Centrum voor Wiskunde en Informatica

REPORTRAPPORT

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA

On the correlation structure of Gaussian queues

A. Es-Saghouani, M.R.H. Mandjes

REPORT PNA-R0706 MAY 2007

Centrum voor Wiskunde en Informatica (CWI) is the national research institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organisation for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3703

On the correlation structure of Gaussian queues

ABSTRACT
In this paper we study Gaussian queues (that is, queues fed by Gaussian processes, such as fractional Brownian motion (fBm) and the integrated Ornstein-Uhlenbeck (iOU) process), with a focus on the correlation structure of the workload process. The main question is: to what extent does the workload process inherit the correlation properties of the input process? We first present an alternative definition of correlation that allows (in asymptotic regimes) explicit analysis. For the special cases of fBm and iOU we analyze the behavior of this metric under a many-sources scaling. Relying on (the generalized version of) Schilder's theorem, we are able to characterize its decay. We observe that the correlation structure of the input process essentially carries over to the workload process.

2000 Mathematics Subject Classification: 60K25
Keywords and Phrases: Gaussian processes, fractional Brownian motion, workload process
Note: Both authors are with the Korteweg-de Vries Institute for Mathematics, Plantage Muidergracht 24, 1018 TV Amsterdam, the Netherlands; {aessagho|mmandjes}@science.uva.nl. M. Mandjes is also with CWI, Amsterdam, the Netherlands, and Eurandom, Eindhoven, the Netherlands.

On the correlation structure of Gaussian queues

Abdelghafour Es-Saghouani & Michel Mandjes*

May 14, 2007

Abstract

In this paper we study Gaussian queues (that is, queues fed by Gaussian processes, such as fractional Brownian motion (fBm) and the integrated Ornstein-Uhlenbeck (iOU) process), with a focus on the correlation structure of the workload process. The main question is: to what extent does the workload process inherit the correlation properties of the input process? We first present an alternative definition of correlation that allows (in asymptotic regimes) explicit analysis. For the special cases of fBm and iOU we analyze the behavior of this metric under a many-sources scaling. Relying on (the generalized version of) Schilder's theorem, we are able to characterize its decay. We observe that the correlation structure of the input process essentially carries over to the workload process.

1 Introduction

Traffic measurement studies have provided convincing statistical evidence that in various networking environments traffic exhibits strong dependence over a wide range of time-scales. These studies, starting off in the early 1990s with the famous article by Leland *et al.* [12] on Ethernet traffic, showed that traffic rate process was *long-range dependent*: with $X(t)$ the traffic rate at time t, the autocorrelation function $c(T)$ of the traffic rate (i.e., the correlation coefficient between $X(0)$ and $X(T)$) vanishes extremely slowly as a function of the lag T – more precisely: $c(T)$ decays so slowly that $\sum_{T\in\mathbb{N}} c(T) = \infty$. The measurements also indicated that often the traffic rate behaves,

*Both authors are with Korteweg-de Vries Institute for Mathematics, Plantage Muidergracht 24, 1018 TV Amsterdam, the Netherlands; {aessagho|mmandjes}@science.uva.nl. M. Mandjes is also with CWI, Amsterdam, the Netherlands, and EURANDOM, Eindhoven, the Netherlands.

approximately, according to a *self-similar* pattern: irrespective of the time-scale considered, the nature of traffic rate fluctuations looks strikingly similar.
This explains the search for statistical models that are in line with the properties mentioned above. Notably, the traffic models that were predominantly used till the mid-1990s did not allow for any long-range dependence: they usually corresponded to short-range dependent traffic processes (such as Poisson processes, Markov-modulated Poisson processes, or exponential on-off sources). In the late 1990s, Gaussian traffic models have gained more interest and popularity for modeling network traffic. One of their attractive features is that they cover a broad variety of correlation structures, ranging from short-range (e.g., the integrated Ornstein-Uhlenbeck process, Brownian motion) to long-range dependent (e.g., fractional Brownian motion with *Hurst parameter* $H > \frac{1}{2}$, see [12].) The use of Gaussian traffic models is justified as long as the aggregation is sufficiently large, both in number of flows and time, as argued in [11].
The fact that network traffic is long-range dependent is of crucial importance from the perspective of traffic engineering in communication networks. Where short-range dependent models usually lead to buffer overflow probabilities that decay exponentially in the buffer size, long-range dependent models are considerably less benign: in case of fractional Brownian motion input with Hurst parameter H this decay is 'Weibullian' [10, 17] (that is, roughly like $\exp(-\alpha B^{2-2H})$, for some $\alpha > 0$, which is slower than exponential for $H > \frac{1}{2}$), or even polynomial [19, 22] (e.g., for on-off sources with regularly varying on-times). In other words: modeling traffic by short-range dependent process would lead to estimates of the overflow probability that are considerably too optimistic.

For Gaussian queues (that is, queues fed by Gaussian processes), so far primary interest lied in the characterization of the buffer overflow probability. Notably, in two limiting regimes asymptotic results were obtained: in the large-buffer regime (where the buffer threshold grows large), and in the many-sources regime (in which the number of Gaussian inputs grows large, and the buffer and service speed are scaled accordingly [21]). Without exhaustively mentioning all relevant contributions, logarithmic asymptotics for the large-buffer case are due to [5, 10], whereas exact asymptotics can be found in, e.g., [9, 16], and in the many-sources regime logarithmic asymptotics are in [1, 4] and the exact asymptotics in [6].
Remarkably, to our best knowledge, no attention was paid to the characterization of the *correlation structure of the workload process* of Gaussian queues, despite its evident relevance for engineering purposes. Seen from a more mathematical angle, an interesting fundamental question is: to what extent the characteristics of the input process are inherited by the workload process? Or, put

differently, does long-range dependent input give rise to a long-range dependent workload process? Our paper shows that indeed for fractional Brownian motion (in the sequel abbreviated to fBm) and integrated Ornstein-Uhlenbeck (iOU) the correlation structure of the workload process strongly resembles that of the input process: it exhibits Weibullian decay for fBm, and exponential decay for iOU.
When analyzing the correlation structure of queues with Gaussian inputs, one major difficulty is that no explicit expression is available for the probability that the queue content buffer reaches some level p (except for the case of Brownian motion input). This means that it will be hard, if not impossible, to characterize, with Q_t denoting the workload at time t, the covariance

$$\mathbb{C}\mathrm{ov}(Q_0, Q_T) = \mathbb{E}(Q_0 Q_T) - \mathbb{E}Q_0 \cdot \mathbb{E}Q_T = \mathbb{E}(Q_0 Q_T) - (\mathbb{E}Q_0)^2, \tag{1}$$

or the corresponding correlation coefficient. To overcome this problem we have chosen the following solution:

- We work in the many-sources asymptotic regime [21] that was mentioned above. As a consequence, we can use an extensive set of useful techniques, most notably (sample-path) large-deviations results, in particular (the generalized version of) Schilder's theorem.

  More precisely, the setting we consider is as follows: we let n i.i.d. Gaussian processes $A_1(\cdot), \ldots, A_n(\cdot)$ feed into a queue in which both the service speed and the buffer content are scaled by n; we denote the workload of the resulting queueing system at time t by Q_t^n. Our results are asymptotic in n.

- We choose a measure for correlation that is more practical than (1). This new metric measures the difference between $\log \mathbb{P}(Q_0^n > np, Q_T^n > nq)$ and $\log\left(\mathbb{P}(Q_0^n > np)\mathbb{P}(Q_T^n > nq)\right)$, for given $p, q > 0$; popularly speaking, the more independent $\{Q_0^n > np\}$ and $\{Q_T^n > nq\}$ are, the smaller the distance. More specifically, we characterize for fBm and iOU how our metric decays to 0 when T grows to infinity.

In our analysis, we specialize to the important cases of fBm and iOU input. We observe that the correlation structure of the input process essentially carries over to the workload process.

The organization of this paper is as follows. In Section 2 we recall some results about Gaussian processes and the large deviations theorems that we need. In Section 3 we give the main results of this paper. Section 4 is devoted to the proofs of the results. In the last section we give a heuristic approach that extends our results to queues fed by more general Gaussian processes, and a number of concluding remarks.

2 Preliminaries

This section consists of two subsections. In Section 2.1 we recall basic properties of Gaussian processes, while in Section 2.2 we state two important tools from large-deviations theory: (the multivariate version of) Cramér's theorem and (the generalized version of) Schilder's theorem.

2.1 Gaussian processes

Let $A_i(\cdot)$ denote a sequence of i.i.d. centered Gaussian processes with continuous sample paths and stationary increments, $i = 1, \ldots, n$; it is assumed that $A_i(0) \equiv 0$ for all i. For $s < t$, we interpret $A_i(s,t) := A_i(t) - A_i(s)$ as the amount of the traffic generated by the i-th Gaussian source in the time interval $(s,t]$. We denote by $A(t)$ the generic Gaussian process corresponding to a single source, and $A(s,t) := A(t) - A(s)$. A (centered) Gaussian process is characterized by its variance function $v(\cdot)$ (which is necessarily continuous); because of the stationarity of the increments of our process, we have $\mathbb{V}\text{ar}A(s,t) = v(t-s)$ for $s < t$.
In the sequel we frequently work with the bivariate random variable $(A(-s,0), A(T-t,T))$ (for large values of T). Its distribution is a bivariate Normal distribution with zero mean vector and covariance matrix $\Sigma(s,t)$ given by

$$\Sigma(s,t) := \begin{pmatrix} v(s) & \Gamma_T(s,t) \\ \Gamma_T(s,t) & v(t) \end{pmatrix},$$

with

$$\Gamma_T(s,t) := \mathbb{C}\text{ov}(A(-s,0), A(T-t,T)) = \frac{v(T+s) - v(T) - v(T-t+s) + v(T-t)}{2}.$$

Gaussian sources have the intrinsic inconvenience that in principle negative traffic can be generated: $A(s,t)$ (with $t > s$) is not necessarily non-negative. When using the representation for the workload at time t (take for ease a queue fed by a single Gaussian source, with service rate $c > 0$)

$$Q_t := \sup_{s \geq 0} \left\{ A(t-s,t) - cs \right\},$$

this turns out to be not an issue: the probabilistic properties of the above functional of the Gaussian process $A(\cdot)$ can be evaluated, irrespective of whether the input process allows negative increments. In our study we focus, without loss of generality, on centered Gaussian processes, but it is straightforward to adapt the results to the case of non-centered Gaussian processes.
In this paper we focus on two special Gaussian processes: (standard) fractional Brownian motion (or *fBm*; $v(t) = t^{2H}$, with $H \in (0,1)$), and integrated Ornstein-Uhlenbeck (or *iOU*; $v(t) = t - 1 + e^{-t}$).

Lemma 2.1. *Fix $s, t > 0$.*

- fBm. *For $H > \frac{1}{2}$ ($H < \frac{1}{2}$, respectively), it holds that for T larger than a threshold $T^\star$, $\Gamma_T(s,t)$ is positive (negative). In addition, $\lim_{T\to\infty} \Gamma_T(s,t) = 0$.*
- iOU. *$\Gamma_T(s,t)$ is positive, and decreases to 0 when $T \to \infty$.*

Proof. First focus on fBm. It is immediate that

$$\Gamma_T(s,t) = \gamma_T^{(\mathrm{fBm})}(s,t) := \frac{1}{2}\left((T+s)^{2H} - T^{2H} - (T-t+s)^{2H} + (T-t)^{2H}\right).$$

It is easy to verify that, with $v(t) = t^{2H}$,

$$\lim_{T\to\infty} \frac{\gamma_T^{(\mathrm{fBm})}(s,t)}{T^{2H-2}} = \lim_{h\downarrow 0} \frac{v(1+sh) - v(1) - v(1-(t-s)h) + v(1-th)}{2h^2}.$$

Applying L'Hôpital's rule twice yields $\frac{1}{2}st \cdot (2H)(2H-1)$. This implies what was stated on fBm. For iOU,

$$\Gamma_T(s,t) = \gamma_T^{(\mathrm{iOU})}(s,t) := \frac{1}{2}\left(e^{-T-s} - e^{-T} - e^{-T+t-s} + e^{-T+t}\right) = \frac{1}{2}(1-e^{-s})(e^t-1)e^{-T},$$

which is indeed positive and decreasing in T. □

2.2 Large deviations results

In this subsection we give a brief description of the main results from the large-deviations theory for Gaussian processes. The proofs of the theorems presented here can be found in [7, 8]; see for more background [13]. We first state Cramér's theorem, that relates to d-dimensional random variables, and then Schilder's theorem, that describes the sample-path large deviations of Gaussian processes.

Let $X \in \mathbb{R}^d$ be a d-dimensional random vector. We denote the moment generating function of X by $M(\theta) := \mathbb{E}\big(\exp(\langle\theta, X\rangle)\big)$ and its logarithm by $\Lambda(\theta) := \log M(\theta)$. Its convex conjugate $\Lambda^\star$ is defined by $\Lambda^\star(x) := \sup_{\theta\in\mathbb{R}^d}\big(\langle\theta, x\rangle - \Lambda(\theta)\big)$, with $\langle\cdot,\cdot\rangle$ denoting the usual inner product: $\langle a, b\rangle := a^{\mathrm{T}}b = \sum_{i=1}^d a_i b_i$. We first state (the multivariate version of) Cramér's theorem which characterizes the logarithmic rate of the convergence of the empirical mean of i.i.d. random vectors in $\mathbb{R}^d$.

Theorem 2.2 (Multivariate Cramér). *Let $X_i \in \mathbb{R}^d$ be i.i.d. d-dimensional random vectors, distributed as a random vector X. Then the following ldp applies:*

(a) *For any closed set* $F \subset \mathbb{R}^d$*,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i \in F\Big) \le -\inf_{x\in F} \Lambda^\star(x);$$

(b) *For any open set* $G \subset \mathbb{R}^d$*,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i \in G\Big) \ge -\inf_{x\in G} \Lambda^\star(x),$$

where the large deviations rate function $\Lambda^\star(\cdot)$ *is as given above.*

Remark 2.3. Consider the case that X has a multivariate Normal distribution with mean vector 0 and $d \times d$-non-singular matrix covariance matrix Σ. Then using $\Lambda(\theta) = \frac{1}{2}\theta^T\Sigma\theta$ we obtain

$$\theta^\star = \Sigma^{-1}x \quad \text{and} \quad \Lambda^\star(x) = \frac{1}{2}x^T\Sigma^{-1}x, \tag{2}$$

where θ^* is the optimizer in the definition of $\Lambda^\star$. ♠

Before stating the generalized Schilder's theorem we first sketch the framework of the Schilder's sample path large deviations principal as established in [3] for the Brownian motion, see also [8]. We use the same setup and notation as in [13, 14]. We consider n i.i.d. centered Gaussian processes $A_i(\cdot)$ and define the path space Ω as

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

which becomes a Banach space by equipping it with the norm

$$||\omega||_\Omega := \sup_{t\in\mathbb{R}} \frac{|\omega(t)|}{1+|t|}.$$

In Addie *et al.* [1] it is shown that $A(\cdot)$ can be realized in Ω under Assumption 2.4; it is clear that both fBm and iOU satisfy this requirement.

Assumption 2.4. *The variance function* $v(\cdot)$ *of the process* $A(\cdot)$ *is continuous and it satisfies*

$$\lim_{t\to\infty} \frac{v(t)}{t^\alpha} = 0 \tag{3}$$

for some $\alpha \in (0, 2)$.

Next we introduce the *reproducing kernel Hilbert space* $\mathrm{R} \subset \Omega$, with the property that its elements are roughly as smooth as the covariance function $\Gamma(s, \cdot)$, see Adler [2] for more details. We start from a subspace $\mathrm{R}^\star \subset \Omega$, defined by

$$\mathrm{R}^\star := \left\{ \omega \in \Omega,\ \omega(\cdot) = \sum_{i=1}^{n} a_i \Gamma(s_i, \cdot),\ a_i, s_i \in \mathbb{R},\ n \in \mathbb{N} \right\}$$

The inner product on this space $\mathrm{R}^\star$ is defined as follows, for $\omega_a, \omega_b \in \mathrm{R}^\star$

$$\langle \omega_a, \omega_b \rangle_{\mathrm{R}} := \left\langle \sum_{i=1}^{n} a_i \Gamma(s_i, \cdot), \sum_{j=1}^{n} b_j \Gamma(s_j, \cdot) \right\rangle_{\mathrm{R}} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j \Gamma(s_i, s_j). \tag{4}$$

Now we can introduce the norm $||\omega||_{\mathrm{R}} := \sqrt{\langle \omega, \omega \rangle_{\mathrm{R}}}$. The closure of $\mathrm{R}^\star$ under this norm is defined as the space R. Now we can define the rate function of the sample-path large-deviations principle (ldp):

$$I(\omega) := \begin{cases} \frac{1}{2}||\omega||_{\mathrm{R}}^2 & \text{if } \omega \in \mathrm{R}; \\ \infty & \text{otherwise.} \end{cases} \tag{5}$$

For Gaussian processes the following sample-path ldp holds.

Theorem 2.5 (Generalized Schilder). *The following sample-path ldp applies:*

(a) For any closed set $F \subset \Omega$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n} \sum_{i=1}^{n} A_i(\cdot) \in F \Big) \leq - \inf_{\omega \in F} I(\omega);$$

(b) For any open set $G \subset \Omega$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n} \sum_{i=1}^{n} A_i(\cdot) \in G \Big) \geq - \inf_{\omega \in G} I(\omega),$$

3 Main results

As mentioned in the introduction of this paper, our main interest lies in the characterization of the correlation structure of the workload. Since only for the case of Brownian motion input the workload distribution has been found explicitly, we resort to an asymptotic framework, viz. the so-called many-sources regime. In this regime the number of Gaussian inputs, say n, grows large, and the service rate is scaled accordingly. In this framework, the stationary workload process is given by

$$Q_t^n := \sup_{s\leq t} \sum_{i=1}^n A_i(s,t) - n(t-s) = \sup_{s\geq 0} \sum_{i=1}^n A_i(t-s,t) - ncs. \tag{6}$$

As we wish to investigate the correlation structure of the workload process, we could try to characterize the autocorrelation

$$\delta_n(T) := \frac{\mathbb{E}Q_0^n Q_T^n - (\mathbb{E}Q_0^n)(\mathbb{E}Q_T^n)}{\sqrt{\mathbb{V}\mathrm{ar}(Q_0^n)}\sqrt{\mathbb{V}\mathrm{ar}(Q_T^n)}} = \frac{\mathbb{E}Q_0^n Q_T^n - (\mathbb{E}Q_0^n)(\mathbb{E}Q_0^n)}{\mathbb{V}\mathrm{ar}(Q_0^n)}.$$

It is evident that $\delta_n(T) \downarrow 0$ as $T \uparrow \infty$, but the question is how fast it vanishes.
Unfortunately, this notion of autocorrelation is hard to handle — not even an explicit expression for $\mathbb{E}Q_0^n$ is known for non-Brownian Gaussian input processes. We therefore introduce an alternative notion of correlation. The following metric describes the degree of correlation between the events $\{Q_0^n > np\}$ and $\{Q_T^n > nq\}$ for positive p, q.

Definition 3.1. For given positive numbers p, q define

$$\kappa_n(T) := \frac{\mathbb{P}\left(Q_0^n > np, Q_T^n > nq\right)}{\mathbb{P}\left(Q_0^n > np\right)\mathbb{P}\left(Q_T^n > nq\right)}. \tag{7}$$

Furthermore, let $\kappa(T)$ be the limit of $\log \kappa_n(T)/n$ as $n \to \infty$ if it exists.

Before stating the main theorems of this section we first give the logarithmic asymptotics of the marginal probabilities involved in Definition 3.1. They are given by

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(Q_0^n > np\big) \quad = \quad -\inf_{s>0} \frac{(p+cs)^2}{2v(s)} \tag{8}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(Q_T^n > nq\big) \quad = \quad -\inf_{t>0} \frac{(q+ct)^2}{2v(t)}; \tag{9}$$

see for instance [1]. In [6] the following lemma was proved; it entails that the infima over s and t are attained and are unique under a specific assumption on the variance function.

Lemma 3.2. *Suppose that the standard deviation function $\sigma(t) := \sqrt{v(t)}$ of the generic input process $A(\cdot)$ is such that $\sigma(t) \in \mathcal{C}^2([0,\infty))$ is strictly increasing and strictly concave. Then the right-hand sides of* (8) *and* (9) *have unique minimizers. Concavity of $\sigma(t)$ is equivalent to requiring that*

$$2v(t)v''(t) - (v'(t))^2 \leq 0. \tag{10}$$

We denote the minimizers by $s^\star$ and $t^\star$. It is readily checked that they solve

$$\begin{cases} 2cv(s) = (p+cs)v'(s); \\ 2cv(t) = (q+ct)v'(t). \end{cases} \tag{11}$$

Now we give the main results of this paper. Theorem 3.3 states that for fBm input $\kappa(T)$ decays to zero and its decay rate is T^{2H-2} as $T \to \infty$, which indicates that the workload process has essentially the same correlation structure as the input process. As will be discussed in more detail in Section 5, this means that the workload process is (in our metric) long-range dependent if the *Hurst parameter* H is greater than $\frac{1}{2}$. For fBm, $\sigma(t) = t^H$ is concave, so Lemma 3.2 applies, and (11) has a unique solution; in fact, $s^\star$ and $t^\star$ can be explicitly calculated, and are given through

$$s^\star := \frac{p}{c}\frac{H}{1-H};\;\; t^\star := \frac{q}{c}\frac{H}{1-H}. \tag{12}$$

Theorem 3.3 (fBm input). *If the input process is fBm we have the following logarithmic asymptotics for* $\kappa_n(T)$*:*

$$\begin{aligned} \kappa(T) &= \lim_{n\to\infty}\frac{1}{n}\log\kappa_n(T) = \frac{(p+cs^\star)(q+ct^\star)}{v(s^\star)v(t^\star)}\cdot\frac{1}{2}s^\star t^\star(2H)(2H-1)T^{2H-2} + o(T^{2H-2}) \\ &= \frac{(2H-1)c^2}{H}s^{\star 2-2H}t^{\star 2-2H}T^{2H-2} + o(T^{2H-2}). \end{aligned} \tag{13}$$

Now consider the case of iOU input. In this case $s^\star$ and $t^\star$ cannot be explicitly calculated. They are uniquely determined though, as can be seen as follows. Criterion (10) reduces to

$$\varphi(t) := 2te^{-t} + e^{-2t} - 1 \le 0,$$

which is true because $\varphi(0) = 0$ and $\varphi'(t) = e^{-t}(2-2t-2e^{-t}) \le 0$.
Theorem 3.4 states that for iOU input the speed of convergence of $\kappa(T)$ to 0 as $T \to \infty$ is e^{-T} (just like the correlation of the input process) indicating that the workload process is short-range dependent as well.

Theorem 3.4 (iOU input). *If the input process is iOU we have the following logarithmic asymptotics for* $\kappa_n(T)$*:*

$$\begin{aligned} \kappa(T) &= \lim_{n\to\infty}\frac{1}{n}\log\kappa_n(T) = \frac{(p+cs^\star)(q+ct^\star)}{v(s^\star)v(t^\star)}\cdot\frac{1}{2}(1-e^{-s^\star})(e^{t^\star}-1)e^{-T} + o(e^{-T}) \\ &= 2c^2e^{-(T-t^\star)} + o(e^{-T}). \end{aligned} \tag{14}$$

4 Proofs

In this section we give the proofs of the results we stated in the previous section. In the first subsection we derive a number of generic results, while we specialize to fBm and iOU in the last part of the section.

4.1 General results

The results of this subsection hold for any type of Gaussian sources (i.e., we do not restrict ourselves to fBm and iOU), the only exception being Proposition 4.4. We first define two sets of paths in Ω that play a crucial role in our analysis.

$$\mathcal{S}_T \;:=\; \{f \in \Omega : \exists s > 0, \exists t > 0 : -f(-s) > p + cs,\ f(T) - f(T-t) > q + ct\}\,; \tag{15}$$

$$\mathcal{S}_T(s,t) \;:=\; \{f \in \Omega : -f(-s) > p + cs,\ f(T) - f(T-t) > q + ct\}\,; \tag{16}$$

Observe that $\mathcal{S}_T$ is the union (over all $s, t > 0$) of the $\mathcal{S}_T(s,t)$. Interestingly, the set of paths $\mathcal{S}_T$ directly relates to the 'joint overflow event' $\{Q_0^n > np, Q_T^n > nq\}$, as follows from the next lemma.

Lemma 4.1. *For any $p, q > 0$,*

$$\mathbb{P}\left(Q_0^n > np, Q_T^n > nq\right) = \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n A_i(\cdot) \in \mathcal{S}_T\right).$$

Proof. This follows by applying (6):

$$\begin{aligned}
&\mathbb{P}\left(Q_0^n > np, Q_T^n > nq\right) = \\
&= \mathbb{P}\left(\sup_{s>0}\left\{\sum_{i=1}^n A_i(-s,0) - ncs\right\} > np,\ \sup_{t>0}\left\{\sum_{i=1}^n A_i(T-t,T) - nct\right\} > nq\right) \\
&= \mathbb{P}\left(\exists s > 0 : \sum_{i=1}^n A_i(-s,0) - ncs > np, \exists t > 0 : \sum_{i=1}^n A_i(T-t,T) - nct > nq\right) \\
&= \mathbb{P}\left(\exists s > 0 : \sum_{i=1}^n \frac{A_i(-s,0)}{n} > p + cs, \exists t > 0 : \sum_{i=1}^n \frac{A_i(T-t,T)}{n} > q + ct\right) \\
&= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n A_i(\cdot) \in \mathcal{S}_T\right),
\end{aligned}$$

which proves the claimed. □

In the sequel we frequently use the following bivariate Normal large-deviations rate function:

$$\Lambda^\star_T(p+cs,q+ct) := \frac{1}{2}(p+cs,q+ct)\,(\Sigma_T(s,t))^{-1}\begin{pmatrix} p+cs \\ q+ct \end{pmatrix}.$$

By explicitly calculating the matrix inverse, we obtain that $\Lambda^\star_T(p+cs,q+ct)$ can be written in the following alternative form:

$$\frac{1}{2}\frac{v(s)v(t)}{v(s)v(t)-\Gamma_T(s,t)^2}\left(\frac{(p+cs)^2}{v(s)}+\frac{(q+ct)^2}{v(t)}-2\frac{(p+cs)(q+ct)\Gamma_T(s,t)}{v(s)v(t)}\right). \tag{17}$$

The next lemma determines the decay rate of the most likely path in $\mathcal{S}_T(s,t)$, for fixed values of s and t. It turns out that there are three different regimes.

Lemma 4.2. *For any $p,q>0$,*

$$\inf_{f\in\mathcal{S}_T(s,t)} I(f) = \bar\Lambda^\star_T(p+cs,q+ct),$$

where $\bar\Lambda^\star_T(p+cs,q+ct)$ equals

$$\frac{(p+cs)^2}{2v(s)} \quad \textit{if} \quad \frac{\Gamma_T(s,t)}{v(s)}(p+cs) > q+ct; \tag{18}$$

$$\frac{(q+ct)^2}{2v(t)} \quad \textit{if} \quad \frac{\Gamma_T(s,t)}{v(t)}(q+ct) > p+cs; \tag{19}$$

$$\Lambda^\star_T(p+cs,q+ct) \;\; \textit{otherwise}.$$

Proof. First observe that Cauchy-Schwarz implies that the conditions in (18) and (19) cannot apply simultaneously. Recognize

$$\begin{aligned}\frac{\Gamma_T(s,t)}{v(s)}(p+cs) &= \mathbb{E}(A(T-t,T)\mid A(-s,0)=p+cs);\\ \frac{\Gamma_T(s,t)}{v(t)}(q+ct) &= \mathbb{E}(A(-s,0)\mid A(T-t,T)=q+ct).\end{aligned}$$

The stated now follows from the bivariate version of Cramér's theorem analogously to [13, Exercise 4.1.9]. □

The proof of the next proposition relies on Lemma A.1, that is stated and proven in the appendix.

Proposition 4.3. *For any $p, q > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Q_0^n > np, Q_T^n > nq) = -\inf_{f\in\mathcal{S}_T} I(f) = -\inf_{s,t>0} \bar{\Lambda}_T^\star(p+cs, q+ct).$$

Proof. From 'Schilder' and Lemma 4.1 we have

$$-\inf_{f\in\mathcal{S}_T} I(f) \le \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Q_0^n > np, Q_T^n > nq) \le -\inf_{f\in\overline{\mathcal{S}_T}} I(f).$$

We first show that the above inequalities are actually equalities, by establishing that $\mathcal{S}_T$ is an I-continuity set, that is,

$$\inf_{f\in\mathcal{S}_T} I(f) = \inf_{f\in\overline{\mathcal{S}_T}} I(f), \tag{20}$$

where the $\overline{\mathcal{S}_T}$ denotes the closure of $\mathcal{S}_T$, and is given in Lemma A.1.
This can be done in the same way as in Appendix of [15]. Choose an arbitrary path f in $\overline{\mathcal{S}_T}\cap\mathrm{R}$, and approximate it by a path in $\mathcal{S}_T$, as follows. We use the sets $\mathcal{S}(s), \mathcal{S}_T(t), \overline{\mathcal{S}(s)}$, and $\overline{\mathcal{S}_T(t)}$ as defined in the appendix. Due to Lemma A.1 we have that $f \in \overline{\mathcal{S}(s)} \cap \overline{\mathcal{S}_T(t)}$ for some $s, t > 0$. Let $\eta(\cdot)$ be a path in R that is strictly increasing and taking negative values for $u \in (-\infty, 0)$ and positive values for $u \in (0, \infty)$ (for instance $\eta(u) :=$ sgn$(u)\sqrt{|u|}$ or $\arctan u$). Define

$$f_n(u) := f(u) + \frac{\eta(u)}{n}.$$

Then $f_n \in \mathcal{S}(s) \cap \mathcal{S}_T(t)$ as, for any $s > 0$, it holds that

$$-f_n(-s) = -f(-s) - \frac{\eta(-s)}{n} \ge p + cs - \frac{\eta(-s)}{n} > p + cs$$

and, for any $t > 0$,

$$\begin{aligned} f_n(T) - f_n(T-t) &= f(T) - f(T-t) + \frac{\eta(T) - \eta(T-t)}{n} \\ &\ge q + ct + \frac{\eta(T) - \eta(T-t)}{n} > q + ct. \end{aligned}$$

Moreover, we have, for $n \to \infty$,

$$\|f_n\|_{\mathrm{R}}^2 = \|f + \frac{1}{n}\eta\|_{\mathrm{R}}^2 \to \|f\|_{\mathrm{R}}^2,$$

which proves (20) and therefore also the first equality of the proposition.

From the above we conclude that the decay rate of our interest equals

$$\inf_{s,t>0} \inf_{f\in(\mathcal{S}(s)\cap\mathcal{S}_T(t))} I(f).$$

Recall from (15) and (16) that $\mathcal{S}_T$ is the union over all $s \geq 0$ and $t \geq 0$ of the $\mathcal{S}_T(s,t)$, and observe that $\mathcal{S}_T(s,t) = \mathcal{S}(s) \cap \mathcal{S}_T(t)$. The second equality of the proposition now follows directly from Lemma 4.2. □

Proposition 4.4. *Consider fBm or iOU. For any $p, q > 0$, and T large enough*

$$\inf_{s,t>0} \bar{\Lambda}^\star_T(p+cs, q+ct) = \inf_{s,t>0} \Lambda^\star_T(p+cs, q+ct). \tag{21}$$

Proof. Lemma 2.1 states that, for any fixed s, t, $\Gamma_T(s,t) \to 0$ as $T \to \infty$. It can be checked that this implies that also $\bar{\Lambda}^\star_T(p+cs, q+ct) \to \Lambda^\star_\infty(p+cs, q+ct)$ as $T \to \infty$, where

$$\Lambda^\star_\infty(p+cs, q+ct) = \frac{(p+cs)^2}{v(s)} + \frac{(q+ct)^2}{v(t)}$$

(to this end, observe that, for any fixed s, t the conditions in (18) and (19) are not fulfilled for T sufficiently large). It is clear that, when taking the infimum of $\Lambda^\star_\infty(p+cs, q+ct)$ over $s, t > 0$, the expression decouples into the sum of an infimum over s and an infimum over t. Both individual infima have a unique minimizer, namely $s^\star$ and $t^\star$ as introduced earlier.
It follows, relying on the continuity properties of the functions involved, that a sequence of local optimizers of the left-hand side of (21), say $\bar{s}^\star_T$ and $\bar{t}^\star_T$, converge to $s^\star$ and $t^\star$ as $T \to \infty$. This means that we can restrict ourselves, for an appropriately chosen $M < \infty$ and T large enough, to optimizing over $s, t \in (0, M)$. For any $s, t \in (0, M)$, and for T large enough, the conditions in (18) and (19) are not satisfied (by virtue of Lemma 2.1), and hence $\bar{\Lambda}^\star_T(p+cs, q+ct) = \Lambda^\star_T(p+cs, q+ct)$. This implies the stated. □

It also holds that $\Lambda^\star_T(p+cs, q+ct) \to \Lambda^\star_\infty(p+cs, q+ct)$. Then a sequence of local optimizers of the right-hand side of (21), say $s^\star_T$ and $t^\star_T$, converge to $s^\star$ and $t^\star$ as $T \to \infty$. The vector $s^\star_T, t^\star_T$ at which the function $\Lambda^\star_T(p+cs, q+ct)$ is minimum in the neighborhood of $s^\star, t^\star$ is solution of the following system:

$$\begin{aligned}
&(p+cs)\left(2cv(s) - (p+cs)v'(s)\right) \\
&= 2\left(\frac{q+ct}{v(t)}\right)\left(\left(cv(s) - (p+cs)v'(s)\right)\Gamma_T(s,t) + (p+cs)v(s)\frac{\partial \Gamma_T}{\partial s}(s,t)\right);
\end{aligned} \tag{22}$$

$$
\begin{aligned}
&(q+ct)\left(2cv(t)-(q+ct)v'(t)\right) &(23)\\
&= \quad 2\left(\frac{p+cs}{v(s)}\right)\left(\left(cv(t)-(q+ct)v'(t)\right)\Gamma_T(s,t)+(q+ct)v(t)\frac{\partial\Gamma_T}{\partial t}(s,t)\right)
\end{aligned}
$$

where the partial derivatives of $\Gamma_T(s,t)$ with respect to s and t are given by

$$
\frac{\partial\Gamma_T}{\partial s}(s,t)=\frac{1}{2}\left(v'(T+s)-v'(T-t+s)\right);\quad \frac{\partial\Gamma_T}{\partial t}(s,t)=\frac{1}{2}\left(v'(T-t+s)-v'(T-t)\right).
$$

In the next two subsections we study the system (22)-(23), for both fBm and iOU , by analyzing the behavior of $s_T^\star, t_T^\star$ in detail. This yields the desired information, needed in order to characterize the decay rate $\kappa(T)$ for T large.

4.2 Proof for fBm input

As we have seen in the proof of Lemma 2.1, for $T\to\infty$,

$$
\gamma_T^{(\mathrm{fBm})}(s,t)=st\cdot H(2H-1)\cdot T^{2H-2}+o(T^{2H-2}).
$$

For large T we obtain in the same way

$$
\frac{\partial\gamma_T^{(\mathrm{fBm})}}{\partial s}=t\cdot H(2H-1)\cdot T^{2H-2}+o(T^{2H-2});\quad \frac{\partial\gamma_T^{(\mathrm{fBm})}}{\partial t}=s\cdot H(2H-1)\cdot T^{2H-2}+o(T^{2H-2}).
$$

Inserting these into (22)-(23) we obtain

$$
(2cs-2H(p+cs))=\frac{2H(2H-1)(q+ct)(cs-(2H-1)(p+cs))st}{t^{2H}(p+cs)}T^{2H-2}+o(T^{2H-2}); \quad (24)
$$

$$
(2ct-2H(q+ct))=\frac{2H(2H-1)(p+cs)(ct-(2H-1)(q+ct))st}{s^{2H}(q+ct)}T^{2H-2}+o(T^{2H-2}). \quad (25)
$$

Note that if we let $T\to\infty$ in the last system, we retrieve (11), which has a unique solution (12). Observe that in the system of equations (24)-(25), the right-hand side of the equations decays to 0 with speed T^{2H-2} as T grows to infinity. This observation, in conjunction with the fact that $s_T^\star, t_T^\star$ converges to $s^\star, t^\star$, entails that we can express $s_T^\star, t_T^\star$ as follows:

$$
\begin{cases}
s_T^\star=s^\star+f(s^\star,t^\star)T^{2H-2}+o(T^{2H-2});\\
t_T^\star=t^\star+g(s^\star,t^\star)T^{2H-2}+o(T^{2H-2}).
\end{cases}
$$

To determine the values of f and g at $s^\star, t^\star$ we proceed as follows. Using Taylor expansions we obtain for the left-hand side of (24), after tedious calculus,

$$\begin{aligned}&(p+cs)\big(2cv(s)-(p+cs)v'(s)\big)\\ &\quad= \; 2H(p+cs^\star)s^{\star 2H-2}\left(cs^\star-(2H-1)(p+cs^\star)\right)f(s^\star,t^\star)T^{2H-2}+o(T^{2H-2}),\end{aligned}$$

and for the right-hand side

$$\begin{aligned}&2\left(\frac{q+ct}{v(t)}\right)\left(\big(cv(s)-(p+cs)v'(s)\big)\,\Gamma_T(s,t)+(p+cs)v(s)\frac{\partial\Gamma_T}{\partial s}(s,t)\right)\\ &\quad= \; 2H(2H-1)(q+ct^\star)s^{\star 2H}t^{\star 1-2H}\left(cs^\star-(2H-1)(p+cs^\star)\right)T^{2H-2}+o(T^{2H-2})\end{aligned}$$

Doing the same for (25), and inserting (12), we find the following expressions for f and g at $s^\star, t^\star$:

$$f(s^\star,t^\star)=(2H-1)\frac{q}{p}s^{\star 2}t^{\star 1-2H}=(2H-1)s^\star t^{\star 2-2H};$$

$$g(s^\star,t^\star)=(2H-1)\frac{p}{q}t^{\star 2}s^{\star 1-2H}=(2H-1)t^\star s^{\star 2-2H}.$$

Inserting these expressions into $\Lambda^\star_T(p+cs,q+ct)$, we can evaluate the components of (17):

$$\begin{aligned}\frac{(p+cs)^2}{s^{2H}} &= \frac{(p+cs^\star)^2}{s^{\star 2H}}\left(1+2\left(\frac{c}{(p+cs^\star)}-\frac{H}{s^\star}\right)f(s^\star,t^\star)T^{2H-2}+o(T^{2H-2})\right)\\ &= \frac{(p+cs^\star)^2}{s^{\star 2H}}+o(T^{2H-2});\end{aligned}$$

$$\begin{aligned}\frac{(q+ct)^2}{t^{2H}} &= \frac{(q+ct^\star)^2}{t^{\star 2H}}\left(1+2\left(\frac{c}{(q+ct^\star)}-\frac{H}{t^\star}\right)g(s^\star,t^\star)T^{2H-2}+o(T^{2H-2})\right)\\ &= \frac{(q+ct^\star)^2}{t^{\star 2H}}+o(T^{2H-2});\end{aligned}$$

$$\begin{aligned}&2H(2H-1)\cdot(p+cs)(q+ct)(st)^{1-2H}\cdot T^{2H-2}+o(T^{2H-2})=\\ &\qquad= 2H(2H-1)(p+cs^\star)(q+ct^\star)s^{\star 1-2H}t^{\star 1-2H}T^{2H-2}+o(T^{2H-2})\\ &\qquad= 2\frac{(2H-1)c^2}{H}s^{\star 2-2H}t^{\star t-2H}T^{2H-2}+o(T^{2H-2}).\end{aligned}$$

We thus obtain the desired result, i.e.,

$$\kappa(T)=\lim_{n\to\infty}\frac{1}{n}\log\kappa_n(T)=\frac{(2H-1)c^2}{H}s^{\star 2-2H}t^{\star 2-2H}T^{2H-2}+o(T^{2H-2}).$$

4.3 Proof for iOU input

As in the fBm case, denote by $s^\star, t^\star$ the minimizing point when there is no correlation, i.e., the solution of (11). We follow the same arguments as in the case of fBm. For fixed (s,t) the covariance $\Gamma_T(s,t)$ is decreasing exponentially in T. The solution of system (22)-(23), say $s^\star_T, t^\star_T$, converges to $s^\star, t^\star$, and its convergence speed is of the order e^{-T} for large T. These observations entail that

$$\begin{cases} s^\star_T = s^\star + k(s^\star,t^\star)e^{-T} + o(T^{2H-2}); \\ t^\star_T = t^\star + \ell(s^\star,t^\star)e^{-T} + o(T^{2H-2}). \end{cases}$$

To determine k and ℓ at $s^\star, t^\star$, we proceed as in the above subsection. We find

$$\begin{aligned} k(s^\star,t^\star) &= \frac{q+ct^\star}{p+cs^\star}(e^{t^\star}-1)\frac{cv(s^\star)(1-e^{-s^\star}) - (p+cs^\star)\left(v'(s^\star)(1-e^{-s^\star}) - v(s^\star)e^{-s^\star}\right)}{\left(cv'(s^\star) - (p+cs^\star)v''(s^\star)\right)v(t^\star)} \\ &= \frac{(q+ct^\star)v'(t^\star)}{v''(t^\star)(p+cs^\star)v(t^\star)} \cdot \frac{cv(s^\star)v'(s^\star) - (p+cs^\star)v'(s^\star)^2 + (p+cs^\star)v(s^\star)v''(s^\star)}{(cv'(s^\star) - (p+cs^\star)v''(s^\star))} \\ &= \frac{2cv(s^\star)}{v''(t^\star)(p+cs^\star)} \cdot \frac{(-cv'(s^\star) + (p+cs^\star)v''(s^\star))}{(cv'(s^\star) - (p+cs^\star)v''(s^\star))} = -\frac{v'(s^\star)}{v''(t^\star)}; \end{aligned}$$

$$\begin{aligned} \ell(s^\star,t^\star) &= \frac{p+cs^\star}{q+ct^\star}(1-e^{-s^\star})\frac{cv(t^\star)(e^{t^\star}-1) - (q+ct^\star)\left(v'(t^\star)(e^{t^\star}-1) - v(t^\star)e^{t^\star}\right)}{\left(cv'(t^\star) - (q+ct^\star)v''(t^\star)\right)v(s^\star)} \\ &= \frac{(p+cs^\star)v'(s^\star)}{(q+ct^\star)v(s^\star)} \cdot \frac{cv(t^\star)v'(t^\star)e^{t^\star} - (q+ct^\star)v'(t^\star)^2e^{t^\star} + (q+ct^\star)v(t^\star)e^{t^\star}}{(cv'(t^\star) - (q+ct^\star)v''(t^\star))} \\ &= \frac{2cv(s^\star)}{v''(t^\star)(q+ct^\star)} \cdot \frac{(-cv'(t^\star) + (q+ct^\star))}{(cv'(t^\star) - (p+ct^\star)v''(t^\star))} = \frac{v'(t^\star)}{v''(t^\star)} \cdot \frac{q+cv(t^\star)}{(cv'(t^\star) - (q+ct^\star)v''(t^\star))} \end{aligned}$$

Now we insert this in the objective function (17), and similarly to the fBm case we obtain

$$\frac{(p+cs)^2}{v(s)} = \frac{(p+cs^\star)^2}{v(s^\star)}\left(1+o(e^{-T})\right);$$

$$\frac{(q+ct)^2}{v(t)} = \frac{(q+ct^\star)^2}{v(t^\star)}\left(1+o(e^{-T})\right);$$

$$\frac{(p+cs)(q+ct)(1-e^{-s})(e^t-1)e^{-T}}{v(s)v(t)} = \frac{(p+cs^\star)(q+ct^\star)(1-e^{-s^\star})(e^{t^\star}-1)e^{-T}}{v(s^\star)v(t^\star)} + o(e^{-T}).$$

Thus we get for iOU input the desired result:

$$\kappa(T) = \lim_{n\to\infty}\frac{1}{n}\log\kappa_n(T) = \frac{(p+cs^\star)(q+ct^\star)}{v(s^\star)v(t^\star)} \cdot \frac{1}{2}(1-e^{-s^\star})(e^{t^\star}-1)e^{-T} + o(e^{-T}),$$

which simplifies to $2c^2e^{-(T-t^\star)}$.

5 Discussion and concluding remarks

A. Generalizations. Theorems 3.3 and 3.4 suggest that our results can be generalized considerably. First observe that expression (17) can alternatively be written as

$$\frac{1}{2}\left(\frac{(p+cs)^2}{v(s)}+\frac{(q+ct)^2}{v(t)}-2\frac{(p+cs)(q+ct)\Gamma_T(s,t)}{v(s)v(t)}\right)+o(\Gamma_T(s,t)). \tag{26}$$

Consider the following lemma.

Lemma 5.1. *Let $f(t)$ and $g(t)\in\mathcal{C}^2([0,\infty))$, and let $f(\cdot)$ have a unique minimizer $t^\star$. Then*

$$\inf_t(f(t)+\varepsilon g(t))-f(t^\star)=\varepsilon g(t^\star).$$

Proof. The Taylor expansion of $f'(t)+\varepsilon g'(t)$ in $t^\star_\varepsilon=t^\star+\varepsilon\bar{t}$ reads

$$f'(t^\star)+\varepsilon\bar{t}f''(t^\star)+\varepsilon g'(t^\star)+O(\varepsilon^2)=f'(t^\star)+\varepsilon\left(\bar{t}f''(t^\star)+g'(t^\star)\right)+O(\varepsilon^2),$$

so that we obtain $\bar{t}=-g'(t^\star)/f''(t^\star)$. Hence

$$\inf_t(f(t)+\varepsilon g(t))=f(t^\star)-\varepsilon\frac{g'(t^\star)}{f''(t^\star)}f'(t^\star)+\varepsilon g(t^\star)+O(\varepsilon^2).$$

The claim follows now from $f'(t^\star)=0$. □

This claim can easily be extended to dimension 2. Now suppose that (for large T) $\Gamma_T(s,t)$ decouples as $\Gamma(s,t)\cdot\varepsilon(T)$; here $\varepsilon(T)$ does not depend on s and t, and converges to 0 as $T\to\infty$. Applying then the two-dimensional version of Lemma 5.1 to (26), we obtain that

$$\begin{aligned}\kappa(T) &= \lim_{n\to\infty}\frac{1}{n}\log\kappa_n(T)=\frac{(p+cs^\star)(q+ct^\star)}{v(s^\star)v(t^\star)}\Gamma(s^\star,t^\star)\varepsilon(T)+o(\varepsilon(T))\\ &= 4c^2\frac{\Gamma(s^\star,t^\star)}{v'(s^\star)v'(t^\star)}\varepsilon(T)+o(\varepsilon(T)),\end{aligned}$$

as long as the above mentioned factorization applies.

B. Correlation structure input vs. correlation structure queue. The above arguments show that in general one would expect that $\kappa(T)$ decays as fast as $\varepsilon(T)$, at least when the factorization $\Gamma_T(s,t)=\Gamma(s,t)\cdot\varepsilon(T)$ applies (for large T). Interestingly, for small s,t, applying Taylor expansions yields

$$\Gamma_T(s,t)=\frac{1}{2}\cdot st\cdot v''(T),$$

which indeed obeys the desired factorization. This suggests that one could expect that $\kappa(T)$ will be proportional to $v''(T)$ under rather general conditions. Noticing that, with, for $\varepsilon > 0$ small,

$$\Gamma_T(\varepsilon) := \mathbb{C}\mathrm{ov}(A(0,\varepsilon), A(T, T+\varepsilon)) \approx \frac{1}{2}\varepsilon^2 v''(T),$$

the bivariate version of Cramér's theorem implies that, for large T,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \left(\frac{\mathbb{P}\left(\sum_{i=1}^n A_i(0,\varepsilon) > np, \sum_{i=1}^n A_i(T,T+\varepsilon) > nq\right)}{\mathbb{P}\left(\sum_{i=1}^n A_i(0,\varepsilon) > np\right) \mathbb{P}\left(\sum_{i=1}^n A_i(T,T+\varepsilon) > nq\right)} \right) \\
&= -\frac{1}{2}(p,q) \begin{pmatrix} v(\varepsilon) & \Gamma_T(\varepsilon) \\ \Gamma_T(\varepsilon) & v(\varepsilon) \end{pmatrix}^{-1} \begin{pmatrix} p \\ q \end{pmatrix} + \frac{p^2}{2v(\varepsilon)} + \frac{q^2}{2v(\varepsilon)} \\
&\approx \frac{1}{2} pq \left(\frac{\varepsilon}{v(\varepsilon)} \right)^2 v''(T),
\end{aligned}$$

i.e., *also* decaying as $v''(T)$! In other words, the above arguments provide support for the claim that, in this metric, the queueing process has essentially the same correlation structure as the input process. In other words: the queue inherits the correlation of the input process.

C. Long-range dependence. First consider fBm. Heuristically reasoning, Theorem 3.3 entails that, for some constant κ_0,

$$\kappa_n(T) \approx \exp\left(n\kappa_0 T^{2H-2}\right),$$

and hence we have that the correlation coefficient of the indicator functions $1\{Q_0^n > np\}$ and $1\{Q_T^n > nq\}$ looks like

$$\frac{\mathbb{P}\left(Q_0^n > np, Q_T^n > nq\right) - \mathbb{P}\left(Q_0^n > np\right) \mathbb{P}\left(Q_T^n > nq\right)}{\mathbb{P}\left(Q_0^n > np\right) \mathbb{P}\left(Q_T^n > nq\right)} \approx e^{n\kappa_0 T^{2H-2}} - 1 \approx n\kappa_0 T^{2H-2},$$

which is non-summable for $H > \frac{1}{2}$. This intuitive argument suggests that long-range dependence of the input process causes long-range dependence of the queueing process.
Likewise, for iOU we find that the correlation coefficient introduced above is roughly proportional to e^{-T}, and hence corresponds to a short-range dependent process.

D. Remarks on asymptotics for iOU. It may be surprising, at first glance, that the asymptotics of $\kappa(T)$ for iOU, that is $2c^2 e^{-T-t^\star}$, depend on q, but *do not depend on* p. This can be understood as follows. First observe that for iOU input (unlike for fBm input) there is a notion of a *traffic rate* process $X(\cdot)$, where $X(t) = A'(t)$. It can be checked easily that (i) $X(t)$ is Normally distributed with mean

0 and variance $\frac{1}{2}$, (ii) $\mathbb{C}\text{ov}(X(0), X(T)) = \frac{1}{2}e^{-T}$, (iii) the conditional distribution of $A(T-t, T)$ given $X(0) = x$ is Normal with mean and variance, respectively,

$$\mu_T(t \mid x) = \mathbb{E}(A(T-t,T) \mid X(0) = x) = x(e^t - 1)e^{-T},$$

$$v_T(t \mid x) = \mathbb{V}\text{ar}(A(T-t,T) \mid X(0) = x) = v(t) - e^{-2T}(e^t - 1)^2,$$

as follows from standard formulae for conditional Normal distributions (cf. Section 4.3 in [14]). Also, rewrite $\kappa_n(t)$ as the ratio of $\mathbb{P}(Q_T^n > nq \mid Q_0^n > np)$ and $\mathbb{P}(Q_0^n > np)$. The decay rate of the latter probability is given by (8). Now focus on the decay rate of the former (i.e., conditional) probability. Realize that, as the condition $Q_0^n > np$ is binding, the most likely path (in the 'Schilder sense') must be such that the traffic rate at time 0 is c (which means that the aggregate input process is generating traffic at a rate nc); otherwise the queue grows even beyond np. Also notice that the most likely path is such that the buffer has been empty between 0 and T. These observations, in conjunction with the Markovian nature of iOU, entail that all the information about the system at time 0 that has impact on the system at time T, is contained in the fact that the rate is (most likely) nc at time 0. To find the decay rate of $\mathbb{P}(Q_T^n > nq \mid Q_0^n > np)$, we therefore have to solve

$$\inf_{t>0} \frac{(q + ct - \mu_T(t \mid c))^2}{2v_T(t \mid c)}.$$

The above formula for the conditional mean and variance entail that this optimization problem reduces to

$$\inf_{t>0} \left(\frac{(q+ct)^2}{2v(t)} - \frac{(q+ct)c(e^t - 1)e^{-T}}{v(t)} + o(e^{-T}) \right).$$

Applying Lemma 5.1 once again, inserting (11), and using that $v'(t) = 1 - e^{-t} = e^{-t}(e^t - 1)$, we indeed obtain that $\kappa(T)$ equals $2c^2 e^{-T-t^\star} + o(e^{-T})$, as expected.
The above reasoning explains why the decay rate does not depend on p; as an aside we mention that also $\ell(s^\star, t^\star)$ does not depend on p.

E. Further research. In this paper we have focused on the metric $\kappa(T)$ that relates to the many-sources scaling, and that was intended to express the level of correlation between the workloads at time 0 and T. Then we studied the asymptotics of $\kappa(T)$ for large T. Evidently, many other measures for correlation can be thought of. One could for instance consider similar measures, but then in the large-buffer regime.

In this respect, we could consider a queue fed by a single Gaussian input, emptied at a constant rate $C > 0$. Then an interesting measure could be, for fixed p, q, T,

$$\lambda_B := \frac{\mathbb{P}\left(Q_0 > pB, Q_{TB} > qB\right)}{\mathbb{P}\left(Q_0 > pB\right)\mathbb{P}\left(Q_{TB} > qB\right)} = \frac{\mathbb{P}\left(Q_0 > pB, Q_{TB} > qB\right)}{\mathbb{P}\left(Q_0 > pB\right)\mathbb{P}\left(Q_0 > qB\right)},$$

and its asymptotics for large B. The analysis of λ_B is radically different from that of $\kappa_n(T)$; the reason for this is that in the many-sources regime the most likely timescales to overflow are more or less constant in the scaling parameter (i.e., n), whereas in the large-buffer one would expect that these timescales are roughly proportional to the scaling parameter (i.e, B).
In this case we expect, when analyzing $\mathbb{P}(Q_0 > pB, Q_{TB} > qB)$, different regimes. More precisely: for B large it is not always true that, in the most likely scenario, both constraints are tightly met; for some values of p, q, T this will be the case, while for others just one constraint will be tightly met (and the other event 'comes for free'). In case both constraints are tightly met, again two cases can be distinguished: a first in which the queue has not become empty between 0 and TB (which we expect is the case for T smaller than some critical timescale $T^\star$), and a second in which epochs 0 and TB lie in different busy periods (for T larger than $T^\star$), cf. [20, Section 11.2].

Acknowledgments

The second author wants to thanks Peter Glynn (Stanford, US) and Ilkka Norros (VTT, Finland) for stimulating discussions.

References

[1] Addie, R., Mannersalo, P., and Norros, I. Most probable paths and performance formulae for buffers with Gaussian input traffic. *European Trans. Telecommun. 13* (2002), 183–196.

[2] Adler, R. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Lecture Notes-Monograph Series, Vol. 12.* Institute of Mathematical Statistics, Hayward, CA, USA, 1990.

[3] Bahadur, R., and Zabell, S. Large deviations of the sample mean in general vector spaces. *Ann. Probab. 7* (1979), 587–621.

[4] Botvich, D., and Duffield, N. Large deviations, the shape of the loss curve, and econimies of large scale multiplexers. *Queueing Syst. 20* (1995), 293–320.

[5] DĘBICKI, K. A note on LDP for supremum of Gaussian processes over infinite horizon. *Stat. Probab. Lett. 44* (1999), 211–220.

[6] DĘBICKI, K., AND MANDJES, M. Exact overflow asymptotics for queues with many Gaussian inputs. *J. Appl. Probab. 40* (2003), 702–720.

[7] DEMBO, A., AND ZEITOUNI, O. *Large deviations techniques and applications, 2nd edition*. Springer, New York, NY, USA, 1998.

[8] DEUSCHEL, J.-D., AND STROOCK, D. *Large Deviations*. Academic Press, Boston, MA, USA, 1989.

[9] DIEKER, T. Extremes of Gaussian processes over an infinite horizon. *Stoch. Proc. Appl. 115* (2005), 207–248.

[10] DUFFIELD, N., AND O'CONNELL, N. Large deviations and overflow probabilities for general single-server queue, with applications. *Math. Proc. Camb. Phil. Soc. 118* (1995), 363–374.

[11] KILPI, J., AND NORROS, I. Testing the Gaussian approximation of aggregate traffic. In *Proc. 2nd Internet Measurement Workshop* (2002), pp. 49–61.

[12] LELAND, W., TAQQU, M., WILLINGER, W., AND WILSON, D. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw. 2* (1994), 1–15.

[13] MANDJES, M. *Large Deviations for Gaussian Queues*. Wiley, Chichester, UK, 2007.

[14] MANDJES, M., MANNERSALO, P., NORROS, I., AND VAN UITERT, M. Large deviations of infinite intersections of events in Gaussian processes. *Stoch. Proc. Appl. 116* (2006), 1269–1293.

[15] MANDJES, M., AND VAN UITERT, M. Sample-path large deviations for tandem and priority queues with Gaussian inputs. *Ann. Appl. Probab. 15* (2005), 1193–1226.

[16] NARAYAN, O. Exact asymptotic queue length distribution for fractional Brownian traffic. *Adv. Perf. Anal. 1* (1998), 39–63.

[17] NORROS, I. A storage model with self-similar input. *Queueing Syst. 16* (1994), 387–396.

[18] NORROS, I. Busy periods of fractional Brownian storage: a large deviations approach. *Adv. Perf. Anal. 2* (1999), 1–20.

[19] PAKES, A. On the tails of waiting-time distributions. *J. Appl. Probab. 12* (1975), 555–564.

[20] SHWARTZ, A., AND WEISS, A. *Large Deviations for Performance Analysis. Queues, Communications, and Computing*. Chapman & Hall, London, UK, 1995.

[21] WEISS, A. A new technique for analyzing large traffic systems. *Adv. Appl. Probab. 18* (1986), 506–532.

[22] ZWART, B., BORST, S., , AND MANDJES, M. Exact asymptotics for fluid queues fed by multiple heavy-tailed on-off sources. *Ann. Appl. Probab. 14* (2004), 903–957.

A Appendix

In this appendix we prove a lemma that is needed to establish Proposition 4.3. We first determine the closure of the set $\mathcal{S}_T$. We define

$$\mathcal{S}(s) := \{f \in \Omega : -f(-s) > p + cs\};$$
$$\mathcal{S}_T(t) := \{f \in \Omega : f(T) - f(T-t) > q + ct\};$$

also

$$\overline{\mathcal{S}(s)} := \{f \in \Omega : -f(-s) \geq p + cs\};$$
$$\overline{\mathcal{S}_T(t)} := \{f \in \Omega : f(T) - f(T-t) \geq q + ct\}.$$

Notice that evidently

$$\mathcal{S}_T = \bigcup_{s,t>0} \left(\mathcal{S}(s) \cap \mathcal{S}_T(t)\right).$$

Lemma A.1. *For any T, we have that the closure $\overline{\mathcal{S}_T}$ of $\mathcal{S}_T$ is given by*

$$\bigcup_{s,t>0} \left(\overline{\mathcal{S}(s)} \cap \overline{\mathcal{S}_T(t)}\right).$$

Proof. The proof is similar to those in [15, 18] We prove both inclusions separately.

- We show first the inclusion "$\subseteq$". For any $f \in \overline{\mathcal{S}_T}$ there exists a sequence $f_n \in \mathcal{S}_T$ such that $||f_n - f||_\Omega \to 0$ as $n \to \infty$. Now since $f_n \in \mathcal{S}_T$ there is an $s_n > 0$ and a $t_n > 0$ such that $f_n \in \mathcal{S}(s_n) \cap \mathcal{S}_T(t_n)$, so that we have

$$-f_n(-s_n) > p + cs_n \quad \text{and} \quad f_n(T) - f_n(T - t_n) > q + ct_n.$$

The sequence s_n is bounded because if not we would have

$$0 = \lim_{n\to\infty} ||f - f_n||_\Omega \geq \lim_{n\to\infty} \frac{f(-s_n) - f_n(-s_n)}{1 + s_n} \geq \lim_{n\to\infty} \left(\frac{f(-s_n)}{1 + s_n} + \frac{p + cs_n}{1 + s_n} \right) = c,$$

(use that $f \in \Omega$!), which gives a contradiction (recall that $c > 0$). Along the same lines it can be shown that t_n is bounded. Hence there are subsequences $s_{n_k} \to s_0$ and $t_{n_k} \to t_0$, for finite s_0 and t_0. Hence for large enough k

$$-f_{n_k}(-s_{n_k}) \geq p + cs_{n_k} \quad \text{and} \quad f_{n_k}(T) - f_{n_k}(T - t_{n_k}) \geq q + ct_{n_k}.$$

We conclude that

$$f \in \left(\overline{\mathcal{S}(s_0) \cap \mathcal{S}_T(t_0)} \right) = \left(\overline{\mathcal{S}(s_0)} \cap \overline{\mathcal{S}_T(t_0)} \right).$$

- For the other inclusion, "$\supseteq$", let

$$f \in \bigcup_{s,t>0} \left(\overline{\mathcal{S}(s)} \cap \overline{\mathcal{S}_T(t)} \right).$$

Then there exist $s, t > 0$ such that $f \in \overline{\mathcal{S}(s)} \cap \overline{\mathcal{S}_T(t)}$. But then there exists a sequence of paths $f_n \in \mathcal{S}(s) \cap \mathcal{S}_T(t)$ such that $||f_n - f||_\Omega \to 0$. This implies that we have for $s, t > 0$

$$-f_n(-s) > p + cs \quad \text{and} \quad f_n(T) - f_n(T - t) > q + ct.$$

This shows that $f_n \in \mathcal{S}_T$, and at the same time we have $||f_n - f||_\Omega \to 0$ as $n \to \infty$, so that $f \in \overline{\mathcal{S}_T}$. □