Centrum voor Wiskunde en Informatica

REPORT *RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Large deviations of Gaussian tandem queues and resulting performance formulae

M.R.H. Mandjes, P. Mannersalo, I. Norros

REPORT PNA-E0412 JULY 2004

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization 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 1386-3711

Large deviations of Gaussian tandem queues and resulting performance formulae

ABSTRACT

This paper considers a two-node tandem queue where the cumulative input traffic is modeled as a Gaussian process with stationary increments. By applying (the generalized version of) Schilder's sample-path large-deviations theorem, we derive the many-sources asymptotics of the overflow probabilities in the second queue; `Schilder' reduces this problem into finding the most probable path along which the second queue reaches overflow. The general form of these paths is described by recently obtained results on infinite intersections in Gaussian processes; for the special cases of fractional Brownian motion and integrated Ornstein-Uhlenbeck input, they can be explicitly determined, as well as the corresponding exponential decay rate. As the computation of this decay rate is numerically involved, we introduce an explicit approximation (`rough full-link approximation'). Based on this approximation, we propose performance formulae that could be used, for instance, for network provisioning purposes. Simulation is used to assess the accuracy of the formulae.

2000 Mathematics Subject Classification: 60F10; 60G15; 60K25

Keywords and Phrases: Tandem queues; Gaussian processes; sample-path large deviations; performance analysis

Large deviations of Gaussian tandem queues and resulting performance formulae

Michel Mandjes Petteri Mannersalo Ilkka Norros

***Abstract*— This paper considers a two-node tandem queue where the cumulative input traffic is modeled as a Gaussian process with stationary increments. By applying (the generalized version of) Schilder's sample-path large-deviations theorem, we derive the many-sources asymptotics of the overflow probabilities in the second queue; 'Schilder' reduces this problem into finding the most probable path along which the second queue reaches overflow. The general form of these paths is described by recently obtained results on infinite intersections in Gaussian processes; for the special cases of fractional Brownian motion and integrated Ornstein-Uhlenbeck input, they can be explicitly determined, as well as the corresponding exponential decay rate. As the computation of this decay rate is numerically involved, we introduce an explicit approximation ('rough full-link approximation'). Based on this approximation, we propose performance formulae that could be used, for instance, for network provisioning purposes. Simulation is used to assess the accuracy of the formulae.**

***Index Terms*— Tandem queues, Gaussian processes, sample-path large deviations, performance analysis**

I. Introduction

Traffic engineering greatly benefits from models that are capable of accurately describing and predicting the performance of the system. The network nodes are usually modeled as *queues*, and queueing theory can be used to analyze the performance (in terms of loss, delay, throughput, etc.) of the nodes. However, most studies address performance issues for single nodes. This is evidently an oversimplification of reality, and justifies research on traffic streams traversing *concatenations* of hops.

Gaussian traffic. There are good reasons for assuming that network traffic is Gaussian. In particular, an application of the central limit theorem leads us to believe that the traffic on communication links will become closer to a Gaussian process as more independent sources add their contribution to the network [1]; see also [2]. The Gaussian traffic model is also popular due to the fact that it covers both short-range (for instance so-called integrated Ornstein-Uhlenbeck) and long-range dependent models. The latter type of dependence was discovered in several measurement studies in real networks: over a wide range of lags, the correlation of traffic follows a power law. This is most succinctly expressed in terms of the variance of the traffic arriving in an interval of length t, which is observed to be proportional to t^{2H} over a wide range of values of t. The parameter H is referred to as the Hurst parameter [3] and typically takes values in the range 0.7 to 0.9. Gaussian process with stationary increments and the variance function of the form t^{2H} is called *fractional Brownian motion* (fBm).

Gaussian tandem queues; negative traffic. As argued above, Gaussian traffic models naturally describe a wide variety of relevant input processes. There is, however, a conceptual difficulty of the use of Gaussian traffic models, namely the fact that negative traffic is not explicitly ruled out, as opposed to 'classical' input processes, such as (compound) Poisson processes or on-off sources. As we will discuss now, for the case of tandem queues with Gaussian input, this does not lead to any practical problems.

- First consider the single-node model, emptied at a constant rate c, where A_t denotes the traffic arriving in $[0,t)$. Then the stationary distribution of the queue is given by the well-known Reich's formula $\sup_{t>0}(A_{-t} - ct)$. Clearly, the distribution of such functionals can be evaluated regardless of the possibility of negative arrivals, and hence also for Gaussian input, see e.g. [4] and [5].
- Now consider the tandem system; for ease we restrict ourselves to a two-node system shown in Figure 1, but the argument can be extended to tandems of any size. To avoid trivialities, we assume the (constant) link speed of queue 1 (c_1) larger than the link speed of queue 2 (c_2). A 'reduction principle' applies: the total queue length is unchanged when the tandem network is replaced by its

Michel Mandjes is with CWI Centre for Mathematics and Computer Science, P.O. Box 94079, NL-1090 GB Amsterdam, the Netherlands. E-mail: michel.mandjes@cwi.nl

Petteri Mannersalo is with VTT Technical Research Centre of Finland, P.O. Box 1202, FI-02044 VTT, Finland. E-mail: petteri.mannersalo@vtt.fi. This work was done while he was at CWI.

Ilkka Norros is with VTT Technical Research Centre of Finland, P.O. Box 1202, FI-02044 VTT, Finland. E-mail: ilkka.norros@vtt.fi

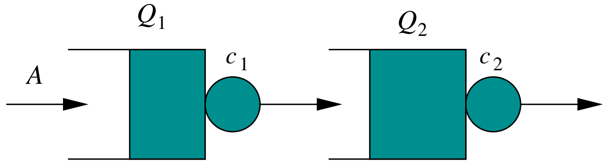

Fig. 1. Two node tandem queue.

slowest queue, see [6], [7]. Hence the total queue is given by $\sup_{t>0}(A_{-t} - c_2 t)$, and the second queue by

$$Q_2 := \sup_{t>0}(A_{-t} - c_2 t) - \sup_{t>0}(A_{-t} - c_1 t). \tag{1}$$

The fact that $c_1 > c_2$ implies that Q_2 is nonnegative. Hence definition (1) is 'proper' (despite the possibility of negative traffic): it cannot lead to negative queue lengths.

Contribution & literature. This paper concentrates on the evaluation of tail asymptotics in tandem queues. Exact analysis of systems with Gaussian input is usually hard (explicit results are only available for standard cases, such as the single queue with Brownian motion and Brownian bridge input), and hence we have to resort to asymptotic regimes. In this paper, we assume that n i.i.d. Gaussian sources feed into the queueing system, where the (deterministic) service rates of the queues as well as the buffer thresholds are scaled by n, too. We now let n go to infinity; the resulting framework is often referred to as the *many-sources* scaling, as was introduced in [8].

A vast body of results exists for single FIFO queues under the many-sources scaling. Most notably, under very mild conditions on the source behavior, it is possible to calculate the *exponential* decay rate of the probability $p_n(b,c)$ that the queue (fed by n sources, and emptied at a deterministic rate nc) exceeds level nb. Logarithmic asymptotics are found in, e.g., [9], [10]; recently exact asymptotics for Gaussian inputs were found by [11]. For Gaussian sources the logarithmic asymptotics of [9] read

$$\lim_{n\to\infty} \frac{1}{n} \log p_n(b,c) = -\inf_{t>0} \frac{(b+(c-\mu)t)^2}{2v(t)}, \tag{2}$$

where μ is the mean input rate per source, and $v(t)$ is the variance of the amount of traffic generated by a single source in a time interval of length t.

Results as (2) cannot be easily generalized to the tandem case. In [12] a lower bound was derived for the decay rate of overflow in the second queue, and this lower bound was under certain conditions 'tight' (in the sense that the lower bound actually equals the decay rate), but there was no tightness for the relevant case of fBm input. [13], [14] provide heuristics for the decay rate of overflow in priority and generalized-processor-sharing systems, such as the (rough) full-link approximation. Numerical studies have shown that these approximations are remarkably accurate.

Our paper has two significant contributions:

1. We have characterized the decay rate of overflow in the second queue, i.e., the 'tandem equivalent' of (2). This was done by first rewriting the event of overflow in the two-node tandem as an infinite intersection of events, and then exploiting recently obtained results [15] on large deviations of these infinite intersections. For the relevant cases of both integrated Ornstein-Uhlenbeck (which we abbreviate to iOU) and fBm input, we found explicit, exact solutions. The techniques applied stem from large-deviations theory, particularly sample-path large deviations, based on (the generalized version of) Schilder's theorem.
2. As the computation of this decay rate is numerically involved, we introduce an explicit approximation ('rough full-link approximation'). Using this decay rate approximation, we propose performance formulae for, for instance, network provisioning purposes. We have performed extensive numerical experiments to assess the accuracy of the decay rate approximations and resulting performance formulae.

This paper is organized as follows. In Section II, first the basic results on the sample path large deviations for Gaussian processes are reviewed, and then special case of the events defined by infinite intersections is considered. Section III studies asymptotics of the tandem queues with general Gaussian input, whereas in Section IV we concentrate on two special processes: fBm and iOU. Section V is devoted to numerical studies. In the end, some conclusions are drawn in Section VI.

II. Preliminaries

This section describes our prerequisites: some fundamental results on Gaussian processes, and a number of results from our earlier work [15].

A. Gaussian processes and Schilder's theorem

The following framework will be used throughout the paper. First we introduce Gaussian processes, and explain that these processes could have different 'degrees of smoothness'. Then we state Schilder's theorem, after having introduced a number of required notions.

Gaussian processes. Let $Z = (Z_t)_{t\in\mathbb{R}}$ be a centered, i.e., $\mathrm{E}Z_t = 0$ for all t, Gaussian process with stationary increments, completely characterized by its variance function

$v(t) \doteq \mathrm{Var}(Z_t)$. A canonical long-range dependent Gaussian process is fBm, with a variance function that is proportional to t^{2H}, with Hurst parameter $H \in (\frac{1}{2}, 1)$. A classical example for a short-range dependent Gaussian process is the iOU process, where the variance function is of the form $t - 1 + e^{-t}$. In general, loosely stated, the more convex the variance function, the stronger the positive correlations.

It is easily verified that the covariance function of Z can be written in terms of the variance function:

$$\Gamma(t,s) \doteq \mathrm{Cov}(Z_t, Z_s) = \frac{1}{2}(v(s) + v(t) - v(s-t)).$$

For a finite subset S of $\mathbb{R}$, denote by $\Gamma(S,t)$ the column vector $\{\Gamma(s,t) : s \in S\}$, by $\Gamma(t,S)$ the corresponding row vector, and by $\Gamma(S)$ the matrix

$$\Gamma(S) \doteq \{\Gamma(s,t) : s \in S, t \in S\}.$$

In addition to the basic requirement that $v(t)$ results in a positive semi-definite covariance function, a number of (technical) assumptions have to be imposed on $v(t)$, see [15]. It is noted that these are fulfilled for the two classical examples (fBm and iOU) mentioned above.

As indicated above, different Gaussian processes could have different 'degrees of smoothness'. We call the Gaussian process Z *smooth* at t, if it has a mean-square derivative at t, that is, there exists a random variable $Z'_t \in G$ such that

$$\lim_{h\to 0} \mathrm{E}\left(\frac{Z_{t+h} - Z_t}{h} - Z'_t\right)^2 = 0.$$

It follows from the stationarity of increments that if Z is smooth at 0, then it is smooth at all $t \in \mathbb{R}$. On the other hand, applying the above definition at $t = 0$, we see that process Z is non-differentiable if $\lim_{h\to 0} v(h)/h^2 = \infty$. It can be shown that fBm is non-smooth, whereas iOU has a mean-square derivative. This difference is crucial in this paper, as it implies that the solutions for fBm and iOU are essentially different.

Schilder's theorem. The remainder of this subsection is devoted to the statement of the main 'tool' used in this paper: Schilder's large-deviations result for Gaussian processes. In this framework a central role is played by the norm $||f||$ of paths f in the reproducing kernel Hilbert space of the underlying Gaussian process. More precisely, 'Schilder' states that the probability of the Gaussian process being in some closed set A has exponential decay rate $\frac{1}{2}||f^*||^2$, where f^* is the path in A with minimum norm, i.e., $\mathrm{argmin}_{f\in A}||f||$. It is noted that for closed and convex A, there exists a unique minimizer. This f^* has the interpretation of the *most probable path* (MPP) in A: if the Gaussian process happens to fall in A, with overwhelming probability it will be close to f^*. An MPP can be intuitively understood as a point of maximum likelihood.

To state Schilder's theorem, we first introduce a number of relevant notions. The path space Ω corresponding to the Gaussian process Z is defined as in [12], [13], and leads to a unique probability measure P. The *reproducing kernel Hilbert space* R related to Z is defined by starting from the functions $\Gamma(t,\cdot)$ and defining an inner product by

$$\langle \Gamma(s,\cdot), \Gamma(t,\cdot)\rangle = \Gamma(s,t). \tag{3}$$

The space is then closed with linear combinations, and completed with respect to the norm $\|\cdot\|^2 = \langle\cdot,\cdot\rangle$. The inner product definition generalizes to the *reproducing kernel property*:

$$\langle f, \Gamma(t,\cdot)\rangle = f(t), \quad f \in R. \tag{4}$$

The generalization of Schilder's theorem on large deviations of Brownian motion to Gaussian measures in a Banach space is originally due to Bahadur and Zabell [16] (see also [17], [18]). Here is a formulation appropriate to our case; for the definition of *good* rate function, see, e.g., [18, Section 2.1].

Theorem 1: The function $I : \Omega \to [0,\infty]$,

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

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

for F closed in Ω:

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left(\frac{Z}{\sqrt{n}} \in F\right) \quad \le -\inf_{\omega\in F} I(\omega);$$

for G open in Ω:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left(\frac{Z}{\sqrt{n}} \in G\right) \quad \ge -\inf_{\omega\in G} I(\omega).$$

Remark 1: With $Z^{(i)}$, $i = 1,\ldots,n$, being i.i.d. copies of Z, it is noted that

$$\frac{Z}{\sqrt{n}} \stackrel{\mathrm{d}}{=} \frac{1}{n}\sum_{i=1}^{n} Z^{(i)}.$$

This implies Schilder's theorem can also be interpreted as a statement on the probability that the 'empirical mean process' of n i.i.d. Gaussian sources.

B. Results on infinite intersections

The central problem dealt with in [15] is of the following form: given a function $\zeta \in R$ and a set of timepoints S, what is the most probable path in the event $\{Z \geq \zeta \text{ on } S\}$? In the tandem setting, we choose a specific form of ζ, as is discussed in Section III.

For any set $S \subset \mathbb{R}$, denote

$$\begin{aligned} B_S &\doteq \{f \in R : f(t) \geq \zeta(t)\ \forall t \in S\}, \\ L_S &\doteq \{f \in R : f(t) = \zeta(t)\ \forall t \in S\}. \end{aligned}$$

and let $\mathrm{Span}\, A$ be the smallest closed linear subspace of R containing the set $A \subseteq R$. The following result was proven in [15]. The theorem implies that in order to determine the MPP it is enough to find the set where the ζ and the optimal path are congruent, i.e., overlapping.

Theorem 2: Let $\zeta \in R$ and $S \subseteq \mathbb{R}$ be compact. Then there exists a function $\beta^* \in B_S$ with minimal norm, i.e.,

$$\beta^* \doteq \operatorname{argmin}\{\|f\| : f \in B_S\}.$$

Moreover, $\beta^* \in \mathrm{Span}\bigcup_{t \in S^*} R_{t\pm}$ where

$$\begin{aligned} S^* &= \{t \in S : \beta^*(t) = \zeta(t)\}, \\ R_{t\pm} &= \bigcap_{u>0} \mathrm{Span}\{\Gamma(s,\cdot) : s \in [t-u, t+u]\}. \end{aligned}$$

In general, the nature of the most probable path β^* depends crucially on the smoothness of Z, as will appear in Section IV. If Z is non-differentiable, like fBm, $R_{t\pm}$ is usually spanned by $\Gamma(t,\cdot)$. In case of iOU, Z has one derivative, and consequently $R_{t\pm}$ contains $(\mathrm{d}/\mathrm{d}t)\,\Gamma(t,\cdot)$. In general, for smooth processes with derivatives up to order k, $R_{t\pm}$ contains also all the derivatives, i.e.,

$$\frac{\mathrm{d}^j}{\mathrm{d}t^j}\Gamma(t,\cdot), \quad j = 1,\ldots,k,$$

see [19] for details.

For any finite $V \subset R$, let the unique element with smallest norm in B_V and L_V be, respectively,

$$\varphi^V \doteq \operatorname{argmin}_{\varphi \in B_V}\|\varphi\|, \quad \overline{\varphi}^V \doteq \operatorname{argmin}_{\varphi \in L_V}\|\varphi\|.$$

By the reproducing kernel properties (3) and (4), we find that $\overline{\varphi}^V(\cdot)$ can be written as linear combination of covariance functions and its norm using the inverse of the covariance matrix:

$$\begin{aligned} \overline{\varphi}^V(\cdot) &= \sum_{v \in V} \theta_v \Gamma(v,\cdot), \qquad (5)\\ \|\overline{\varphi}^V\|^2 &= \zeta(V)\,\Gamma(V)^{-1}\zeta(V) \end{aligned}$$

where the vector $\theta(V) = (\theta_v)_{v \in V}$ is given by $\theta(V) = \Gamma(V)^{-1}\zeta(V)$ with $\zeta(V) = (\zeta(v))_{v \in V}$. Note that for any $V \subseteq S$, $\|\varphi^V\|$ is a lower bound on $\|\beta^*\|$, but it is possible that $\|\overline{\varphi}^V\| > \|\beta^*\|$.

Next, another result from [15] shows that the coefficients of the $\Gamma(v,\cdot)$, $v \in V$ in the representation of $\overline{\varphi}^V$ are strictly positive, as long as every v is needed to make function φ^V feasible.

Proposition 1: Assume a finite V. If for each $v \in V$ it holds that $\overline{\varphi}^{V\setminus\{v\}}(v) < \zeta(v)$, then the coefficients θ_v in the representation $\overline{\varphi}^V = \sum_{v \in V}\theta_v \Gamma(v,\cdot)$ are all strictly positive.

III. Tandem queues

Consider a two-queue tandem model with infinite buffers at both nodes. The input process $A_t = Z_t + \mu t$ is modeled as a Gaussian process with stationary increments, where μ is the mean rate and Z is a centered Gaussian process. The queues are served with deterministic service rates c_1 for the first queue and c_2 for the second queue. We assume $c_1 > c_2$, in order to exclude the trivial case where the second queue cannot build up. Moreover, we restrict ourselves to centered A by setting $\mu = 0$ and $A \doteq Z$, since the constant drift can be included in the server rates [12, Remark 2.6].

As argued in the introduction, the stationary queue length of the first queue reads $Q_1 = \sup_{t \geq 0}(Z_{-t} - c_1 t)$. Also, the total queue length behaves as a queue with link rate c_2, i.e., $Q_1 + Q_2 = \sup_{t \geq 0}(Z_{-t} - c_2 t)$. Therefore, expressing the occupancy of the second queue as the difference of the total buffer content and the content of the first queue, we find

$$\{Q_2 \geq b\} = \{\exists t \geq 0 : \forall s \geq 0 : Z_{-t} - Z_{-s} - c_2 t + c_1 s \geq b\};$$

it is easily seen that we can restrict ourselves to $s \in [0,t]$, and $t \geq t_b \doteq b/(c_1 - c_2)$, see [12, Lemma 2.4].

In this paper, our (first) aim is to determine

$$I(b) \doteq -\lim_{n\to\infty}\frac{1}{n}\log \mathrm{P}(Q_{2,n} \geq nb),$$

where $Q_{2,n}$ is the steady state queue length of the second queue in the system with n i.i.d. Gaussian sources served at the rates nc_1 and nc_2. In [12, Thm. 3.1], it was shown that $I(b)$ equals the rate function of $\mathrm{P}(Q_{2,n} > nb)$. We now rephrase $I(b)$ by applying Schilder's theorem. To this end, define the function

$$\zeta_{t,b}(s) \doteq -\alpha_1(t,b) + \alpha_2 s,$$

where $\alpha_1(t,b) = (c_1 - c_2)t - b$ and $\alpha_2 = c_1$, and denote

$$U_{t,b} \doteq \{f \in R : f(s) \geq \zeta_{t,b}(s) \forall s \in [0,t]\}.$$

Invoking the relation between Schilder and the many-sources setting (as in Remark 1), the following result is due to a straightforward time-shift, see also [12, Remark 2.5]. It shows that determination of the rate function can be partitioned into two steps: first find the most probable paths and their norms in the $U_{t,b}$ with fixed t and b, and after that optimize t. We omit the proof.

Theorem 3:

$$I(b) = \inf_{f\in U_b} \frac{1}{2}\|f\|^2 = \inf_{t\geq t_b}\inf_{f\in U_{t,b}} \frac{1}{2}\|f\|^2 = \inf_{t\geq t_b} \frac{1}{2}\|\beta^*_{t,b}\|^2,$$

where $U_b \doteq \bigcup_{t\geq t_b} U_{t,b}$ *and* $\beta^*_{t,b} = \mathrm{argmin}\,\{\|f\|:\ f\in U_{t,b}\}$.

When determining the MPP in $U_{t,b}$, we find that there are two regimes. In the 'simple regime', the MPP is just a single scaled covariance function. In the complementary situation, we need more than one (possible infinitely many) covariance functions and their derivatives if process is differentiable.

To precisely introduce both regimes, we first note that if $\zeta_{t,b}(s) > 0$, then, by (5),

$$\overline{\varphi}^s(\cdot) = \varphi^s(\cdot) = \frac{\zeta_{t,b}(s)}{v(s)}\Gamma(s,\cdot), \quad \|\varphi^s\|^2 = \frac{\zeta_{t,b}(s)^2}{v(s)}.$$

If we assume a variance function $v(\cdot)$ such that the above norm has its maximum on the interval $[0,t]$ at t, then we can consider the function

$$\frac{\zeta_{t,b}(s)}{v(s)}\Gamma(s,\cdot) - \zeta_{t,b}(s) \quad \forall s\in[0,t], \tag{6}$$

leading to the following classification.

Case 1: If

$$\frac{\alpha_1(t,b)}{\alpha_2} \geq \alpha^F(t) \doteq \sup_{s\in[0,t]}\left\{\frac{sv(t)-t\Gamma(t,s)}{v(t)-\Gamma(t,s)}\right\},$$

then

$$\beta^*_{t,b}(\cdot) = \frac{\zeta_{t,b}(t)}{v(t)}\Gamma(t,\cdot), \quad \|\beta^*_{t,b}\|^2 = \frac{\zeta_{t,b}(t)^2}{v(t)}. \tag{7}$$

Case 2: If $\alpha_1(t,b)/\alpha_2 < \alpha^F(t)$, then the path (6) is not feasible, and consequently a single covariance function is not enough.

Note that $\zeta_{t,b} \notin R$. However, this is not a problem since $\zeta_{t,b}(0) \leq 0$ and

$$\zeta_{t,b}\left(\frac{\alpha_1(t,b)}{\alpha_2} + \cdot\right) \in R,$$

so that we can approximate $\zeta_{t,b}$ on $(0,t)$ by a sequence of R-functions. Thus the results (Thm. 2 and Prop. 1) on infinite intersections hold.

By Theorem 2, if we could find the set

$$S^*_{t,b} \doteq \left\{s\in[0,t]:\ \beta^*_{t,b}(s) = \zeta_{t,b}(s)\right\},$$

then the most probable path would be known also in Case 2. Unfortunately, we do not have any general recipe for that at our disposal, and determining $S^*_{t,b}$ can be a difficult task. In Section IV, we solve this problem for the special cases of fBm and iOU input. Before that, we consider three approximations that we developed earlier and that will serve as benchmarks in Section V.

A. *Lower bound on the decay rate*

Mandjes and van Uitert [20], [12] find a lower bound $I^L(b)$ of $I(b)$. Let t^*_b minimize

$$J_b(t) \doteq \frac{(b+c_2t)^2}{2v(t)},$$

and let $k_b(s,t) \doteq \Gamma(s,t)(b+c_2t)/v(t)$. Then it is proven that for

$$c_1 \geq c_1^F(b) \doteq \sup_{s\in(0,t^*_b)} \frac{k_b(s,t^*_b)}{s},$$

it holds that $I(b) = I^L(b) \doteq J_b(t^*_b)$. The above condition is, of course, equivalent to $\alpha_1(t^*_b,b)/\alpha_2 \geq \alpha^F(t^*_b)$.

Now consider the opposite case. For $c_2 < c_1 < c_1^F(b)$), we have that $I(b) \geq I^L(b)$, with

$$I^L(b) \doteq \inf_{t\geq t_b}\sup_{s\in K_b} \frac{1}{2} x(s,t)^{\mathrm{T}} \begin{pmatrix} v(t) & \Gamma(s,t) \\ \Gamma(s,t) & v(s) \end{pmatrix}^{-1} x(s,t),$$

where $K_b = \left\{s\in S:\ k_b(s,t^*_b) < \zeta_{t^*_b,b}(s)\right\}$ and $x(s,t)$ is the two-dimensional vector $(b+c_2t, b+c_2t-c_1(t-s))^{\mathrm{T}}$.

Hence, in the regime $c_1 \geq c_1^F(b)$ the lower bound is always tight, in that $I(b) = I^L(b)$. However, also for the regime $c_2 < c_1 < c_1^F(b)$ [12] presents an explicit condition under which the above lower bound is tight; this condition is *not* fulfilled in the case of fBm.

B. *Rough full-link approximation*

As we have the exact decay rate for $c_1 \geq c_1^F(b)$, special interest is in the other regime. We here consider an approximation for $I(b)$ for the case $c_1 \in (c_2, c_1^F)$. In the context of priority and generalized-processor-sharing queues, Mannersalo and Norros [13], [14] proposed the *rough full-link approximation*. Here we extend this approximation to the tandem case. The idea is that if $c_1 \in (c_2, c_1^F)$, the source transmits at approximately a rate c_1 during t_b units of time, thus causing exceedance of level b in the second queue. This leads to the approximation $I^R(b) = I^L(b)$ for $c_1 \geq c_1^F(b)$, and

$$I^R(b) = \frac{(b + c_2 t_b)^2}{2v(t_b)}$$

for $c_2 < c_1 < c_1^F(b)$.

Mathematically, this approximation can be motivated by replacing the $U_{t,b}$ by the larger sets $B_{t,b} = \{f \in R : f(t) \geq \zeta_{t,b}(t)\}$; hence a requirement is imposed for time t, rather than for all $s \in [0,t]$. Consequently,

$$\begin{aligned} I(b) &= \inf_{t \geq t_b} \inf_{f \in U_{t,b}} \frac{1}{2}\|f\|^2 \geq \inf_{t \geq t_b} \inf_{f \in B_{t,b}} \frac{1}{2}\|f\|^2 \\ &= \inf_{t \geq t_b} \frac{(b + c_2 t)^2}{2v(t)}. \end{aligned} \tag{8}$$

Thus $I^R(b)$ is a lower bound to $I(b)$, if $t_b^* \leq t_b$ and if the variance function v is such that the minimum is attained at $t = t_b$ in (8).

C. *Upper bound for the decay rate*

Any feasible path, i.e., $u \in U_b$ gives an upper bound to the decay, since, according to 'Schilder',

$$I(b) = \inf \left\{ \frac{1}{2}\|f\|^2 : f \in U_b \right\} \leq \frac{1}{2}\|u\|^2.$$

For tandem queues, the natural upper bound comes from the most probable path for a busy period of length t_b in the first queue (as a busy period of t_b in the first queue implies that $t_b(c_1 - c_2) = b$ traffic is built up in the second queue). Hence,

$$I^B(b) \doteq \inf_{f \in U_{t_b,b}} \frac{1}{2}\|f\|^2 \geq \inf_{t \geq t_b} \inf_{f \in U_{t,b}} \frac{1}{2}\|f\|^2 = I(b)$$

The evaluation of $I^B(b)$ is done as in [15].

D. *Performance formulae*

So far we have concentrated on determining (approximations of) the exponential decay rate $I(b)$. In the many-sources setting, these can be also used to (roughly) characterize the probability distribution itself:

$$\mathrm{P}(Q_{2,n} > nb) \approx \exp(-nI(b)).$$

In practice, numerical evaluation of the exact decay $I(b)$ is a difficult task, as follows from our explicit formulae in the next section.

Therefore, for engineering purposes, we propose to rely on an approximation based on the rough full-link approximation:

$$\mathrm{P}(Q_{2,n} > nb) \approx \exp(-nI^R(b)).$$

As in most cases $I(b) \geq I^R(b)$, this approximation tends to be conservative, and is consequently appropriate for, e.g., provisioning purposes.

For Gaussian processes, the decay rate of the many-sources asymptotics is also useful in approximating the queue-length distribution of a tandem queue fed by a single source. The following performance formula was originally introduced for single-node queues in [21], [22]:

$$\mathrm{P}(Q_2 > b) \approx \exp(-I^R(b)). \tag{9}$$

According to numerical studies with single-node queues, it seems that $\exp(-I(b))$ is an upper bound for the tail distribution. The same holds for tandem queues, as seen in Section V. Unfortunately, no formal proof for such property exists.

IV. Most probable paths for FBM and Integrated Ornstein-Uhlenbeck

Let us now focus on the inner minimization problem in $I(b) = \inf_{t \geq t_b} \inf_{f \in U_{t,b}} \frac{1}{2}\|f\|^2$, i.e., minimization over $U_{t,b}$ with fixed t and b. We can restrict ourselves to $\alpha_1(t,b)/\alpha_2 < \alpha^F(t)$ (Case 2), since the complementary case (Case 1) was already solved by (7). In the following two subsections, the most probable paths and their norms are determined for the special cases of fBm and iOU input.

After fixing t and b, we can simplify our notation by denoting $\alpha_2 = \alpha_2(t,b)$, $\zeta(\cdot) = \zeta_{t,b}(\cdot)$, and $\alpha^F = \alpha^F(t)$. More precisely, we consider the set $B_{[0,t]} \doteq U_{t,b}$, and the corresponding MPP

$$\beta^* = \operatorname{argmin}\left\{\|f\| : f \in B_{[0,t]}\right\}.$$

A. *Fractional Brownian Motion*

Consider fractional Brownian motion which is a centered Gaussian process with stationary increments and variance function $v(t) = t^{2H}$, $H \in (0,1)$.

We first state some results from [15] without proofs. The first theorem shows that we can construct a sequence of sets $S^n \doteq \{\mathbf{s} \in \mathbb{R}^n : 0 < s_1 < \cdots < s_n \leq t\}$ and corresponding sequence of the functions φ^{S^n} such that φ^{S^n} converges to β^*. In addition to fBm, the same also holds for a large family of non-differentiable Gaussian processes.

Theorem 4: Let Z be a centered fBm and denote

$$\begin{aligned} h^n &\doteq \sup\left\{\|\varphi^V\| : V \subseteq [0,t], |V| \le n\right\} \\ n^* &\doteq \inf\left\{n \in \mathbb{N} : h^n = h^{n+1}\right\}. \end{aligned}$$

Then,

- *For each n, there exists a set* $S^n \subseteq [0,t]$ *with at most n elements such that* $\|\varphi^{S^n}\| = h^n$*;*
- *If* $\|\varphi^{S^n}\| = \|\varphi^{S^{n+1}}\|$ *for some n, then* $\beta^* = \varphi^{S^{n^*}}$*;*
- *If* $n \le n^*$*, then* $\varphi^{S^n} = \overline{\varphi}^{S^n}$*;*
- $\lim_{n\to\infty} \varphi^{S^n} = \beta^*$.

From now on, we denote

$$\varphi_n(\cdot) \doteq \varphi^{S^n}(\cdot).$$

The following properties are crucial for the explicit determination of the MPP, as they show that φ_n touches ζ at points s_i (with $i = 1,\ldots,n-1$) from below. Again, the proofs can be found in [15]. Note that if the set S^n exists, then Equation (10) holds for any Gaussian process whose variance function is differentiable on the whole real line.

Theorem 5: Let Z be a centered fBm and assume $n \le n^*$*. For* $H \in (\frac{1}{2}, 1)$ *and for all* $s_i \in S^n \cap (0,t)$*,*

$$\varphi_n'(s_i) = \zeta'(s_i). \tag{10}$$

For $H \in (0, \frac{1}{2})$ *and for all* $s_i \in S^n$*,*

$$\lim_{s \nearrow s_i} \varphi_n'(s) = \infty, \quad \lim_{s \searrow s_i} \varphi_n'(s) = -\infty. \tag{11}$$

For $H \in (0,1)$ *and for all* $s_i \in S^n$*,*

$$\lim_{s \to s_i} \varphi_n''(s) = -\infty. \tag{12}$$

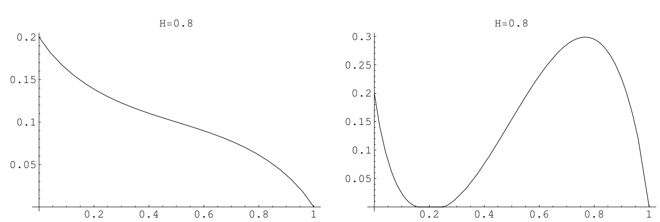

Fig. 2. $\beta^*(s) - \zeta(s)$ for fBm with $H = 0.8$ and $t = 1$. On the left, $\alpha_1/\alpha_2 > \alpha^F$, and on the right $\alpha_1/\alpha_2 < \alpha^F$.

Using Theorems 4 and 5, we can show that the MPP in $B_{[0,t]}$ has one of the shapes shown in Figure 2: either the path touches the condition ζ only at the 'end point' t, or, in addition to the 'end point', the paths coincide over an interval. In the latter case, the MPP is determined by a linear combination of the covariance functions over an infinite index set and thus calculating the norm involves (complicated) integrations. Denote

$$\zeta^*_{[s_1,s_2]} = \operatorname{argmin}\left\{\|f\| : f \in R,\ f(s) = \zeta(s)\ \forall s \in [\underline{s}^*, \overline{s}^*]\right\},$$

i.e., the most probable path which follows ζ on $[s_1, s_2]$. The following theorem is our main result for tandem queues with fBm input.

Theorem 6: Assume $H \in (\frac{1}{2}, 1)$ *and* $\frac{\alpha_1}{\alpha_2} < \alpha^F$*. Then* S^* *is of the form* $[\underline{s}^*, \overline{s}^*] \cup \{t\}$*, where* $\frac{\alpha_1}{\alpha_2} \le \underline{s}^* < \overline{s}^* < t$ *and the function* β^* *has the expression*

$$\begin{aligned} \beta^*(s) &= \mathrm{E}\left[Z_s \mid Z_\tau = \zeta(\tau)\ \forall \tau \in [\underline{s}^*, \overline{s}^*],\ Z_t = t\right] \\ &= \zeta^*_{[\underline{s}^*,\overline{s}^*]}(t) + \frac{\mathrm{Cov}\left[Z_t, Z_s \mid \mathscr{F}\right]}{\mathrm{Var}\left[Z_t \mid \mathscr{F}\right]}\left(\zeta(t) - \zeta^*_{[\underline{s}^*,\overline{s}^*]}(t)\right), \end{aligned}$$

where $\mathscr{F} = \mathscr{F}_{[\underline{s}^*,\overline{s}^*]} = \sigma(Z_s : s \in [\underline{s}^*, \overline{s}^*])$*, and*

$$\|\beta^*\|^2 = \|\zeta^*_{[\underline{s}^*,\overline{s}^*]}\|^2 + \frac{\left(\zeta(t) - \zeta^*_{[\underline{s}^*,\overline{s}^*]}(t)\right)^2}{\mathrm{Var}\left(Z_t - \mathrm{E}\left[Z_t \mid \mathscr{F}\right]\right)}.$$

Proof: This is a slight modification of the proof for [15, Thm. 5] where the case $\alpha_1 = 0$ (equivalent to the busy period problem) is solved. Thus we can assume $\alpha_1 > 0$. The proof is partitioned into two parts: first we show that β^* has the claimed shape, and then we determine its norm.

Shape of β^*: Let us study the properties of the sequence φ_n which converges to β^* by Theorem 4. Since $\alpha_1/\alpha_2 < \alpha^F$ there exists $s \in (0,t)$ such that $\varphi_1(s) < \zeta(s)$ and we have $\|\varphi_1\| < \|\varphi_2\|$. Thus the case $n^* = 1$ is ruled out.

Now assume $n < n^*$. Then $\varphi_n(\cdot) = \sum_{s \in S^n} \theta_s \Gamma(s, \cdot)$. Since $\theta_{s_i} > 0$ (Proposition 1) and $\Gamma(s_i, \cdot) \ge 0$ (positive correlations), $\varphi_n(s) \ge 0$ for all $s \in \mathbb{R}$. Thus we can restrict ourselves to the set

$$\{s \in [0,t] : \zeta(s) \ge 0\} = \left[\frac{\alpha_1}{\alpha_2}, t\right]$$

and necessarily $S^n \subseteq [\alpha_1/\alpha_2, t]$.

Now consider function φ_n and its derivatives:

$$\begin{aligned} \varphi_n'(s) &= C\left[t^\alpha + \sum_{\substack{s_i \in S^n \\ s_i > s}} \rho_{s_i}(s_i - s)^\alpha - \sum_{\substack{s_i \in S^n \\ s_i < s}} \rho_{s_i}(s - s_i)^\alpha\right], \\ \varphi_n''(s) &= \alpha C\left[t^{\alpha-1} - \sum_{s_i \in S^n} \rho_{s_i}(s - s_i)^{\alpha-1}\right], \end{aligned}$$

where

$$\alpha \doteq 2H - 1, \quad C \doteq H \sum_{s_i \in S^n} \theta_{s_i}, \quad \rho_i \doteq \frac{\theta_{s_i}}{\sum_{s_j \in S^n} \theta_{s_j}} \in (0,1).$$

If $s_n = \max\{S^n\}$ and $s_n < t$, then $\varphi_n(s) < \zeta(s)$ for all $s > s_n$, since $\varphi_n(s_n) = \zeta(s_n)$, $\varphi_n'(s_n) = \zeta'(s_n)$ (by (10)), and $\varphi_n''(s) < 0$ for all $s \ge s_n$ (by the fact that the ρ_i sum up to 1, and $\alpha - 1 < 0$). Thus

$$\max\{s \in S^n\} \to t \quad \text{as } n \to n^*.$$

Similarly as in [15, Thm. 5], with lengthy calculations, one can show that for each n there exist $\underline{u}_n$ and $\bar{u}_n$,

$$\frac{\alpha_1}{\alpha_2} \le \underline{u}_n < s_1 < s_{n-1} < \bar{u}_n < s_n,$$

such that φ_n is at or below the condition ζ on $[\underline{u}_n, \bar{u}_n] \cup \{s_n\}$ and strictly above ζ on $[0, \underline{u}_n)$ and $[\bar{u}_n, s_n)$, see Figure 3 for the shapes of φ_n, $n = 1,2,3$. Since the above holds for any n, we have $n^* = \infty$. The convergence $\varphi_n \to \beta^*$ implies that β^* has the claimed shape.

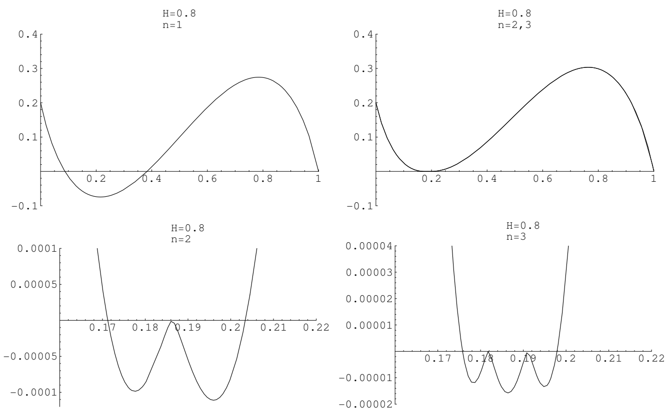

Fig. 3. $H = 0.8$. The shape of $\varphi_n(s) - \zeta(s)$, $n = 1,2,3$. At the scale $[0,1]$, cases $n = 2,3$ are indistinguishable. The lower pictures are zoomed in around the point s_1 (and s_2 for $n = 3$).

Norm of β^*: For any function $f \in R$, define

$$\begin{aligned} \varphi_f(s) &= \mathrm{E}\left[Z_s \,|\, Z_\tau = f(\tau)\ \forall\tau \in [\underline{s}^*, \bar{s}^*]\right], \\ \psi_f(s) &= \mathrm{E}\left[Z_s \,|\, Z_\tau = f(\tau)\ \forall\tau \in [\underline{s}^*, \bar{s}^*];\ Z_t = \zeta(t)\right]. \end{aligned}$$

The conditional distribution of the pair (Z_s, Z_t) w.r.t. $\mathscr{F}$ is a two-dimensional Gaussian distribution with (random) mean $\mathrm{E}\left[(Z_s, Z_t) \,|\, \mathscr{F}\right]$. Thus, a further conditioning on the event $\{Z_t = \zeta(t)\}$ can be computed according to the standard formula of conditional expectation in a bivariate Gaussian distribution:

$$\begin{aligned} \psi_f(s) &= \varphi_f(s) + \frac{\mathrm{Cov}\left[Z_s, Z_t \,|\, \mathscr{F}\right]}{\mathrm{Var}\left[Z_t \,|\, \mathscr{F}\right]}\left(\zeta(t) - \varphi_f(t)\right) \\ &= \varphi_f(s) + c(s)(\zeta(t) - \varphi_f(t)), \end{aligned}$$

where $c(s) = \mathrm{Cov}\left[Z_s, Z_t \,|\, \mathscr{F}\right] / \mathrm{Var}\left[Z_t \,|\, \mathscr{F}\right]$ does not depend on f. Applying this to the function $f(t) \equiv 0$ yields $c(t) = \psi_0(t)$. One can show (see [15]) that

$$\|\psi_0\|^2 = \mathrm{Var}\left(Z_t - \mathrm{E}\left[Z_t \,|\, \mathscr{F}\right]\right)^{-1}.$$

Now, note that

$$\begin{aligned} \beta^*(s) &= \mathrm{E}\left[Z_s \,|\, Z_s = \zeta(s),\ \forall s \in [\underline{s}^*, \bar{s}^*],\ Z_t = \zeta(t)\right] \\ &= \psi_{\zeta^*_{[\underline{s}^*,\bar{s}^*]}}, \end{aligned}$$

$\varphi_{\zeta^*_{[\underline{s}^*,\bar{s}^*]}} = \zeta^*_{[\underline{s}^*,\bar{s}^*]}$, and ψ_0 is orthogonal to $\zeta^*_{[\underline{s}^*,\bar{s}^*]}$. Thus,

$$\|\beta^*\|^2 = \|\zeta^*_{[\underline{s}^*,\bar{s}^*]}\|^2 + \frac{(\zeta(t) - \zeta_{[\underline{s}^*,\bar{s}^*]}(t))^2}{\mathrm{Var}\left(Z_t - \mathrm{E}\left[Z_t \,|\, \mathscr{F}\right]\right)}$$

as claimed. ■

When $H \in (0, \frac{1}{2})$ we are always in Case 2, and consequently a single covariance function never suffices. Moreover, a finite number of the covariance functions is never enough and we need to condition with respect to an interval, i.e., the MPP is determined by a linear combination of the covariance functions over an infinite index set. The shape of the MPP is shown in Figure 4.

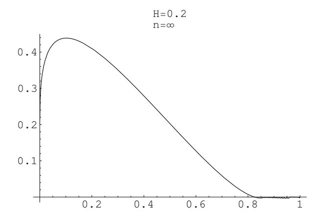

Fig. 4. $\beta^*(s) - \zeta(s)$ for fBM with $H = 0.2$ and $t = 1$.

Theorem 7: Assume $H \in (0, \frac{1}{2})$*. Then the set* S^* *has the form* $[s^*, t]$*, where* $0 < s^* < t$ *and the function* β^* *has the expression*

$$\begin{aligned} \beta^*(s) &= \mathrm{E}\left[Z_s \,|\, Z_\tau = \zeta(\tau) \forall\tau \in [s^*, t]\right] \\ &= \zeta^*_{[s^*,t]}(t) \end{aligned}$$

and

$$\|\beta^*\|^2 = \|\zeta^*_{[s^*,t]}\|^2.$$

Proof: Let us first show that $\alpha^F = \infty$. It is easy to see that $S^1 = \{t\}$ so that $\varphi_1(t) = \zeta(t)$. On the other hand, $\lim_{s \nearrow t} \varphi_1'(s) = \infty$ by (11). Thus, whenever α_2 is finite, $\varphi_n(t - \varepsilon) < \zeta(t - \varepsilon)$ for some $\varepsilon > 0$ and $\varphi_1 \notin B_{[0,t]}$. Hence, the 'easy' Case 1 solution never exists.

More generally, using equations (11) and (12) and similar type of argument as for $H > 1/2$, it is seen that the shapes of the φ_n are such that the limiting path must satisfy: $\beta^*(s) > \zeta(s)$ if $s \in (0, s^*)$ and $\beta^*(s) = \zeta(s)$ if $s \in [s^*, t]$ for some $s^* \in (0, t)$. See Figure 5 for the shapes of φ_n, $n = 1,2,3$. ■

Note that, in principle, all the quantities in the expressions for β^* could be computed. For example when $H \in (1/2, 1)$, an equivalent formulation is

$$\begin{aligned} \beta^* = \ & \operatorname{argmin}_{f \in R}\{\|f\| :\ f(s) = \alpha_2 s,\ s \in [0, \bar{s}^* - \underline{s}^*], \\ & f(-\underline{s}^*) = -\alpha_2 \underline{s}^*,\ f(t - \underline{s}^*) = \alpha_2(t - \underline{s}^*)\} \end{aligned}$$

As a consequence, we can consider the MPP following the line $\alpha_2 s$ on $[0, \bar{s}^* - \underline{s}^*]$ and and hitting two other points

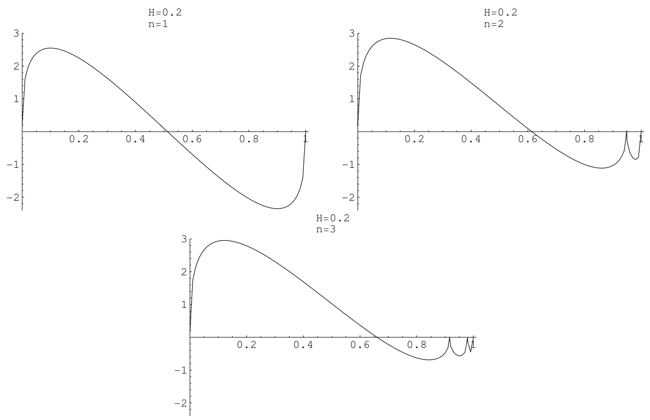

Fig. 5. $H = 0.2$. The shape of $\varphi^{S^n}(s) - \zeta(s)$, $n = 1, 2, 3$.

at $t - \underline{s}^*$ and $-\underline{s}^*$. The straight-line part is equivalent to χ-path of [23], [24]. Also the (two-dimensional) conditioned Gaussian random variable has semi-explicit representation (see e.g. [25]), though consisting of multiple integrals. We have not succeeded in finding explicit expressions for the numbers $\underline{s}^*$ and $\bar{s}^*$. However, by knowing the structure of S^*, or even by just knowing from Theorem 2 that the MPP is determined by a set where it touches ζ, it is easy to obtain arbitrarily accurate discrete approximations of the MPPs and their norms using some graphical mathematical tool.

B. Integrated Ornstein-Uhlenbeck input

Unfortunately, the results of Theorem 4 do not hold generally for smooth processes. The main reason for that is the larger infinitesimal spaces $R_{t\pm}$ containing also the derivatives of the covariance functions. This means that mapping $\mathbf{t} \mapsto \varphi^{\mathbf{t}}$, with $\mathbf{t} \in \mathbb{R}^n$, is not always continuous and thus the existence of S^n is not guaranteed.

Consider a Gaussian process Z_t with stationary increments and variance $v(t) = t - 1 + e^{-t}$. This is an iOU model, which can be interpreted as the Gaussian counterpart of the Anick-Mitra-Sondhi model [22]. Since the *rate process* is defined by the stochastic differential equation

$$\mathrm{d}X_t = -\gamma X_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t,$$

where W denotes the standard Brownian motion, Z is exactly once differentiable. In the above differential equation both γ and σ should be equated to 1 to get the desired variance function.

Our main result for tandem queues with iOU input says that the lower bound of [12] is tight: $I^L(b) = I(b)$. Again, Case 1 ($\alpha_1/\alpha_2 \geq \alpha^F$) is trivial and the decay rate is given by (7).

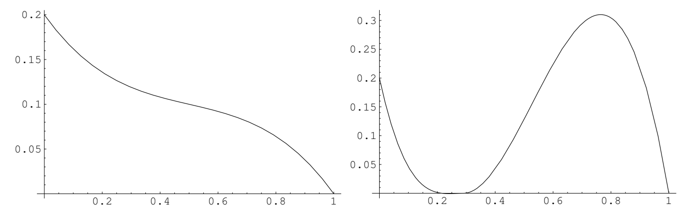

Fig. 6. Integrated Ornstein-Uhlenbeck input: $\beta^*(t) - \zeta(t)$, when $\alpha_1 = 0.2$, $\alpha_2 = 2$ (left) and $\alpha_2 = 20$ (right).

Theorem 8: Assume $\alpha_1/\alpha_2 < \alpha^F$. If $\alpha_1 > 0$, then $S^* = \{s^*, t\}$, with $s^* \in (0, t/2)$, and the function β^* has the expression

$$\beta^*(s) \quad = \quad (\zeta(s^*), \zeta(t))\Gamma((s^*, t))^{-1}(\Gamma(s^*, s), \Gamma(t, s))^T$$

where s^* minimizes the norm

$$(\zeta(s), \zeta(t))\Gamma(s, t)^{-1}(\zeta(s), \zeta(t))^T$$

with $s \in \left\{ s \in [0, t] : \overline{\varphi}^t(s) < \zeta(s) \right\}$. If $\alpha_1 = 0$, then

$$\beta^*(\cdot) = (\zeta'(0), \zeta(t)) \begin{pmatrix} \frac{1}{2}v''(0) & \frac{1}{2}v'(t) \\ \frac{1}{2}v'(t) & v(t) \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2}v'(\cdot) \\ \Gamma(t, \cdot) \end{pmatrix}.$$

and

$$\|\beta^*\|^2 = (\zeta'(0), \zeta(t)) \begin{pmatrix} \frac{1}{2}v''(0) & \frac{1}{2}v'(t) \\ \frac{1}{2}v'(t) & v(t) \end{pmatrix}^{-1} \begin{pmatrix} \zeta'(0) \\ \zeta(t) \end{pmatrix}.$$

Proof: The case $\alpha_1 = 0$ is solved by finding the minimal element of the set

$$\left\{ f \in R : \; f'(0) = \zeta'(0), \; f(t) = \zeta(t) \right\}$$

and showing that it is also an element of $B_{[0,t]}$; for details see [15].

Now assume $\alpha_1 > 0$. Although Theorem 4 does not hold in general, we can still apply a similar method. Consider the set

$$B_s \doteq \{ f \in R : \; f(s) \geq \zeta(s), \; f(t) \geq \zeta(t) \}$$

and define the 'least likely' point

$$s^* = \operatorname{argmax}_{s \in [0,t]} \{ \|f\| : \; f \in B_s \},$$

which exists whenever $\alpha_1 > 0$. Clearly,

$$s^* \in \left\{ s \in [0, t] : \; \overline{\varphi}^t(s) < \zeta(s) \right\} \subset [\alpha_1/\alpha_2, t/2],$$

where the right end point of the interval follows from the antisymmetry of $s \mapsto \Gamma(t, s) - s$ around point $t/2$.

Denote

$$g \doteq \operatorname{argmin} \{ \|f\| : \; f \in B_{s^*} \} = \theta_1 \Gamma(t, \cdot) + \theta_2 \Gamma(s^*, \cdot).$$

Since $B_{[0,t]} \subset B_{s^*}$, we have $\|\beta^*\| \geq \|g\|$ and we only need to prove that the path g is feasible, i.e., $g \in B_{[0,t]}$. The basic idea is to show that path g is convex on $[0,s^*]$, and first convex and then concave on $[s^*,t]$ (see the right picture in Figure 6).

Assume first $s \in (s^*,t]$. Then

$$g'(s) = \frac{1}{2}\theta_1(2 - e^{-s} - e^{-(t-s)}) + \frac{1}{2}\theta_2(e^{-s} - e^{-(s-s^*)}),$$
$$g''(s) = \frac{1}{2}\theta_1(e^{-s} - e^{-(t-s)}) + \frac{1}{2}\theta_2(-e^{-s} + e^{-(s-s^*)}).$$

The second derivative has at most one zero on $[s^*,t]$. However, since $g(s^*) = -\alpha_1 + \alpha_2 s^*$, $g(t) = -\alpha_1 + \alpha_2 t$ and $g'(s^*) = \alpha_2$ (similarly to (10)), there is exactly one point $s_0 \in [s^*,t]$ such that $g''(s_0) = 0$. On the other hand, $s^* < t/2$ implies that $g''(s^*+) > 0$. Thus $g(s) \geq \zeta(s)$ for all $[s^*,t]$.

Next consider $s \in [0,s^*]$, when

$$g'(s) = \frac{1}{2}\theta_1(2 - e^{-s} - e^{-(t-s)}) + \frac{1}{2}\theta_2(2 - e^{-s} - e^{-(s^*-s)}),$$
$$g''(s) = \frac{1}{2}\theta_1(e^{-s} - e^{-(t-s)}) + \frac{1}{2}\theta_2(e^{-s} + e^{-(s^*-s)}).$$

Similarly as above, the second derivative has at most one zero. On the other hand, we have $g''(s^*-) > 0$ and $g''(0) > 0$. This implies that $g''(s) > 0$ for all $s \in [0,s^*]$. Thus $g(s) \geq \zeta(s)$ for all $s \in [0,s^*]$ and $g \in B_{[0,t]}$ as claimed. ■

V. Numerical studies

A. Comparison of the decay rates

As seen in the previous section, the computation of the exact decay rate is numerically involved. Fortunately, the proposed approximate decay rates are close enough to the exact outcome, at least from the performance analysis viewpoint.

For iOU input, we have already proved in Theorem 8 that $I^L = I$. In Figure 7, we compare the exact decay rate I with the rough full-link approximation I^R. Note that scaling the variance by C means scaling the pictures with the same number. In the top picture, the decay rates are almost indistinguishable. When plotting the difference, we see that it is more or less bounded.

For fBm, we do not calculate the exact decay rate, but rather an upper bound based on the busy period solution, i.e., $I^B(b)$. Because of the self-similarity of the busy period problem with fBm input – see [24, Prop. 4.1] – it is enough to determine $I^B(1)$. Other values are given by $I^B(b) = b^{2-2H} I^B(1)$. The lower bounds I^R and I^L are then compared to the upper bound I^B. Again, when plotting the

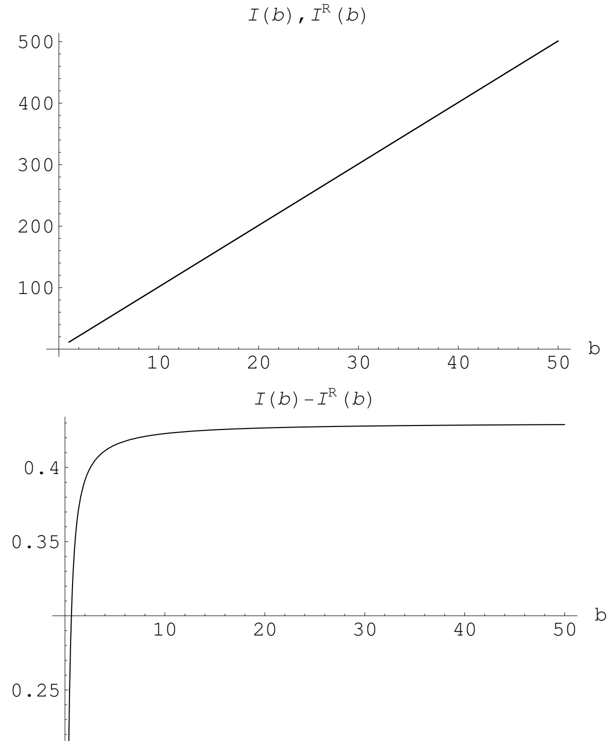

Fig. 7. Comparison of decay rates I and I^R for Ornstein-Uhlenbeck input. Parameters: $v(t) = t - 1 + e^{-t}$, $c_1 = 1$, $c_2 = 0.9$.

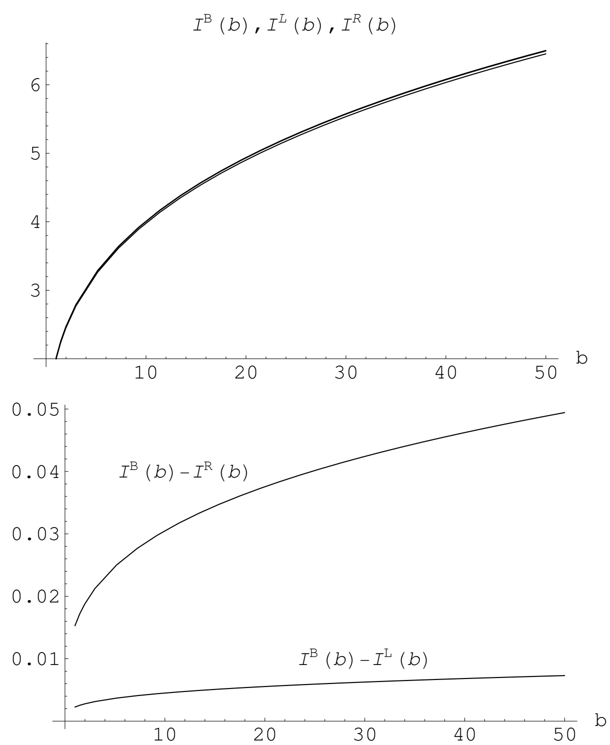

Fig. 8. Comparison of decay rates I and I^R for fBm input. Parameters: $v(t) = t^{2H}$, $H = 0.85$, $c_1 = 1$, $c_2 = 0.9$.

decay rates in the same picture, one hardly sees any difference (see Figure 8). The behavior is qualitatively the same for all $H \in (\frac{1}{2}, 1)$.

B. *Estimating the queue length distribution*

In order to check the accuracy of the performance estimate (9), we have compared it to estimates obtained by simulation. The simulation traces were generated using an extension of random midpoint displacement algorithm RMD_{mn}, see [26]. Each simulation was 2^{24} steps at the resolution 2^{-5}.

In all scenarios, $c_1 = 1$ and $c_2 = 0.9$ so that $c_1 < c^F(b)$ for almost all b. Note that if $c_1 \geq c_1^F(b)$ then the MPP-based approximations are as good as those of the single queue; see some examples in [21], [22].

We consider three different traffic models. The first one is the iOU input with scaled variance functions $v(t) = C(t - 1 + \exp(-t))$, for $C = 5, 10, 25, 50$. The results are shown in Figure 9. Performance of a tandem queue serving fBm with different Hurst parameters, i.e., $v(t) = t^{2H}$, $H = 0.6, 0.7, 0.8, 0.9$, is shown in Figure 10. Finally, Gaussian input with variance $v(t) = C((t+1)^{3/2} - \frac{3}{2}t - 1)$, $C = 1, 2, 5, 25$ is studied in Figure 11. This process is the Gaussian counterpart of the so-called M/G/∞-input with Pareto distributed session durations, see [12]. It is a smooth process (like iOU) and has long-range correlations (like fBm).

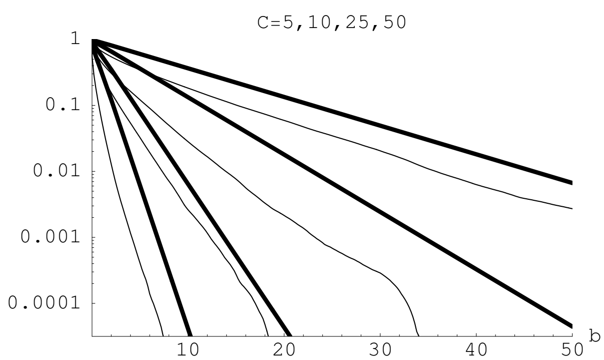

Fig. 9. Tail probabilities $\mathrm{P}(Q_2 > b)$ for a tandem queue fed by an integrated Ornstein-Uhlenbeck source: $v(t) = C(t - 1 + e^{-t})$, $c_1 = 1$, $c_2 = 0.9$. The thick lines are the approximations by (9) and the thin lines the results from the simulations.

All these simulations show that the accuracy of estimate (9) is about the same order as in single-node queues. Moreover, it seems to give an upper bound for the tail distributions. This demonstrates that it might have use in coarse performance analysis and dimensioning, although the overall accuracy is not very good. One possible improvement is to scale by the non-idle probability as in [22], i.e., $\mathrm{P}(Q_2 > b) \approx \mathrm{P}(Q_2 > 0)\exp(-I(b))$. This

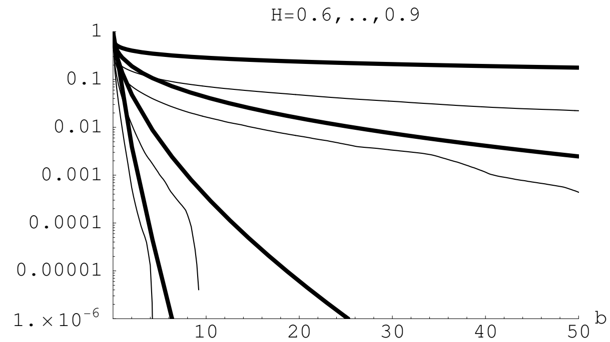

Fig. 10. Tail probabilities $\mathrm{P}(Q_2 > b)$ for a tandem queue fed by FBM source: $v(t) = t^{2H}$, $c_1 = 1$, $c_2 = 0.9$. The thick lines are the approximations by (9) and the thin lines the results from the simulations.

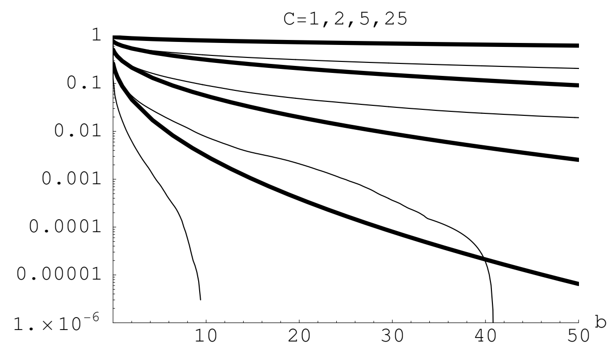

Fig. 11. Tail probabilities $\mathrm{P}(Q_2 > b)$ for a tandem queue fed by Gaussian process with $v(t) = C((t+1)^{\frac{3}{2}} - \frac{3}{2}t - 1)$: $c_1 = 1$, $c_2 = 0.9$. The thick lines are the approximations by (9) and the thin lines the results from the simulations.

would eliminate most of the gap between the estimate and the simulation results.

VI. Concluding remarks and outlook

In this paper, we have considered Gaussian tandem queues as an application of the recently obtained results on infinite intersections in Gaussian processes. In principle, the same approach could have been applied, for example, in priority and generalized processor sharing queues. We studied two 'classical' input processes in order to demonstrate the basic ideas: fractional Brownian motion and integrated Ornstein-Uhlenbeck. These processes have either zero or one derivative; the existence of higher-order derivatives would cause substantial complications.

The second topic considered was the performance measures based on the most probable paths. It turns out that the decay rate $I(b)$ is very well approximated by the lower bound $I^L(b)$ and the rough full-link approximation $I^R(b)$. These expressions can be used to generate approximations for the probability distribution (rather than its exponential decay rate). We see that, as in the single-node case, they show the correct qualitative behavior, but from a quantitative perspective there is room for improvements; scaling by the non-idle probability might yield better approximations. It would be desirable to find well-founded improved approximations based on our MPP identification, but this remains a challenge for future work.

ACKNOWLEDGMENT

This work was done while P. Mannersalo was visiting CWI as an ERCIM fellow. He is grateful for the generous financial support of CWI and Academy of Finland (#202218).

REFERENCES

[1] R.G. Addie, "On the weak convergence of long range dependent traffic processes," *J. Stat. Plann. Inference*, vol. 80, pp. 155–171, 1999.

[2] J. Kilpi and I. Norros, "Testing the Gaussian approximation of aggregate traffic," in *Proceedings of the 2nd Internet Measurement Workshop*, Marseille, France, 2002, pp. 49–61.

[3] W. Leland, M.S. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, 1994.

[4] I. Norros, "A storage model with self-similar input," *Queueing Syst.*, vol. 16, pp. 387–396, 1994.

[5] J. Hüsler and V. Piterbarg, "Extremes of a certain class of Gaussian processes," *Stochastic Processes Appl.*, vol. 83, pp. 257–271, 1999.

[6] B. Avi-Itzhak, "A sequence of service stations with arbitrary input and regular service times," *Manage. Sci.*, vol. 11, pp. 565–571, 1965.

[7] H. Friedman, "Reduction methods for tandem queueing systems," *Operations Research*, vol. 13, pp. 121–131, 1965.

[8] A. Weiss, "A new technique for analyzing large traffic systems," *Adv. Appl. Probab.*, vol. 18, pp. 506–532, 1986.

[9] D. Botvich and N. Duffield, "Large deviations, the shape of the loss curve, and economies of large scale multiplexers," *Queueing Syst.*, vol. 20, pp. 293–320, 1995.

[10] C. Courcoubetis and R. Weber, "Buffer overflow asymptotics for a buffer handling many traffic sources," *J. Appl. Probab.*, vol. 33, pp. 886–903, 1996.

[11] K. Dębicki and M. Mandjes, "Exact overflow asymptotics for queues with many Gaussian inputs," *J. Appl. Probab.*, vol. 40, pp. 702–720, 2003.

[12] M. Mandjes and M. van Uitert, "Sample-path large deviations for tandem and priority queues with Gaussian inputs," *Ann. Appl. Probab.*, 2004, To appear, available from `http://ftp.cwi.nl/CWIreports/PNA/PNA-R0221.pdf`.

[13] P. Mannersalo and I. Norros, "Approximate formulae for Gaussian priority queues," in *Proceedings of ITC 17*, Salvador, Brazil, 2001, pp. 991–1002.

[14] P. Mannersalo and I. Norros, "GPS schedulers and Gaussian traffic," in *Proceedings of IEEE Infocom*, New York, USA, 2002, pp. 1660–1667.

[15] M. Mandjes, P. Mannersalo, I. Norros, and M. van Uitert, "Large deviations of infinite interesections of events in Gaussian processes," Report PNA-E0409, CWI, 2004, Available from `http://ftp.cwi.nl/CWIreports/PNA/PNA-E0409.pdf`.

[16] R.R. Bahadur and S.L. Zabell, "Large deviations of the sample mean in general vector spaces," *Ann. Probab.*, vol. 7, no. 4, pp. 587–621, 1979.

[17] R. Azencott, *Ecole d'Eté de Probabilités de Saint-Flour VIII-1978*, vol. 774 of *Lecture Notes in Mathematics*, chapter Grandes deviations et applications, pp. 1–176, Springer, Berlin, 1980.

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

[19] L.I. Piterbarg, "On the structure of the infinitesimal σ-algebra of Gaussian processes and fields," *Theory Probab. Appl.*, vol. 31, pp. 484–492, 1987.

[20] M. Mandjes and M. van Uitert, "Sample-path large deviations for tandem queues with Gaussian inputs," in *Proceedings of ITC 18*, Berlin, Germany, 2003, pp. 521–530.

[21] R.G. Addie, P. Mannersalo, and I. Norros, "Performance formulae for queues with Gaussian input," in *Proceedings of ITC 16*, Edinburgh, UK, 1999, pp. 1169–1178.

[22] R.G. Addie, P. Mannersalo, and I. Norros, "Most probable paths and performance formulae for buffers with Gaussian input traffic," *Eur. Trans. Telecommun.*, vol. 13, no. 3, pp. 183–196, 2002.

[23] I. Norros, E. Valkeila, and J. Virtamo, "An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions," *Bernoulli*, vol. 5, pp. 571–587, 1999.

[24] I. Norros, "Busy periods of fractional Brownian storage: a large deviations approach," *Adv. Perf. Anal.*, vol. 2, no. 1, pp. 1–20, 1999.

[25] G. Gripenberg and I. Norros, "On the prediction of fractional Brownian motion," *Journal of Applied Probability*, vol. 33, pp. 400–410, 1996.

[26] I. Norros, P. Mannersalo, and J.L. Wang, "Simulation of fractional Brownian motion with conditionalized random midpoint displacement," *Adv. Perf. Anal.*, vol. 2, no. 1, pp. 77–101, 1999.