$ and the zero-value be ν.

Suppose that $H = G$. Since $\mathcal{X}$ is a regular Hausdorff space, the rate function is unique, so $I = J$. From the definition of non-triviality, $I(\nu) = J(\nu) = 0$ so $\nu = \mu$.

So assume (without loss of generality) that $G > H$. By Lemma 2, X^N satisfies an LDP at speed $N^{2(1-G)}$ with good rate function I'. By the uniqueness of the rate function at speed $N^{2(1-G)}$ it must be that $I' = J$. Yet I' is trivial and J is not, a contradiction. This contradicts the assumption that X^N has Hurstiness H and also Hurstiness $G > H$. □

3 Notation for traffic processes

In Section 2 we gave an abstract definition of Hurstiness, for arbitrary random variables. In Section 4 we will explore the interpretation of Hurstiness for traffic processes. In this section, we first establish some notation for talking about traffic.

Let $\mathcal{C}$ be the set of continuous functions

$$\mathcal{C} = \left\{ x : \mathbb{R}^+ \to \mathbb{R},\ x \text{ is continuous},\ x(0) = 0 \right\}$$

and equip it with the norm

$$\|x\| = \sup_{t\geq 0}\left|\frac{x(t)}{1+t}\right|$$

Define also the set of truncated functions

$$\mathcal{C}^T = \big\{x : [0,T] \to \mathbb{R},\ x \text{ is continuous},\ x(0)=0\big\}$$

Since $(\mathcal{C}, \|\cdot\|)$ is a metric space, it is a regular Hausdorff space. The same is true of $(\mathcal{C}^T, \|\cdot\|)$.

When we wish to discuss traffic processes we will interpret $x(t)$ as the amount of work arriving in the interval $(-t, 0]$, and use a more suggestive syntax: we will write

$x(-t,0]$	for	$x(t)$
$x(-u,-v]$	for	$x(u)-x(v)$
$x\|_{(-T,0]}$	for	x restricted to $[0,T]$, i.e. the projection of x into $\mathcal{C}^T$
$\dot{x}_t$	for	the rate $dx(t)/dt$, where this derivative exists
e	for	the constant-rate process $e(-t,0]=t$.

Define the *mean rate* of $x \in \mathcal{C}$ to be

$$\text{mean rate}\,(x) = \lim_{t\to\infty}\frac{x(t)}{t} \quad \text{(when the limit exists)}$$

and let

$$\mathcal{C}_\mu = \big\{x \in \mathcal{C} : x \text{ has mean rate } \mu\big\}.$$

4 Hurstiness of traffic processes

Now we will relate the abstract definition of Hurstiness in Section 2 to the idea of long-range dependence in traffic processes. Long-range dependence concerns the scaling properties of traffic over long timescales, and so we make the following definitions:

Given a sample path $x \in \mathcal{C}$, define the *speeded-up* version $x^{\circlearrowleft N}$ for $N \in \mathbb{R}^+$ by

$$x^{\circlearrowleft N}(t) = x(Nt).$$

Let X be a random traffic process, i.e. a $\mathcal{C}$-valued random variable. In the following sections, we will be interested in the sequence

$$\frac{1}{N}X^{\circlearrowleft N}.$$

Say that the traffic process X has Hurstiness H if this sequence has Hurstiness H.

The prototypical example of a long-range dependent traffic process is fractional Brownian motion (fBm). In this section we will show that fBm with

Hurst parameter H has Hurstiness H (hence the name Hurstiness). We will also demonstrate that a process can have high Hurstiness yet be arbitrarily smooth; for good intuition about LRD it is important to understand this point that 'Hurstiness is not burstiness'. In Section 5 we will go on to calculate the Hurstiness of derived quantities, such as queue size, in a network fed by Hursty traffic.

4.1 Fractional Brownian motion

Let Z be a standard fractional Brownian motion with Hurst parameter H, so that $Z(t)$ is distributed as $\mathrm{N}(0, t^{2H})$. There is a standard traffic model obtained from fBm:

$$A(-t, 0] = \mu t + \sigma Z(t).$$

Now, it is well-known that fBm is self-similar, i.e. that

$$\frac{1}{a^H} Z^{\circlearrowleft a} \text{ is distributed as } Z \quad \text{for all } a > 0.$$

From this fact, and using a standard result concerning large deviations for Gaussian processes (Theorem 10 in the appendix), we can show the following about A:

Lemma 4 (Hurstiness of fBm) *The process A, defined above, has Hurstiness H. The LDP for $N^{-1} A^{\circlearrowleft N}$ holds in the space $\mathcal{C}_\mu$, and the zero-value of the LDP is μe. The rate function I is such that*

$$I\big(\mu e + \xi(a - \mu e)\big) = \xi^2 I(a) \tag{5}$$

for any $\xi \in \mathbb{R}$ and $a \in \mathcal{C}_\mu$.

We have deliberately not given the rate function here, since the whole point of this paper is that one can prove a great deal without knowing the value of the rate function—one need only check that some quantity of interest is non-trivial in the sense of Lemma 1(ii). The above remark about I can sometimes be useful to this end. To be concrete, we know that I is non-trivial, so there exists some $\hat{a} \neq \mu e$ with $I(\hat{a}) < \infty$; thus for some t there are sample paths $\hat{b}$ with $\hat{b}(-t, 0]$ arbitrarily large and $I(\hat{b})$ finite; and as we will see in Section 5.1 this can be useful for finding the Hurstiness of queue size.

Proof. Theorem 10 in the appendix tells us that the sequence $Z/\sqrt{L}$ satisfies an LDP at speed L with some good rate function I in the space $\mathcal{C}_0$. Now let $L = N^{2(1-H)}$. From the self-similarity relationship,

$$\frac{1}{N} Z^{\circlearrowleft N} \text{ is distributed as } \frac{1}{\sqrt{L}} Z.$$

Rewriting the LDP for $Z/\sqrt{L}$ in terms of N, we obtain an LDP for $Z^{\circlearrowleft N}/N$ at speed $N^{2(1-H)}$, with the same good rate function I.

Lemma 11 in the appendix shows that the rate function is non-trivial, and that the zero-value is $0e$. That lemma, combined with the contraction principle, also gives (5).

By applying Lemma 1 to the continuous function $z \mapsto \mu e + \sigma z$, the sequence $N^{-1} A^{\circ N}$ satisfies an LDP in $\mathcal{C}_\mu$ with zero-value μe, and has Hurstiness H. $\square$

4.2 Hurstiness and burstiness

It is worth giving an example to illustrate the point that Hurstiness is not directly related to conventional ideas of 'burstiness'. In essence, Hurstiness is only concerned with long-timescale fluctuations, whereas the intuitive idea of burstiness also takes account of short-timescale fluctuations. We will present a process which has arbitrary Hurstiness $H \in (0, 1)$, but which is arbitrarily smooth (in the sense of being Lipschitz continuous with arbitrarily small constant).

The technical details rely on the departure map, which is not introduced properly until Section 5.3, but the idea is simple: take a fBm with mean rate 0 and feed it into a queue with constant service rate $\varepsilon > 0$. Then the departure process has the same Hurstiness as the input process, but its rate is bounded.

We now make this construction formal. Pick $\varepsilon > 0$. Let $z \in \mathcal{C}_0$. Consider the departure map $d_\varepsilon : \mathcal{C}_0 \to \mathcal{C}_0$, discussed further in Section 5.3. It may be written

$$d_\varepsilon(z)(t) = \varepsilon t + \sup_{s \geq t} \Big\{ z(s) - \varepsilon s \Big\} - \sup_{s \geq 0} \Big\{ z(s) - \varepsilon s \Big\}.$$

Observe that $d_\varepsilon(z)(s) - d_\varepsilon(z)(t) \leq \varepsilon(s - t)$ for $s \geq t$. Consider now

$$y(t) = \varepsilon t - d_\varepsilon(z)(t).$$

Then $y(s) - y(t) \geq 0$ for all $s \geq t$. Finally define

$$x(t) = d_{2\varepsilon}(y)(t).$$

It can readily be verified that $x(s) - x(t) \geq 0$ for $s \geq t$, and also as before that $x(s) - x(t) \leq 2\varepsilon(s - t)$ for $s \geq t$. Finally, let

$$w(t) = x(t) - \varepsilon t$$

so that

$$|w(s) - w(t)| \leq \varepsilon |s - t|.$$

Write f for the map $z \mapsto w$. If $z \in \mathcal{X}_0$ then $d_\varepsilon(z) \in \mathcal{X}_0$, so $y \in \mathcal{X}_\varepsilon$, and $x \in \mathcal{X}_\varepsilon$, and $w \in \mathcal{X}_0$, so $f : \mathcal{X}_0 \to \mathcal{X}_0$. Furthermore, f is continuous since all the constituent maps are continuous. (For the relevant results about d_ε, see Section 5.3 below.)

Now, let Z be a standard fBm with Hurst parameter H, and let $W = f(Z)$. It can be checked that

$$N^{-1} W^{\circ N} = f(N^{-1} Z^{\circ N}).$$

(For the relevant calculation for d_ε, see Section 5.3.) We now want to use Lemma 1 to show that $N^{-1}W^{\circ N}$ has Hurstiness H, which means we need to find some input path z with finite rate function, such that $f(z)$ is not equal to $f(0e)$, $0e$ being the zero-value for fBm. Pick any $\hat{z}$ with finite non-zero rate function. From the proof of Lemma 8, we see that $d_\varepsilon(\hat{z}) \neq d_\varepsilon(0e)$; nor do the trivial maps to y and w make the sample path trivial. Thus $f(\hat{z}) \neq f(0e)$.

So we have exhibited a process W which has Hurstiness H, but for which $|W(s) - W(t)| \leq \varepsilon|s-t|$. By making H large and ε small, we have a process which is arbitrarily Hursty and arbitrarily smooth.

4.3 Form of the rate function

We briefly note a consequence of Hurstiness. If the traffic process X has Hurstiness H and the corresponding LDP has rate function I, then it is shown in [6, Lemma 8.2] that

$$I(\alpha x^{\circ 1/\alpha}) = \alpha^{2(1-H)} I(x) \quad \text{for any } \alpha > 0.$$

We will not use this result here. The proof is very similar to that of Theorem 5 below.

5 Hurstiness in queueing systems

Suppose we feed an arrival process A with Hurstiness H into a queueing system. We can say a great deal about the scaling behaviour of queues in the network, just from the fact of Hurstiness. In this section we will explore this for a selection of increasingly complicated queueing systems.

5.1 Queue size

Consider first a stand-alone queue with an infinite buffer and constant service rate C, fed by an arrival process A which has Hurstiness H. Suppose that the LDP for $N^{-1}A^{\circ N}$ holds in the space $\mathcal{C}_\mu$, and that $C > \mu$. The queue size at time 0 may be written as

$$Q = f(A) = \sup_{t \geq 0} A(-t, 0] - Ct, \tag{6}$$

and the function f is continuous on $\mathcal{C}_\mu$ [6, Theorem 5.3]. It is easy to see from (6) that $Q/N = f(N^{-1}A^{\circ N})$. Using Lemma 1 we find the following LDP for Q/N:

Theorem 5 *Let I be the rate function for $N^{-1}A^{\circ N}$.*

i. Q/N satisfies an LDP with speed $N^{2(1-H)}$ and good rate function $J(q) = \inf_{a: f(a)=q} I(a)$;

ii. $J(\alpha q) = \alpha^{2(1-H)} J(q)$ for any $\alpha > 0$, and $J(q) = 0$ if and only if $q = 0$;

iii. *If there exists some sample path $\hat{a}$ with $I(\hat{a}) < \infty$ and $f(\hat{a}) > 0$ then $J(q) = q^{2(1-H)}\delta$ for some $0 < \delta < \infty$, and Q/N has Hurstiness H.*

Proof. Claim (i). This is a direct consequence of the contraction principle.

Claim (ii). Let $R = Q/\alpha$. Now R/N is a simple continuous function of Q/N, so we can apply the contraction principle to find that R/N satisfies an LDP with speed $N^{2(1-H)}$ and good rate function

$$J'(r) = \inf_{q: q/\alpha = r} J(q) = J(\alpha r).$$

We can find an LDP for R/N in another way. The LD upper bound for Q/N says that for any closed set B

$$\limsup_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}(Q/N \in B) \le -\inf_{q\in B} J(q).$$

Let $N = \alpha L$. Rewriting this equation,

$$\limsup_{L\to\infty} \frac{1}{L^{2(1-H)}} \log \mathbb{P}(R/L \in B) \le -\inf_{q\in B} \alpha^{2(1-H)} J(q).$$

A LD lower bound for R/N can be found similarly. We conclude that R/N satisfies an LDP with speed $N^{2(1-H)}$ and good rate function

$$J''(r) = \alpha^{2(1-H)} J(r).$$

These two LDPs for R/N hold in $\mathbb{R}^+$, which is regular. By uniqueness of the rate function, $J'(r) = J''(r)$. This proves the first part of the claim.

For the second part of the claim, we will first prove that I has zero-value μe. Let $X^N = N^{-1} A^{\circ N}(-1, 0]$ and let $Y^N = N^{-1} A^{\circ N}(-t, 0]$. By the contraction principle, X^N satisfies a large deviations principle with some good rate function $J_1(x)$ and Y^N does too with some good rate function $J_t(y)$. By Lemma 1(i), the zero-value for J_1 is $\hat{a}(-1, 0]$ and that for J_t is $\hat{a}(-t, 0]$, where $\hat{a}$ is the zero-value for I.

Now, $Y^N = tX^{tN}$, and by a similar argument to that above $J_t(y) = t^{2(1-H)} J_1(y/t)$. Pick $y = \hat{a}(-t, 0]$; then $J_t(y) = 0$, hence $J_1(y/t) = 0$, hence $y/t = \hat{a}(-t, 0]$; hence $\hat{a}(-t, 0] = t\hat{a}(-1, 0]$. Indeed, since t was arbitrary, $\hat{a} = \hat{a}(-1, 0]e$. Since $\hat{a} \in \mathcal{C}_\mu$, it must be that $\hat{a} = \mu e$.

Now we can finish proving the claim. By Lemma 1(i), $J(q) = 0$ if and only if $q = f(\mu e)$, which is equal to zero since $\mu < C$.

Claim (iii). The assumption that there exists such a sample path $\hat{a}$ implies, by Lemma 1(ii), that J is non-trivial and that Q/N has Hurstiness H.

It remains to establish the form of the rate function. By (ii), $J(0) = 0$ and $J(q) = q^{2(1-H)} J(1)$ where $J(1) > 0$ and possibly $J(1) = \infty$. But if $J(1) = \infty$ then this rate function would be trivial, which is not the case. Therefore $J(1) < \infty$. □

It's helpful for the intuition to write the LDP for Q/N another way:

Lemma 6 *In the setting of Theorem 5,*

$$\lim_{q\to\infty} \frac{1}{q^{2(1-H)}} \log \mathbb{P}(Q > q) = -\delta$$

for some $0 < \delta \le \infty$*.*

Proof. Consider the open interval $B = (1, \infty)$ and its closure $\bar{B} = [1, \infty)$. From the equation for J,

$$\inf_{r\in B} J(r) = \inf_{r\in \bar{B}} J(r) = J(1),$$

so the LD upper and lower bounds agree, and so

$$\lim_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}(Q/N > 1) = -J(1).$$

Letting $\delta = J(1)$ and relabelling N as q we find

$$\lim_{q\to\infty} \frac{1}{q^{2(1-H)}} \log \mathbb{P}(Q > q) = -\delta.$$

Finally, $\delta > 0$ by Theorem 5(ii). □

Finally, a remark about non-triviality. It may be that $\delta = \infty$, or equivalently $J(q) = \infty$ for all $q > 0$. For example, consider two queues in tandem, let the upstream queue have service rate C and the downstream queue have service rate $C' > C$. Suppose the upstream queue is fed by a fBm source with Hurstiness H and mean rate $\mu < C$. We will see in Section 5.3 that the output process, call it A', also has Hurstiness H and mean rate μ. Since A' is rate-limited, the downstream queue can never overflow. So if Q is the size of the downstream queue, then Q/N satisfies a trivial LDP with rate function $J(q) = \infty$ for all $q > 0$.

In general, though, if $J(q) = \infty$ for all $q > 0$, it does not necessarily mean that the queue never builds up, merely that the tail of the distribution is smaller than can be measured by an LDP at speed $N^{2(1-H)}$. In such cases we need more refined techniques to learn about the actual tail of the distribution.

5.2 Aggregates

Theorem 7 *Let A and B be two traffic processes, the former having Hurstiness H and lying in $\mathcal{C}_\mu$, and the latter having Hurstiness G and lying in $\mathcal{C}_\nu$.*

i. If A and B are independent then $A + B$ has Hurstiness $H \vee G$.

ii. If $H \ne G$ and $(N^{-1}A^{\circ N}, N^{-1}B^{\circ N})$ satisfies an LDP at speed $N^{2(1-H\vee G)}$ with some good rate function K, then $A + B$ has Hurstiness $H \vee G$.

Here, $H \vee G$ means $\max(H, G)$.

Proof. Suppose that the LDP for $N^{-1}A^{\circ N}$ has good rate function I and zero-value a_0, and that that for $N^{-1}B^{\circ N}$ has good rate function J and zero-value b_0.

Claim (i). Consider first the case that $H \neq G$, and without loss of generality suppose $H < G$. By Lemma 2, $N^{-1}A^{\circ N}$ satisfies an LDP at speed $N^{2(1-G)}$ with trivial good rate function

$$I'(a) = \begin{cases} 0 & \text{if } a = a_0 \\ \infty & \text{otherwise.} \end{cases}$$

Since this LDP holds in $\mathcal{C}_\mu$, and that for $N^{-1}B^{\circ N}$ holds in $\mathcal{C}_\nu$, both of which are regular Hausdorff spaces, the pair $(N^{-1}A^{\circ N}, N^{-1}B^{\circ N})$ satisfies an LDP in $\mathcal{C}_\mu \times \mathcal{C}_\nu$ with good rate function $K(a,b) = I'(a) + J(b)$. We can now refer to (ii).

Consider next the case that $H = G$. Again, the pair $(N^{-1}A^{\circ N}, N^{-1}B^{\circ N})$ satisfies an LDP in $\mathcal{C}_\mu \times \mathcal{C}_\nu$, this time with good rate function $K(a,b) = I(a) + J(b)$. Let $C = A + B$. By the contraction principle, $N^{-1}C^{\circ N}$ satisfies an LDP with speed $N^{2(1-H)}$ and good rate function

$$L(c) = \inf_{a,b:a+b=c} I(a) + J(b).$$

We will now argue that L is non-trivial.

Suppose that $L(c) = 0$ for some c. Since L is good, and the set $\{(a,b) : a + b = c\}$ is closed, $c = \hat{a} + \hat{b}$ for some $\hat{a}$ and $\hat{b}$, and $I(\hat{a}) + J(\hat{b}) = 0$, thus $I(\hat{a}) = J(\hat{b}) = 0$. Since I and J are both non-trivial, $\hat{a}$ is equal to the zero-value a_0 of I, and b is equal to the zero-value b_0 of J. So $c = a_0 + b_0$, and this is the unique c for which $L(c) = 0$. This fulfils (ii) of the requirement of non-triviality.

Since J is non-trivial, there exists some arrival process $b \neq b_0$ for which $J(b) < \infty$. Now $L(a_0 + b) \leq J(b) < \infty$, and $L(a_0 + b) > 0$ by the earlier discussion. This fulfils (i) of the requirement of non-triviality. We conclude that L is non-trivial, and thus that $A + B$ has Hurstiness $H = G$.

Claim (ii). Suppose without loss of generality that $H < G$. From Lemma 2, $N^{-1}A^{\circ N}$ satisfies an LDP in $\mathcal{C}_\mu$ at speed $N^{2(1-G)}$ with good rate function

$$I'(a) = \begin{cases} 0 & \text{if } a = a_0 \\ \infty & \text{otherwise.} \end{cases}$$

By applying the contraction principle to the pair $(N^{-1}A^{\circ N}, N^{-1}B^{\circ N})$, we also know that $N^{-1}A^{\circ N}$ satisfies an LDP at speed $N^{2(1-G)}$ with good rate function

$$I''(a) = \inf_b K(a,b).$$

By uniqueness of the rate function, these two forms are equal, and so $K(a,b) = \infty$ if $a \neq a_0$.

Similarly, we find that $N^{-1}B^{\circ N}$ satisfies an LDP at speed $N^{2(1-G)}$ with good rate function $J(b)$, and also with good rate function $\inf_a K(a,b)$. By the above $\inf_a K(a,b) = K(a_0,b)$, and so $K(a_0,b) = J(b)$.

Let $C = A + B$. By the contraction principle, $N^{-1}C^{\circ N}$ satisfies an LDP with good rate function

$$L(c) = \inf_{a,b:a+b=c} K(a,b).$$

From what we have learnt about K, this must be equal to $J(b - a_0)$. It is easy to see that L inherits non-triviality from J. $\square$

This result echoes the remark about aggregates of Gaussian processes made in [3, Section 1]. See however [4] which shows that, if we are concerned with more refined probability estimates than are provided by large deviations theory, such results do not always hold.

5.3 Departures

Consider an infinite-buffer queue fed by a source with Hurstiness H and mean rate μ, where the service rate C is greater than μ. The departure map $d : \mathcal{C}_\mu \to \mathcal{C}$ is defined by

$$d(a)(-t, 0] = a(-t, 0] + q_{-t}(a) - q_0(a)$$

where $q_{-t}(a)$ is the queue size at time $-s$,

$$q_{-t}(a) = \sup_{s \geq t} a(-s, -t] - C(s - t).$$

The following lemma shows that the Hurstiness of the departure process $d(A)$ is the same as the Hurstiness of the arrival process A.

Lemma 8 *Suppose that A has Hurstiness H, and that its LDP holds in $\mathcal{C}_\mu$. Suppose that $\mu < C$ and let $D = d(A)$. Then D has Hurstiness H, and its LDP holds in $\mathcal{C}_\mu$.*

Proof. The function d has a convenient scaling property:

$$\begin{aligned} N^{-1}D^{\circ N}(-t, 0] &= N^{-1}D(-Nt, 0] \\ &= N^{-1}\Big\{A(-Nt, 0] + q_{-Nt}(A) - q_0(A)\Big\} \\ &= N^{-1}A^{\circ N}(-t, 0] + q_{-t}(N^{-1}A^{\circ N}) - q_0(N^{-1}A^{\circ N}) \\ &= d(N^{-1}A^{\circ N}). \end{aligned}$$

Therefore the Hurstiness of $N^{-1}D^{\circ N}$ is exactly the Hurstiness of $d(N^{-1}A^{\circ N})$.

Now, it is shown in [6, Theorem 5.13] that d is a continuous function $\mathcal{C}_\mu \to \mathcal{C}_\mu$. It was argued in the proof of Theorem 5(ii) that the zero-value for I must be μe. We will now seek a sample path $\hat{a}$ with $I(\hat{a}) < \infty$ and $d(\hat{a}) \neq d(\mu e)$. If we manage this, then by Lemma 1(ii) $d(N^{-1}A^{\circ N})$ has Hurstiness H.

First note that, from the definition of d, $d(\mu e) = \mu e$.

We know that I is non-trivial. Therefore there exists some $\hat{a}$ such that $0 < I(\hat{a}) < \infty$. Pick any such $\hat{a}$. We will show that $d(\hat{a}) \neq \mu e$.

Suppose to the contrary that $d(\hat{a}) = \mu e$. Rearranging the defining equation for $d(\hat{a})$,

$$\hat{a}(-s,-t] = \mu(s-t) + q_{-s}(\hat{a}) - q_{-t}(\hat{a}).$$

If $q_{-t}(\hat{a}) = 0$ for all t then $\hat{a}(-t,0] = \mu t$, i.e. $\hat{a}$ is the zero-value for I, so $I(\hat{a}) = 0$, which contradicts our choice of $\hat{a}$. So it must be that $q_{-\hat{t}}(\hat{a}) > 0$ for some $\hat{t}$. Now, by [6, Theorem 5.3], the supremum in

$$q_{-\hat{t}}(\hat{a}) = \sup_{s \geq \hat{t}} \hat{a}(-s,-\hat{t}] - C(s-\hat{t})$$

is attained at some $\hat{s}$, and by [6, Lemma 5.4] $q_{-\hat{s}}(\hat{a}) = 0$. The defining equation for $d(\hat{a})$ then says

$$\begin{aligned} d(\hat{a})(-\hat{s},-\hat{t}] &= \hat{a}(-\hat{s},-\hat{t}] + q_{-\hat{s}}(\hat{a}) - q_{-\hat{t}}(\hat{a}) \\ &= C(\hat{s}-\hat{t}). \end{aligned}$$

This is not equal to $\mu(\hat{s}-\hat{t})$, so $d(\hat{a}) \neq \mu e$, which contradicts our supposition that $d(\hat{a}) = \mu e$. This completes the proof. □

5.4 Other derived quantities

It is worth summarizing the form of the proof from Section 5.3, since the same form can be used for many problems. We wanted to know the Hurstiness of D, i.e. of $N^{-1}D^{\circ N}$. We wrote this as a continuous function of the scaled inputs to the system, $N^{-1}D^{\circ N} = f(N^{-1}A^{\circ N})$. This gives an LDP for $N^{-1}D^{\circ N}$. To show that the LDP is non-trivial, we picked a sample input path a such that $0 < I(a) < \infty$ and $f(a) \neq f(\mu e)$, where μe is the zero-value of A. This proves that D has the same Hurstiness as A.

A snappier summary is this: *any continuous non-trivial function of the input has the same Hurstiness as the input.*

In fact, this is exactly the same argument that we used in Section 5.1 to argue that the queue size has the same Hurstiness as the arrival process. We will now use this technique to look at some more complicated scenarios, where it is not always immediately clear which inputs we need to consider.

Priority queues. Consider a priority queue with an infinite buffer (as described, for example, in [6, Section 5.9]). Suppose that the high priority input has Hurstiness H and the low-priority input has Hurstiness G, and that the queue has constant service rate which is sufficient to serve both arrival processes.

Considering the high-priority queue in isolation, the high-priority queue size has Hurstiness H (assuming that it is not trivial, i.e. assuming that there is some input sample path with finite rate function which causes the high-priority queue size to be non-zero). The high-priority departure process has Hurstiness H.

Suppose that $H < G$, and consider an LDP for the pair of arrival processes at speed $N^{2(1-G)}$. The high-priority queue size has Hurstiness H, and the aggregate queue size has Hurstiness G (assuming that both are non-trivial). By Theorem 7(ii) the low-priority queue size, which is the difference between the aggregate and the high-priority queue sizes, has Hurstiness G. It is also simple to check that the low-priority departure process has Hurstiness G.

Suppose that $H > G$, and consider an LDP for the pair of arrival processes at speed $N^{2(1-H)}$. At this speed, we can take the low-priority input to be deterministic, at a rate equal to its mean rate, which we will assume is strictly positive. Assume that the high-priority queue is non-trivial, i.e. that there is a high-priority input sample path which causes the high-priority queue size to be non-zero. Clearly this sample path also causes the low-priority queue size to be non-zero. Therefore the low-priority queue size has Hurstiness H, and the low-priority departure process has too.

Finally, if $H = G$, then (assuming non-triviality) all quantities of interest have the same Hurstiness.

Queues with stochastic service. So far we have assumed a constant service rate C. Suppose instead that a random amount of service $C(-t, 0]$ is offered in the interval $(-t, 0]$, and that the service process C has Hurstiness G. The queue size at time 0 is then

$$q(X) = \sup_{t \geq 0} X(-t, 0]$$

where $X(-t, 0] = A(-t, 0] - C(-t, 0]$. Suppose that the arrival process A has Hurstiness H, and that the mean arrival rate is strictly less than the mean service rate (so that the system is stable).

Using the results for aggregates in Section 5.2, the 'netput' process X has Hurstiness $H \vee G$. To work out the Hurstiness of the queue size, we need to know something about the rate function for $N^{-1} X^{\circ N}$.

First note that the form of this rate function, call it I, depends on which is larger of H and G. As far as large deviations at speed $N^{2(1-H \vee G)}$ are concerned, if $H > G$ then the the service process is essentially deterministic; if $H < G$ then the arrival process is essentially deterministic; and only if $H = G$ are both processes essentially non-deterministic.

Suppose that there is some sample path x with $I(x) < \infty$ and $q(x) > 0$, i.e. that queue buildup is possible at scale $N^{2(1-H \vee G)}$. Then, as in Section 5.1, the queue size has Hurstiness $H \vee G$.

This is not the only case. Suppose for example that the arrival process has Hurstiness H and that the arrival rate is bounded above by μ; suppose that the service process has Hurstiness $G > H$ and that the service rate is bounded below by μ. (Such processes exist, using the construction from Section 4.2.) It is easy to check that the queue is always empty, and that the departure process is exactly the same as the arrival process. The departure process therefore has Hurstiness H, i.e. the Hurstiness of the service has not 'contaminated' the traffic.

Queues with finite buffers. Consider a queue with a finite buffer $B > 0$ fed by a Hursty arrival process. Obviously the queue size is bounded by B, so it does not make sense to inquire about the tail of the queue size distribution. In order to be able to ask interesting questions about Hurstiness, we therefore consider a *sequence* of queues, indexed by N, where the Nth queue has finite buffer NB and $N \to \infty$. What can we say about the asymptotic probability of overflow?

Let A be a traffic process with Hurstiness H. Suppose the LDP for $N^{-1}A^{\circ N}$ holds in $\mathcal{C}_\mu$. We will additionally assume that A has stationary increments, i.e. that $A(-t, 0] \sim A(-s-t, -s]$ for all s and $t > 0$. Let Q be the queue size in an infinite-buffer queue fed by A, and let Q^N be the queue size for a queue with finite buffer NB. We know from Theorem 5 that Q/N satisfies an LDP at speed $N^{2(1-H)}$ with good rate function $J(0) = 0$ and $J(q) = q^{2(1-H)}\delta$ (possibly infinite). It turns out that Q^N/N is closely related to Q/N:

Lemma 9 *In the above setting, Q^N/N satisfies an LDP at speed $N^{2(1-H)}$ with good rate function*

$$\bar{J}(q) = \begin{cases} J(q) & \textit{if } q \leq B \\ \infty & \textit{otherwise.} \end{cases}$$

If $\delta < \infty$ then $\bar{J}$ is non-trivial and so Q^N/N has Hurstiness H.

Sketch proof. It is shown in [6, Section 5.7] that $Q^N/N = \bar{f}(N^{-1}A^{\circ N})$ where $\bar{f}(a)$ is the queue size at time 0 for a queue with finite buffer B, and that this function is continuous on $\mathcal{C}_\mu$. As A has Hurstiness H, then we obtain an LDP for Q^N/N at speed $N^{2(1-H)}$ with some good rate function $\bar{J}$.

The proof of [6, Theorem 7.10] gives us the form of the rate function (even though that proof was concerned with a different context).

Stationarity is used in the following way. Let I be the rate function for $N^{-1}A^{\circ N}$, and let $\hat{a}$ be an optimizing path for $J(q) = \inf_{a: f(a)=q} I(a)$, so that $f(\hat{a}) = q$ and $I(\hat{a}) = J(q)$. (As before, f is the queue-size function for queue with an infinite buffer.) Suppose that the infinite-buffer queue fed by $\hat{a}$ is empty at time $-u$ and then reaches level q at time $-v \leq 0$. Consider the truncated traffic process a' defined by $a'(-t, 0] = \hat{a}(-t-v, -v]$; then the infinite-buffer queue fed by a' is empty at time $-(u-v)$ and does not subsequently reach level q until time 0. The finite-buffer queue fed by a' must do the same, since it has no chance of reaching buffer capacity. Hence $\bar{J}(q) \leq I(a')$. By stationarity $I(a') \leq I(\hat{a})$, and so $\bar{J}(q) \leq J(q)$. □

The conclusion is that

$$\lim_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}(Q^N/N > B - \varepsilon) = -(B-\varepsilon)^{2(1-H)}\delta.$$

To all intents and purposes, this means

$$\log \mathbb{P}\big(\text{queue with buffer size } B \text{ is full}\big) \approx -B^{2(1-H)}\delta$$

(although the technicalities associated with the large deviations lower bound prevent us from writing this as a limit.)

Departures from queues with finite buffers, processor sharing queues, and other challenges. To analyse the departure process from a queue with a finite buffer, and a number of more complicated systems such as processor sharing, there are several technical obstacles to be surmounted.

The first is that we may need to consider a sequence of queueing systems, as described above.

The second is that often the quantity of interest is not a continuous function on $\mathcal{C}_\mu$. In [6, Example 5.5] it is pointed out that this is the case for departures from a queue with a finite buffer. Often, however, the corresponding 'transient' function is continuous. By the transient function we mean the function defined on $\mathcal{C}^T$ for some $T > 0$ rather than on $\mathcal{C}_\mu$. For example, the transient queue size function for an infinite-buffer queue is

$$f(a|_{(-T,0]}) = \sup_{t \le T} a(-t, 0] - Ct,$$

the queue size when the system is started empty at time $-T$. We might then find an LDP for $f(N^{-1}A^{\circ N}|_{(-T,0]})$, which is the scaled queue size at time 0 when the system is started empty at time $-NT$.

The third technical obstacle is that sometimes we do not have an explicit formula for the quantity of interest, we only have a system of integral equations. For example, the integral equations for the transient queue size function are

$$q_{-T} = 0 \quad \text{and} \quad q_t = \int_{s=-T}^{t} \big[\dot{a}_s - C\big]^{+(q_s=0)}\, ds$$

where the notation $[x]^{+(y=0)}$ means

$$[x]^{+(y=0)} = \begin{cases} x^+ & \text{if } y = 0 \\ x & \text{otherwise.} \end{cases}$$

As described in [6, Section 5.10], in the context of processor-sharing queues, it can sometimes be shown that these equations have a unique solution, as long as a is absolutely continuous. The trouble is that some interesting processes are not absolutely continuous; for example, fBm is (almost surely) nowhere differentiable. One way to get around this is to consider a queueing system fed not by the interesting process A but rather by a piecewise linear approximation $\tilde{A}$. In Lemma 13 in the appendix it is shown that if A is an fBm then $N^{-1}\tilde{A}^{\circ N}$ is exponentially equivalent to $N^{-1}A^{\circ N}$, meaning that it satisfies exactly the same LDP.

These three obstacles do not interfere with the abstract idea of Hurstiness, they simply mean that one must be a little more careful in interpreting what Hurstiness means.

6 Many-flows limit

The prototypical traffic flow in the preceeding sections has been fractional Brownian motion. It is by now well-known that fBm can arise as a limit when many flows are superimposed. We can use the idea of Hurstiness, though with a modification, to study this. The key is to use a mixture of two scaling limits, the $\circlearrowleft$ scaling described above, and the $\oplus$ scaling. If X is a random traffic process, then $X^{\oplus M}$ is the aggregate (i.e. sum) of M independent copies of X.

The modified idea of Hurstiness is this. Consider the doubly-indexed family of random processes

$$\frac{1}{MN} X^{\oplus M \circlearrowleft N}.$$

We will suppose that this family satisfies an LDP in $\mathcal{C}_\mu$ as $M \wedge N \to \infty$ with some good rate function I at speed $MN^{2(1-H)}$. We will also suppose that I is non-trivial in the sense of Definition 1.

(We have chosen to write this in the context of traffic processes rather than giving an abstract definition. It is possible to define Hurstiness abstractly for doubly-indexed sequences of random variables, though we do not at the moment see any benefit in doing so.)

We will first discuss when such an LDP holds, and we will then use the contraction principle to investigate an LDP for queue size. Further elaboration along the lines of sections 2 and 5 is possible, though we do not give it here as the details are much the same.

6.1 Hurstiness of traffic processes

The characteristics of $X^{\oplus M \circlearrowleft N}$ for general traffic processes X are not yet well-understood. For the case where X is Gaussian, however, a comprehensive picture has been provided by Dieker [3]; the result of his we will be using is given in Theorem 12 in the appendix. We will now illustrate how this theorem tells us about Hurstiness. We conjecture that similar results hold for non-Gaussian processes.

By way of example, let X be the Gaussian approximation to a single $M_\lambda/G/\infty$ traffic flow, where the on-time G has distribution $\mathbb{P}(G > t) = (t+1)^{2H-3}$ for $H \in (1/2, 1)$, as described by [7]. Then the mean rate is $\mu = \lambda \mathbb{E} G$ and the variance $\sigma^2(t) = \operatorname{Var} X(-t, 0]$ is

$$\sigma^2(t) = \frac{2\lambda}{2H(2H-1)(2H-2)} \Big[1 - (t+1)^{2H} + 2Ht \Big].$$

This is regularly varying with index H; in fact $\sigma^2(t) \sim \rho^2 t^{2H}$. The following argument works whenever $\sigma^2(t) \sim \rho^2 t^{2H}$, not just for this example.

It is convenient to have zero mean, so let $Y(-t, 0] = X(-t, 0] - \mu t$. By Theorem 12, if L_N is a sequence with $L_N \to \infty$ as $N \to \infty$ then the sequence

$$\frac{1}{L_N \sigma(N)} Y^{\circlearrowleft N}$$

satisfies a large deviations principle in $\mathcal{C}_0$ with speed L_N^2, and the rate function is exactly the rate function for fBm which did not appear in Lemma 4.

To turn this into the form we want, we will reparameterize. Let $L = \sqrt{M}N/\sigma(N)$. Then $L \to \infty$ as $M \wedge N \to \infty$. We obtain an LDP for

$$\frac{1}{\sqrt{M}N} Y^{\circlearrowleft N}$$

with speed $MN^2/\sigma^2(N)$. But Y is Gaussian with zero mean, so $Y^{\oplus M} \sim \sqrt{M}Y$, and so the LDP can be written as an LDP for

$$\frac{1}{MN} X^{\oplus M \circlearrowleft N}$$

at the same speed $MN^2/\sigma^2(N)$. Also, since $\sigma^2(N) \sim \rho^2 N^{2H}$, we can equivalently write the LDP as having speed $MN^{2(1-H)}$. Finally, we can add the mean back: $(MN)^{-1}X^{\oplus M \circlearrowleft N}$ is a continuous function of $(MN)^{-1}Y^{\oplus M \circlearrowleft N}$, so it too satisfies an LDP at speed $MN^{2(1-H)}$, and the rate function is the rate function for an fBm with drift μ; this is a non-trivial rate function.

We conclude that $(MN)^{-1}X^{\oplus M \circlearrowleft N}$ has Hurstiness H.

6.2 Implications for queues

Consider a queue serving an aggregate of M independent copies of some flow A, and let the service rate be MC, i.e. the service rate is scaled up in proportion to the number of flows. Take the buffer to be infinite. The queue size at time 0 is

$$Q^M = \sup_{t \geq 0} A^{\oplus M}(-t, 0] - MCt$$

and so

$$\begin{aligned} \frac{1}{MN} Q^M &= \sup_{t \geq 0} \frac{1}{MN} A^{\oplus M}(-t, 0] - \frac{1}{N} Ct \\ &= \sup_{s \geq 0} \frac{1}{MN} A^{\oplus M \circlearrowleft N}(-s, 0] - Cs \\ &= f\Big((MN)^{-1} A^{\oplus M \circlearrowleft N}\Big) \end{aligned}$$

where f is the usual queue size function for a queue with constant service rate C. Assuming that $(MN)^{-1}A^{\oplus M \circlearrowleft N}$ has Hurstiness H, and assuming that C is greater than the mean rate, we find that Q^M/MN satisfies an LDP at speed $MN^{2(1-H)}$. If the rate function for queue size is non-trivial, the queue size has Hurstiness H.

The argument of Theorem 5 works here too, and implies that the rate function for queue size satisfies

$$J(\alpha q) = \alpha^{2(1-H)} J(q) \quad \text{for any } \alpha > 0.$$

As in Lemma 6, we also have

$$\lim_{M \wedge N \to \infty} \frac{1}{MN^{2(1-H)}} \log \mathbb{P}(Q^M/MN > q) = -\delta q^{2(1-H)} \tag{7}$$

for some $\delta > 0$. If $\delta < \infty$, we may heuristically rewrite this as

$$\log \mathbb{P}(Q^M > Mb) \approx -Mb^{2(1-H)}\delta \tag{8}$$

where $b = Nq$, i.e. b is the amount of buffer space per flow.

This echoes the result of [7, Example 4.4], for the Gaussian approximation to a $M_\lambda/G/\infty$ flow, but is here expressed as a mixed limit rather than as an asymptotic of the rate function. See however [5], which studies asymptotics of the rate function for the actual $M_\lambda/G/\infty$ system, and indicates that when $\mathbb{P}(G > t) = (t+1)^{2H-3}$ (as for the example in Section 6.1) then the estimate in (8) might perhaps be replaced by $-M\log(1+b)\delta$,

Rate function for fBm. We noted in Section 6.1 that, for aggregates of Gaussian traffic processes, the resulting rate function is generally that of an fBm. Therefore any conclusions we drew in Section 5.1 about rate functions for derived quantities for a *single* fBm traffic process apply equally to *aggregates* of Gaussian processes—the optimization problem arising from the contraction principle is exactly the same in each case.

Other derived quantities. The contraction principle can also tell us about other derived quantities, such as the departure process or the queue size in a downstream queue. We only need to be slightly cautious in understanding the scaling involved. For example, if d_C is the departure map for a queue with constant service rate C, then the scaled departure process

$$d_C\Big((MN)^{-1}A^{\oplus M \circ N}\Big)$$

is equivalent to

$$(MN)^{-1}(D^M)^{\circ N} \quad \text{where} \quad D^M = d_{MC}(A^{\oplus M}).$$

In other words, the LDP for the scaled departure process tells us about the aggregate departure process from the scaled-up queue. In other words, just as we did when looking at queues with finite buffers in Section 5.4, we need to bear in mind that we are considering a sequence of queueing systems, in this case indexed by the number of flows, not just a single queueing system.

7 Conclusion

This concept of Hurstiness is a very convenient way of studying long-range dependence in queueing networks. It lends itself better to calculations than does the conventional definition (1). It can tell us about scaling behaviour in networks, without requiring that we bother with probabilistic calculations or that we solve tedious optimization problems.

It leads to a very simple 'calculus of Hurstiness'. Any continuous non-trivial function of an input has the same Hurstiness as that input. To decide if the queue size for a particular queue in a network is a non-trivial function, consider all the flows that might affect it, treat all flows with Hurstiness less than the maximum Hurstiness as being deterministic, and work out whether bursts of traffic in the remaining flows can cause that queue to overflow; if they can then that queue is a non-trivial function. The rate function for the input traffic says which sorts of bursts are possible, and so it may be useful to know a little about the rate function to work out non-triviality (although in this paper we have given abstract arguments which don't depend on the particular rate function). A queue with Hurstiness H has a tail which decays like $q^{2(1-H)}$.

However, Hurstiness is rather limited. The output of a queue with constant service rate is just as Hursty as the input, even though common sense screams at us that it is smoother. Indeed, a flow can be arbitrarily Hursty yet arbitrarily smooth. Given that Hurstiness is not affected by any of the ways in which we might reasonably like to shape traffic, Hurstiness is too coarse an idea.

We believe that Hurstiness and long-range-dependence are interesting theoretical constructions, with some elegant mathematics, but that they are not terribly informative in practice. In practice, we believe, formulae based on simple mean and variance traffic statistics, such as are given in [6, Chapter 10], will prove to be more useful.

Acknowledgements

Thanks to Ton Dieker for his helpful comments.

Appendix

7.1 LDP for Gaussian process

This result concerning Gaussian processes is taken from [1]. Let $Z(t)$, $t \in \mathbb{R}^+$, be a centred Gaussian process with continuous sample paths and stationary increments, and $Z(0) = 0$. Centred means that $\mathbb{E}Z(t) = 0$. Let v be the variance function, $v(t) = \mathbb{E}Z(t)^2$, and let $\gamma(s,t)$ be the covariance function, $\gamma(s,t) = \mathbb{E}Z(s)Z(t)$. Assume that

$$\lim_{t\to\infty} \frac{v(t)}{t^\alpha} = 0$$

for some $\alpha < 2$, and that $v(t) > 0$ for all $t > 0$.

Theorem 10 *In the above setting, the sequence $Z/\sqrt{N}$ satisfies a large deviations principle in $\mathcal{C}_0$ at speed N, with a good rate function. The rate function is*

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

where R and $\|\cdot\|_R$ are defined as follows. Start with the space of functions $\Gamma_t(\cdot) = \gamma(t,\cdot)$, and the inner product $\langle \Gamma_t, \Gamma_s \rangle_R = \gamma(s,t)$. Close this space with *linear combinations and complete it with respect to the norm* $\|\cdot\|_R^2 = \langle\cdot,\cdot\rangle_R$ *to obtain the space* R*. This norm has the property that* $\langle f, \Gamma_t\rangle = f(t)$ *for* $f \in R$*. The space* R *is known as the reproducing kernel Hilbert space.*

For the purposes of this paper, all we need to know is that this rate function is non-trivial:

Lemma 11 *In the above setting,*
i. there exists some $\hat{z}$ *such that* $0 < I(\hat{z}) < \infty$*;*
ii. $I(z) = 0$ *if and only if* $z = 0e$*;*
iii. $I(\xi z) = \xi^2 I(z)$*.*

Proof. Claim (i). Simply pick $\hat{z} = \Gamma_t$ for any $t > 0$.

Claim (ii). It is clear that $I(0e) = 0$. Suppose that $z \neq 0e$, so that $z(\hat{t}) = \xi \neq 0$ for some $\hat{t} > 0$. Consider $Z(\hat{t})/\sqrt{N}$. Since $Z(\hat{t})$ is a normal random variable, $Z(\hat{t})/\sqrt{N}$ satisfies an LDP with good rate function

$$J(\eta) = \frac{\eta^2}{2v(t)}.$$

On the other hand, by the contraction principle from $Z(\cdot)/\sqrt{N}$, $Z(\hat{t})/\sqrt{N}$ satisfies an LDP with good rate function

$$J'(\eta) = \inf_{z \in \mathcal{C}_0 : z(\hat{t}) = \eta} I(z).$$

These two rate functions must be identical. Furthermore, $J(\xi) > 0$ and $J'(\xi) \leq I(z)$, so $I(z) > 0$. We have thus proved that if $z \neq 0e$ then $I(z) > 0$.

Claim (iii). This is trivial, because the rate function is $\|\cdot\|_R^2$. □

7.2 Aggregates of LRD Gaussian flows

Let X be a Gaussian process with mean rate μ and variance function $\sigma^2(t) = \operatorname{Var} X(t)$, and with stationary increments. In fact, it's easier to work with centred processes, so assume $\mu = 0$. Assume that σ^2 is regularly varying with index $0 < 2H < 2$, i.e. that

$$\lim_{N\to\infty} \frac{\sigma^2(Nt)}{\sigma^2(N)} = t^{2H}.$$

Also assume that $\sigma^2(t)$ is continuous, and that

$$\lim_{t\to 0} \sigma^2(t) |\log t|^{1+\varepsilon} < \infty \quad \text{for some } \varepsilon > 0.$$

Then [3, Corollary 3] tells us the following:

Theorem 12 *In the above setting, for any sequence* L_N *with* $L_N \to \infty$ *as* $N \to \infty$*, the sequence*

$$\frac{1}{\sqrt{L_N}\sigma(N)} X^{\circ N}$$

satisfies a large deviations principle in $\mathcal{C}_0$ *with speed* L_N*, and the rate function is the rate function for a standard fractional Brownian motion with Hurst parameter* H*.*

7.3 Piecewise linear approximations to fBm

Let X be a Gaussian process which satisfies the conditions of Theorem 10. Suppose also that X has stationary increments and that $\operatorname{Var} X(t) \to 0$ as $t \to 0$. Let $\tilde{X}$ be a piecewise linear approximation to X, for which each piece has length less than Δ. That is, there is a sequence $0 = t_0 < t_1 < \cdots$ with $t_i \to \infty$ and $t_{i+1} - t_i < \Delta$, and

$$\tilde{X}(t) = \lambda X(t_i) + (1-\lambda) X(t_{i+1})$$

where $\lambda = \lambda(t)$ is chosen such that, for some i, $t = \lambda t_i + (1-\lambda) t_{i+1}$ and $0 \leq \lambda < 1$.

Lemma 13 *In the above setting, the sequence $N^{-1}\tilde{X}^{\circlearrowleft N}$ is exponentially equivalent to $N^{-1}X^{\circlearrowleft N}$ at speed $N^{2(1-H)}$ in the space $\mathcal{C}_0$.*

Proof. We need to show that

$$\limsup_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}\big(\|N^{-1}X^{\circlearrowleft N} - N^{-1}\tilde{X}^{\circlearrowleft N}\| > \delta\big) \tag{9}$$

is equal to $-\infty$ for all $\delta > 0$.

Write p_N for the map which takes a process x to its piecewise linear approximation, approximated at $0 = t_0/N < t_1/N < t_2/N < \cdots$. Then

$$\begin{aligned}
(9) &= \limsup_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}\big(N^{-1}X^{\circlearrowleft N} \in \big\{x : \|x - p_N(x)\| > \delta\big\}\big) \\
&\leq \limsup_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}\big(N^{-1}X^{\circlearrowleft N} \in B_r(\delta)\big) \qquad (10)
\end{aligned}$$

for all r, where

$$B_r(\delta) = \bigcup_{s\geq r} \big\{x : \|x - p_s(x)\| > \delta\big\}$$

We will estimate (10) using the large deviations upper bound

$$\limsup_{N\to\infty} \frac{1}{N^{2(1-H)}} \log \mathbb{P}\big(N^{-1}X^{\circlearrowleft N} \in B_r(\delta)\big) \leq -\inf_{x\in\bar{B}_r(\delta)} I(x). \tag{11}$$

So suppose that $x \in \bar{B}_r(\delta)$. Then there exists a sequence $x_n \to x$, $x_n \in B_r(\delta)$. Since $x_n \in B_r(\delta)$, there exists $s_n \geq r$ such that

$$\delta < \|x_n - p_{s_n}(x_n)\|.$$

But this is

$$\leq \|x_n - x\| + \|x - p_{s_n}(x)\| + \|p_{s_n}(x) - p_{s_n}(x_n)\|.$$

Since $x_n \to x$, and therefore $\|p_{s_n}(x) - p_{s_n}(x_n)\| \to 0$, it must be that $\|x - p_{s_n}(x)\| \geq \delta$ for n sufficiently large. That is to say, there is some $s_n \geq r$ such that $\|x - p_s(x)\| \geq \delta$.

Given that $\|x - p_s(x)\| \geq \delta$ for some $s \geq r$, there must be some t and $u < \Delta/r$ such that $|x(t+u) - x(t)| \geq \delta$. By stationarity, we may as well assume that $t = 0$. The infimum of the rate function for this event is

$$\inf_{x:|x(u)|\geq\delta} I(x) = \frac{\delta^2}{2 \operatorname{Var} X(u)}.$$

Since $\operatorname{Var} X(t) \to 0$ as $t \to 0$, this quantity $\to \infty$ as $r \to \infty$. Taking the limit of (11) as $r \to \infty$ completes the proof. □

References

[1] Ron Addie, Petteri Mannersalo, and Ilkka Norros. Performance formulae for queues with Gaussian input. In *Proceedings of ITC 16, Edinburgh*, pages 1169–1178, 1999. URL `http://www.vtt.fi/tte/staff/ttemat/research.html`.

[2] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, 2nd edition, 1998.

[3] A. B. Dieker. Conditional limit theorems for queues with Gaussian input, a weak convergence approach. URL `http://homepages.cwi.nl/~ton/publications/`. To appear in Stochastic Processes and their Applications. Available as CWI technical report PNA-E0306, 2003.

[4] Ton Dieker. Reduced-load equivalence for queues with Gaussian input. URL `http://homepages.cwi.nl/~ton/publications/`. Submitted for publication, 2004.

[5] N. G. Duffield. Queueing at large resources driven by long-tailed $M/G/\infty$-modulated processes. *Queueing Systems*, 28, 1998. URL `http://www.research.att.com/~duffield/pubs/`.

[6] A. J. Ganesh, N. O'Connell, and D. J. Wischik. *Big Queues*. Lecture Notes in Mathematics. Springer, 2004. URL `http://www.bigqueues.com`.

[7] M. Mandjes. A note on the benefits of buffering. *Stochastic Models*, 2003. URL `http://db.cwi.nl/rapporten/abstract.php?abstractnr=1512`. Available as CWI technical report PNA-E0407.

[8] Neil O'Connell. Large deviations with applications to telecommmunications. Lecture course at Uppsala University, November 1999. URL `http://www.maths.lth.se/matstat/research/asn/ldcourse99.html`.