Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA

Traffic with an FBM limit: convergence of the workload process

K.G. Dębicki, M.R.H. Mandjes

REPORT PNA-R0220 DECEMBER 31, 2002

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

Traffic with an FBM Limit: Convergence of the Workload Process

Krzystof Dębicki[*,**] and Michel Mandjes [*,†]

[*]*CWI*
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

[**]*Mathematical Institute, University of Wrocław*
pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland

[†]*Faculty of Mathematical Sciences,University of Twente*
P.O. Box 217, 7500 AE Enschede, The Netherlands

ABSTRACT
Highly-aggregated traffic in communication networks is often modeled as *fractional Brownian motion* (fBm). This is justified by the theoretical result that the sum of a large number of on-off inputs, with either on-times or off-times having a heavy-tailed distribution with infinite variance, converges to fBm, after rescaling time appropriately. For performance analysis purposes, the key question is whether this convergence carries over to the stationary buffer content process. In this paper it is shown that, in a heavy-traffic queueing environment, this property indeed holds.

2000 Mathematics Subject Classification: 60K25 (primary), 68M20, 90B22, 60G15 (secondary).
Keywords & Phrases: On-off sources, Gaussian processes, fractional Brownian motion, heavy-traffic environment, weak convergence.
Note: Work carried out under the CWI project P1201. K. Dębicki was also supported by KBN under grant 5 P03A 021 20 (2001-2003).

1 Introduction

Recently the statistical modeling of broadband communication networks has become a vivid area of research. Traditionally one assumed either the absence of any significant correlation between consecutive packet arrivals ('renewal input', for instance a Poisson processes), or just a mild form of dependence (for instance Markov modulated Poisson processes). The discovery of significant correlations on a broad range of time-scales, as exhibited in many measurement studies during the 1990s, however, led to the examination of different classes of traffic models. Many models have been proposed for describing this *long-range dependence*; we mention here (i) (the superposition of) *on-off sources*, with heavy-tailed on- or off-times, (ii) the *infinite-source Poisson model*, often referred to as the M/G/∞ input model, in which heavy-tailed jobs arrive according to a Poisson process, and transmit their job at a fixed rate, (iii) *fractional Brownian motion* (fBm).

A key result was derived by Taqqu, Willinger, and Sherman [14], who show that a superposition of on-off sources, whose on- or off-periods have infinite variance, converges in distribution to fBm. Essentially, in this convergence *two* limits are involved: first the number of sources (say N) grows large, whereas the 'outer limit' corresponds to zooming out the time axis (i.e., time is scaled by a factor T, meaning $t \to tT$, and then $T \to \infty$). A similar result holds if the job arrival rate in the infinite-source Poisson model grows large. The analysis in [14] shows that these limit results should be treated with care: when the limits are taken in reverse order, the traffic process converges to an α-stable Lévy motion, rather than fBm. A recent article by Mikosch, Resnick, Rootzén, and Stegeman [10] explores how these two limits relate; in fact it is proven that under *slow growth* (connection rates are modest relative to the connection length distribution tails) the limit is an α-stable Lévy motion, whereas under *fast growth* (the opposite situation) there is an fBm limit.

The discovery of the phenomenon of long-range dependence in broadband network traffic led to the analysis of queueing systems with input traffic models (i), (ii), and (iii). Particular attention was paid to the *tail behavior* of the stationary queue length distribution. For instance for fBm input, characterized by Hurst parameter H, this decay is *Weibullian* [11], i.e., the probability of exceeding buffer level B is roughly of the form $\exp(-B^{2-2H})$. The results mentioned above suggest the following procedure for predicting the performance experienced by a superposition of heavy-tailed on-off sources multiplexed at a buffered resource. First it is noted that the aggregate input process weakly converges in $C([0,\infty))$ to fBm (with a Hurst parameter that follows from the specific on- and off-time distributions). Subsequently the formulae for queues with fBm input are used to evaluate the queuing performance. However, this two stage procedure ignores one essential step: *does the workload distribution of a stationary queue with many heavy-tailed on-off inputs converge to the stationary buffer content distribution of a queue with fBm input?* It is this relation that is established in the present paper.

It is emphasized that this question is not trivial, since the stationary workload has the representation as a supremum of the input traffic. More precisely, the well-known Reich

formula states that the stationary workload of a queue with service rate c is distributed as $\sup_{t>0} A(t)-ct$, where $A(t)$ denotes the amount of traffic accessing the system in $[-t, 0]$. As the supremum operator is not continuous in $C([0,\infty))$, the continuous mapping theorem cannot be applied, and hence it is not obvious that the convergence of the input process carries over to the workload.
Whereas many papers deal with the convergence of arrival processes to some limit, considerably less attention is paid to the relevant question if this convergence carries over to the buffer content distribution. In this respect we mention the contribution by Kulkarni and Rolski [8]: under a specific parameterization the queue fed by N exponential on-off sources (the so-called Anick-Mitra-Sondhi model) converges to a queue with Gaussian input (with an Ornstein-Uhlenbeck type variance structure), as N grows large. This result was generalized by Dębicki and Palmowski [6] to general on-off sources. Another result along the same lines is [5]: motivated by the weak convergence of the superposition of many i.i.d. stationary point processes to a Poisson process, it is verified that this convergence carries over to the workload process.
In the present work we primarily consider the situation in which both N and T go to infinity, similarly to [14]. In addition to this we study the 'short time-scale regime' $T \downarrow 0$, $N \to \infty$. It appears that in this case, although we have the convergence of arrival processes, the convergence does *not* carry over to the buffer content process. The latter result illustrates how delicate this type of convergence properties are.

This paper is organized as follows. In *Section 2* the model and some preliminaries are described. *Section 3* contains the results on the large time-scale regime, whereas *Section 4* covers the short time-scale. Proofs are provided in *Section 5.*

2 Model and preliminaries

I. TRAFFIC SOURCES. We consider N statistically identical, independent fluid sources. The traffic rate of each source alternates between on and off; during the on-times traffic is generated continuously at a (normalized) peak rate of 1. The activity periods constitute an i.i.d. sequence of random variables, each of them distributed as a random variable $T_{\rm on}$ with values in $\mathbb{R}_+$. The silence periods are also an i.i.d. sequence, distributed as a random variable $T_{\rm off}$ with values in $\mathbb{R}_+$. In addition, both sequences are mutually independent.
Let $T_{\rm on}$ and $T_{\rm off}$ have distribution functions $F_{\rm on}(\cdot)$ and $F_{\rm off}(\cdot)$, respectively, and complementary distribution functions $\bar{F}_{\rm on}(\cdot)$ and $\bar{F}_{\rm off}(\cdot)$. Assuming that $\mu_{\rm on} := \mathbb{E}(T_{\rm on}) < \infty$ and $\mu_{\rm off} := \mathbb{E}(T_{\rm off}) < \infty$, we have that the mean input rate equals $N\mu$, with $\mu := \mu_{\rm on}/(\mu_{\rm on} + \mu_{\rm off})$. In our analysis we impose the following two assumptions on the on- and off-times.

Assumption 2.1 *The complementary distribution functions are regularly varying, and* *the on- and off-times have infinite variance, cf.* [7], [10]:

1. $\bar{F}_{\mathrm{on}}(x) = x^{-\alpha_{\mathrm{on}}} L_{\mathrm{on}}(x)$, *with* $\alpha_{\mathrm{on}} \in (1,2)$;
2. $\bar{F}_{\mathrm{off}}(x) = x^{-\alpha_{\mathrm{off}}} L_{\mathrm{off}}(x)$, *with* $\alpha_{\mathrm{off}} \in (1,2)$;
3. $\bar{F}_{\mathrm{on}}(x) = o(\bar{F}_{\mathrm{off}}(x))$ *as* $x \to \infty$ *or* $\bar{F}_{\mathrm{off}}(x) = o(\bar{F}_{\mathrm{on}}(x))$ *as* $x \to \infty$,

where $L_{\mathrm{on}}(\cdot)$ *and* $L_{\mathrm{off}}(\cdot)$ *are slowly varying at infinity.*

Assumption 2.2 *Additional regularity properties, cf.* [6]: $F_{\mathrm{on}}(\cdot)$ *and* $F_{\mathrm{off}}(\cdot)$ *are absolutely continuous with densities* $f_{\mathrm{on}}(\cdot)$ *and* $f_{\mathrm{off}}(\cdot)$, *respectively, such that both* $\lim_{t\to 0^+} f_{\mathrm{on}}(t) < \infty$ *and* $\lim_{t\to 0^+} f_{\mathrm{off}}(t) < \infty$.

Let $\{\xi_i(t), t \in \mathbb{R}\}$, for $i = 1, \ldots, N$, denote the traffic rate of the ith source, i.e., $\xi_i(t) = 1$ (0) if source i is on (off) at time t. Let $R(\cdot)$ denotes the covariance function of a generic source, i.e., $R(s) := \mathbb{C}\mathrm{ov}(\xi_1(s+t), \xi_1(t))$.
We write $\lim_{N\to\infty}\{Y_N(t), t \in T\} =_{\mathrm{d}} \{Y(t), t \in T\}$ in order to denote that sequence of stochastic processes $\{Y_N(t), t \in T\}$ weakly converges in $C(T)$ to $\{Y(t), t \in T\}$ as $N \to \infty$. Moreover, by $\lim_{N\to\infty} Y_N =_{\mathrm{d}} Y$, we mean that sequence of random variables Y_N converges in distribution to random variable Y as $N \to \infty$.

II. Fluid queue. We now introduce the *fluid queue* driven by a stationary ergodic process $\{Z(t), t \in \mathbb{R}\}$ as follows. Consider a stochastic process $\{X(t), t \in \mathbb{R}\}$ whose dynamics can be described by

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \begin{cases} Z(t) - c & X(t) > 0 \\ (Z(t) - c)_+ & X(t) = 0 \end{cases},$$

where $c > 0$ is a constant. Then $\{X(t) : t \in \mathbb{R}\}$ represents the buffer content process of a fluid queue fed with the input rate $Z(t)$ and drained with constant rate c.
Now introduce the aggregate traffic rate at time Ts of the N multiplexed sources, and the traffic generated in the interval $[sT, tT]$, with $s \le t$:

$$Z_{T,N}(s) := \sum_{i=1}^{N} \xi_i(Ts), \quad A_{T,N}(s,t) := \int_s^t Z_{T,N}(u)\mathrm{d}u.$$

Since the process $\{A_{1,1}(0,t), t \in \mathbb{R}\}$ has stationary increments, it follows that

$$\sigma(t) := \sqrt{\mathbb{V}\mathrm{ar}\ (A_{1,1}(0,t))} = \sqrt{2\int_0^t \int_0^s R(u)\mathrm{d}u\mathrm{d}s}. \tag{2.1}$$

We also define the fluid queue $\{X_{T,N}(t), t \in \mathbb{R}\}$ governed by

$$\frac{\mathrm{d}X_{T,N}(t)}{\mathrm{d}t} = \begin{cases} r(T,N)Z_{T,N}(t) - c(T,N) & X_{T,N}(t) > 0; \\ (r(T,N)Z_{T,N}(t) - c(T,N))_+ & X_{T,N}(t) = 0, \end{cases} \tag{2.2}$$

with, for a given constant $c > 0$,

$$r(T,N) := \frac{T}{\sigma(T)\sqrt{N}}, \quad c(T,N) := c + \mu\frac{T\sqrt{N}}{\sigma(T)}. \tag{2.3}$$

III. Stationary workload distribution. Notice that our parameterization (2.3) enforces stability of the queueing system:

$$r(T,N)\cdot \mathbb{E}Z_{T,N}(t) = \mu\frac{T\sqrt{N}}{\sigma(T)} < c(T,N).$$

Thus, following Theorem 13 in Borovkov [4], there exists the *stationary* solution $X^\star_{T,N}(t)$ of the workload (or: *buffer content*) process of the fluid queue (2.2), which obeys the representation

$$\{X^\star_{T,N}(t), t\geq 0\} =_{\rm d} \left\{\sup_{s\leq t}\left(r(T,N)A_{T,N}(s,t) - c(T,N)(t-s)\right), t\geq 0\right\},$$

where $\mathbb{P}(X^\star_{T,N}(t) > x) = \lim_{s\to\infty}\mathbb{P}(X_{T,N}(s) > x)$ for all $t\geq 0$ and $x\geq 0$.

In the further analysis we will apply some known properties of $\{X^\star_{T,N}(t), t\geq 0\}$, in the regime $N\to\infty$. Define $\{Z^\star(t), t\in\mathbb{R}\}$ as a stationary centered Gaussian process having the same covariance structure as a single on-off source, i.e., $\mathbb{C}{\rm ov}(Z^\star(s+t), Z^\star(t)) = R(s)$. Furthermore, we denote by $Z^\star_T(s)$ a stationary centered Gaussian process such that $\{Z^\star_T(s), s\in\mathbb{R}\} =_{\rm d} \{Z^\star(Ts), s\in\mathbb{R}\}$. Observe that $R_T(s) := \mathbb{C}{\rm ov}(Z^\star_T(s+t), Z^\star_T(t)) = R(Ts)$. Again, we can define the corresponding fluid queue through

$$\frac{{\rm d}X_T(t)}{{\rm d}t} = \begin{cases} r_T Z^\star_T(t) - c & X_T(t) > 0;\\ (r_T Z^\star_T(t) - c)_+ & X_T(t) = 0,\end{cases} \tag{2.4}$$

where $r_T := T/\sigma(T)$.
Writing $A_T(s,t) := \int_s^t Z^\star_T(u){\rm d}u$, we define the stationary buffer content process:

$$\{X^\star_T(t), t\geq 0\} =_{\rm d} \left\{\sup_{s\leq t}\left(r_T A_T(s,t) - c(t-s)\right),\ t\geq 0\right\}.$$

The next proposition follows directly from Theorem 3.1 and Corollary 3.1 in [6].

Proposition 2.1 *If Assumption 2.2 is satisfied, then, for any $T > 0$,*

(i) $\lim_{N\to\infty}\left\{\frac{Z_{T,N}(s) - \mu N}{\sqrt{N}}, s\geq 0\right\} =_{\rm d} \{Z^\star_T(s), s\geq 0\}$;

(ii) $\lim_{N\to\infty}\{X^\star_{T,N}(t), t\geq 0\} =_{\rm d} \{X^\star_T(t), t\geq 0\}$.

3 Large time-scale regime

Taqqu, Willinger, and Sherman [14] have shown convergence of the arrival process – in a limiting regime that is made precise in Proposition 3.1 – to fBm. This section deals with the convergence – in the same limiting regime, and parameterization (2.3) – of the corresponding workload process to a queue fed by fBm (i.e., *reflected* fractional Brownian motion).
Reflected fBm is defined as

$$Y^\star_H(0) := \sup_{s\leq 0}(B_H(s) + cs) =_{\rm d} \sup_{s\geq 0}(B_H(s) - cs)$$

for some positive $c > 0$. Here $\{B_H(t), t \in \mathbb{R}\}$ is fBm with Hurst parameter $H \in (1/2, 1]$, or, more explicitly, a centered Gaussian process with stationary increments, continuous sample paths a.s., $B_H(0) = 0$ and variance $\mathbb{V}\mathrm{ar}\,(B_H(t)) = |t|^{2H}$. The stationary reflected fBm process $\{Y_H^\star(t), t \geq 0\}$, is the process

$$\{Y_H^\star(t), t \geq 0\} = \left\{ \sup_{s \leq t} \left(B_H(t) - B_H(s) - c(t-s)\right), t \geq 0 \right\}.$$

In the sequel we use the following identification:

$$H \equiv (3 - \min(\alpha_{\mathrm{on}}, \alpha_{\mathrm{off}}))/2, \tag{3.5}$$

Theorem 1 of Taqqu, Willinger & Sherman [14] (see also Theorem 7.2.5 of Whitt [15] for the tightness arguments needed to prove weak convergence) implies the following proposition, justifying the use of reflected fBm.

Proposition 3.1 *If Assumptions 2.1 and 2.2 hold, then, under parameterization* (2.3),

$$\lim_{T\to\infty} \lim_{N\to\infty} \{r(T,N)A_{T,N}(0,t) - c(T,N)t, t \geq 0\} =_{\mathrm{d}} \{B_H(t) - ct, t \geq 0\}.$$

Remark 3.1 *Proposition 3.1 and the stationarity of the increments of $A_{T,N}(0,t)$ implies directly that, for any $S > 0$,*

$$\lim_{T\to\infty} \lim_{N\to\infty} \{r(T,N)A_{T,N}(t,S) - c(T,N)(S-t), t \leq S\}$$
$$=_{\mathrm{d}} \{B_H(S) - B_H(t) - c(S-t), t \leq S\}. \tag{3.6}$$

Proposition 3.1 implies that several functionals of $A_{T,N}(\cdot)$ may be approximated by the appropriate functionals of $B_H(\cdot)$. In particular, assuming that $X_{T,N}(0) = 0$, the *transient* buffer content $X_{T,N}(t)$ at time t satisfies

$$\lim_{T\to\infty} \lim_{N\to\infty} X_{T,N}(t) =_{\mathrm{d}} \sup_{s\in[0,t]} (B_H(s) - cs). \tag{3.7}$$

The limit (3.7) does not imply directly the convergence of the *stationary* buffer content process $\{X_{T,N}^\star(t), t \geq 0\}$ to reflected fBm, since the sup functional is not continuous in $C([0,\infty))$. This weak convergence, however, is valid, and is the main result of this section:

Theorem 3.2 *If Assumptions 2.1 and 2.2 hold, then, under parameterization* (2.3),

$$\lim_{T\to\infty} \lim_{N\to\infty} \{X_{T,N}^\star(t), t \geq 0\} =_{\mathrm{d}} \{Y_H^\star(t), t \geq 0\}.$$

Proof. The detailed proof is given in Section 5.2. □

An important quantity in performance evaluation of fluid models is the steady state buffer content. Let $X_{T,N}^\star(0)$ be the steady state buffer content for the $\{X_{T,N}(t), t \geq 0\}$ fluid model. The following corollary gives the theoretical justification of approximating the stationary buffer workload $X_{T,N}^\star(0)$ by reflected fBm, as heuristically proposed in [11].

Corollary 3.3 *If Assumptions 2.1 and 2.2 hold, then, under parameterization* (2.3),

$$\lim_{T\to\infty} \lim_{N\to\infty} X_{T,N}^\star(0) =_{\mathrm{d}} \sup_{s\geq 0}(B_H(s) - cs) =: Y_H^\star(0).$$

Proof. The proof immediately follows from Theorem 3.2, applying that $\{B_H(t), t \in \mathbb{R}\}$ is time-reversible. □

4 Small time-scale regime

In Section 3 we analyzed the asymptotics of $\{X^\star_{T,N}(t), t \geq 0\}$ in the *large time-scale regime*, i.e., $T \to \infty$. In this section we focus on the opposite scenario, that is, $T \downarrow 0$. The main point of this section is that the convergence of the input process does *not* carry over to the buffer content process. We begin by stating the weak convergence of the input process.

Theorem 4.1 *If Assumptions 2.1 and 2.2 hold, then, under parameterization* (2.3)*,*

$$\lim_{T\downarrow 0}\lim_{N\to\infty}\{r(T,N)A_{T,N}(0,t) - c(T,N)t, t\geq 0\} =_{\rm d} \{B_1(t) - ct, t \geq 0\}.$$

Proof. A complete proof is presented in Section 5.3. □

Despite the convergence result given in Theorem 4.1, it appears that the stationary workload process $\{X^\star_{T,N}(t), t \geq 0\}$ does *not* converge to the appropriate reflected fBm. Moreover, we do not even have 1-dimensional convergence, as stated in the following theorem.

Theorem 4.2 *If Assumptions 2.1 and 2.2 hold, then, under parameterization* (2.3)*,*

$$\lim_{T\downarrow 0}\lim_{N\to\infty} X^\star_{T,N}(0) \neq_{\rm d} Y^\star_1(0).$$

Proof. A complete proof is presented in Section 5.3. □

Remark 4.1 *From the proof of Theorem 4.2 it follows that*

$$\lim_{T\downarrow 0}\lim_{N\to\infty} X^\star_{T,N}(0) =_{\rm d} \sup_{t\geq 0}\left(\int_0^t Z^\star(s){\rm d}s - c\sqrt{R(0)}t\right),$$

where $\{Z^\star(t), t \geq 0\}$ is a stationary centered Gaussian process with the same covariance function $R(t)$ as the covariance function of the generic on-off process $\{\xi(t), t \geq 0\}$.

5 Proofs

This section presents the proofs of the results of the previous sections. Before doing so, we first establish some lemmas that are also of independent interest.

5.1 Preliminary lemmas

Lemma 5.1 *If Assumptions 2.1 and 2.2 hold, then*

(i) $\lim_{t\downarrow 0}\dfrac{\sigma^2(t)}{t^2} = R(0)$;

(ii) $\sigma^2(t)$ *is ultimately convex;*

(iii) $\sigma^2(t)$ *is regularly varying at* ∞ *with index* $2H$.

Proof. Note that, due to Assumption 2.2, the on-off processes $\{\xi_i(t), t \geq 0\}$ are stochastically continuous, see, e.g., Theorem 2 in Szczotka [13]. This implies that $R(\cdot)$ is continuous at 0. Result (i) is now a consequence of representation (2.1).
In order to prove (ii) it suffices to combine $(\sigma^2)''(t) = 2R(t)$ with

$$R(t) \sim L(t)t^{2H-2} \quad \text{as} \quad t \to \infty, \tag{5.8}$$

where $L(t) > 0$ is a slowly varying function at ∞. The latter statement is due to Assumption 2.1, see e.g. [7].
The proof of (iii) follows from (2.1) and Karamata's theorem applied twice to (5.8). □

Lemma 5.2 *Let $\{X(t), t \in [x, \infty)\}$ be a centered Gaussian process with stationary increments and a* convex *variance function* $\mathbb{V}\mathrm{ar}\ X(t)$*. Then, for each $u > 0$,*

$$\mathbb{P}\left(\sup_{t>x} X(t) - ct > u\right) \leq \mathbb{P}\left(\sup_{t>x} B_{\frac{1}{2}}(\mathbb{V}\mathrm{ar}\ (X(t))) - ct > u\right).$$

Proof. For any convex $f(\cdot)$, it holds that both (assume without loss of generality $s \leq t/2$) $f(t-s) + f(0) \geq 2f(s)$ and $f(s) + f(t) \geq 2f(t-s)$. If in addition $f(0) = 0$, this yields $f(t) - f(s) \geq f(t-s)$. Hence

$$|\mathbb{V}\mathrm{ar}\ X(s) - \mathbb{V}\mathrm{ar}\ X(t)| \geq \mathbb{V}\mathrm{ar}\ X(|s-t|)$$

for each s, t. This implies

$$\begin{aligned} \mathbb{C}\mathrm{ov}(X(s), X(t)) &= \frac{\mathbb{V}\mathrm{ar}\ X(s) + \mathbb{V}\mathrm{ar}\ X(t) - \mathbb{V}\mathrm{ar}\ X(|s-t|)}{2} \qquad (5.9)\\ &\geq \frac{\mathbb{V}\mathrm{ar}\ X(s) + \mathbb{V}\mathrm{ar}\ X(t) - |\mathbb{V}\mathrm{ar}\ X(s) - \mathbb{V}\mathrm{ar}\ X(t)|}{2} \\ &= \min\{\mathbb{V}\mathrm{ar}\ X(s), \mathbb{V}\mathrm{ar}\ X(t)\} \\ &= \mathbb{C}\mathrm{ov}\left(B_{\frac{1}{2}}(\mathbb{V}\mathrm{ar}\ X(s)), B_{\frac{1}{2}}(\mathbb{V}\mathrm{ar}\ X(t)\right), \end{aligned}$$

where (5.9) follows from the stationarity of the increments of $\{X(t), t \in \mathbb{R}\}$. Hence, in order to complete the proof, it is enough to apply Slepian's inequality – see e.g. Theorem C.1 in [12]. □

Lemma 5.3 *For each $x > 0$ there exists a T_0 such that for each $T \geq T_0$ it holds that $r_T^2 \mathbb{V}\mathrm{ar}\ A_T(0, s)$ is* convex *for $s \in [x, \infty)$.*

Proof. The proof follows immediately from part (ii) of Lemma 5.1, in conjunction with the fact that $r_T^2 \mathbb{V}\mathrm{ar}\ A_T(0, s) = \sigma^2(Ts)/\sigma^2(T)$. □

5.2 Proof of the large time-scale results

The proof of the main result of Section 3 is given in three steps: convergence of the stationary buffer content, convergence in finite-dimensional distributions, and the weak convergence result.

Step 1: Convergence of stationary buffer content. Since $X_{T,N}(t)$ is stationary, then it is enough to show that

$$\lim_{T\to\infty}\lim_{N\to\infty} X^{\star}_{T,N}(0) =_{\mathrm{d}} Y^{\star}_{H}(0).$$

Proposition 2.1, part (ii), states that for each $T > 0$, $\lim_{N\to\infty} X^{\star}_{T,N}(0) =_{\mathrm{d}} X^{\star}_{T}(0)$ and hence it is left to prove that

$$\lim_{T\to\infty} X^{\star}_{T}(0) =_{\mathrm{d}} Y^{\star}_{H}(0). \tag{5.10}$$

Note that, due to the reversibility of Gaussian processes, (5.10) is equivalent to

$$\lim_{T\to\infty} \mathbb{P}\left(\sup_{s\geq 0} r_T A_T(0,s) - cs > u\right) = \mathbb{P}\left(\sup_{s\geq 0} B_H(s) - cs > u\right) \tag{5.11}$$

for each $u \in \mathbb{R}$. Since the functional sup is continuous in the space $C([0,S])$ for *finite* S, Proposition 3.1, combined with the continuous mapping theorem, implies that for any $u \in \mathbb{R}$, $S \in (0,\infty)$

$$\lim_{T\to\infty} \mathbb{P}\left(\sup_{s\in[0,S]} r_T A_T(0,s) - cs > u\right) = \mathbb{P}\left(\sup_{s\in[0,S]} B_H(s) - cs > u\right).$$

Hence, a sufficient condition for (5.10) is that, for any $u \in \mathbb{R}$,

$$\mathbb{P}\left(\sup_{s>S} r_T A_T(0,s) - cs > u\right) \to 0, \quad \text{as} \quad S \to \infty$$

uniformly in T. According to Lemma 5.3, for T sufficiently large, $\mathbb{V}\mathrm{ar}\,(r_T A_T(0,s))$ is convex for $s \in [S,\infty)$; $S > 0$. Consequently, from Lemma 5.2 we infer that, for $u \in \mathbb{R}$,

$$\mathbb{P}\left(\sup_{s>S} r_T A_T(0,s) - cs > u\right) \leq \mathbb{P}\left(\sup_{s>S} B_{\frac{1}{2}}\left(\mathbb{V}\mathrm{ar}\,(r_T A_T(0,s))\right) - cs > u\right).$$

But, recalling that $\mathbb{V}\mathrm{ar}\,(r_T A_T(0,s)) = \sigma^2(Ts)/\sigma^2(T)$,

$$\mathbb{P}\left(\sup_{s>S} B_{\frac{1}{2}}(r_T^2 \mathbb{V}\mathrm{ar}\,(A_T(0,s))) - cs > u\right) = \mathbb{P}\left(\sup_{s>S} B_{\frac{1}{2}}(\sigma^2(Ts)/\sigma^2(T)) - cs > u\right)$$

$$= \mathbb{P}\left(\sup_{s>\frac{\sigma^2(TS)}{\sigma^2(T)}} B_{\frac{1}{2}}(s) - c\frac{(\sigma^2)^{-1}(s\sigma^2(T))}{T} > u\right)$$

$$= \mathbb{P}\left(\sup_{s>\frac{\sigma^2(TS)}{\sigma^2(T)}} B_{\frac{1}{2}}(s) - c\frac{(\sigma^2)^{-1}(s\sigma^2(T))}{(\sigma^2)^{-1}(\sigma^2(T))} > u\right).$$

Now, due to (iii) of Lemma 5.1, and Potter's theorem (see e.g. Theorem 1.5.6 in [3], or Proposition 2 in the Appendix of [10]), for sufficiently large S and T

$$\frac{\sigma^2(TS)}{\sigma^2(T)} \geq \frac{1}{2} S^{2H-\epsilon},$$

where $\epsilon < 2H$. Analogously, due to Potter's theorem, for sufficiently large T and s,

$$\frac{(\sigma^2)^{-1}(s\sigma^2(T))}{(\sigma^2)^{-1}(\sigma^2(T))} \geq \frac{1}{2} s^{1/(2H)-\kappa},$$

where $\kappa < 1/(2H) - 1/2$. Hence we have uniformly in T:

$$\mathbb{P}\left(\sup_{s > \frac{\sigma^2(TS)}{\sigma^2(T)}} B_{\frac{1}{2}}(s) - c\frac{(\sigma^2)^{-1}(s\sigma^2(T))}{(\sigma^2)^{-1}(\sigma^2(T))} > u \right) \leq \mathbb{P}\left(\sup_{s > \frac{1}{2}S^{2H-\epsilon}} B_{\frac{1}{2}}(s) - \frac{c}{2}s^{1/(2H)-\kappa} > u \right).$$

The right hand side of the previous display vanishes for any $u \in \mathbb{R}$ as $S \to \infty$, cf. the law of the iterated logarithm for Brownian motion; since $1/(2H) - \kappa > 1/2$. This completes the proof of the convergence of 1-dimensional distributions.

Step 2: Convergence of finite-dimensional distributions. The argumentation of this step is analogous to Step 1. First note that for any $t_1, ..., t_n \geq 0$, $u_1, ..., u_n \in \mathbb{R}$, $n \in \mathbb{N}$, and $x_1 < t_1, ..., x_n < t_n$ it holds that

$$\begin{aligned}
&\mathbb{P}(X_T^\star(t_1) > u_1, ..., X_T^\star(t_n) > u_n) = \\
&= \mathbb{P}\left(\sup_{t \leq t_1} r_T A_T(t, t_1) - c(t_1 - t) > u_1, \ldots, \sup_{t \leq t_n} r_T A_T(t, t_n) - c(t_n - t) > u_n \right) \\
&\leq \mathbb{P}\left(\sup_{t \in [x_i, t_i]} r_T A_T(t, t_i) - c(t_i - t) > u_i, i = 1, \ldots, n \right) + \\
&\quad + \sum_{i=1}^n \mathbb{P}\left(\sup_{t < x_i} r_T A_T(t, t_i) - c(t_i - t) > u_i \right) \\
&= \mathbb{P}\left(\sup_{t \in [x_i, t_i]} r_T A_T(t, t_i) - c(t_i - t) > u_i, i = 1, \ldots, n \right) \\
&\quad + \sum_{i=1}^n \mathbb{P}\left(\sup_{t > t_i - x_i} r_T A_T(0, t) - ct > u_i \right).
\end{aligned}$$

Now the same procedure can be followed as in Step 1.

Step 3: Tightness. Due to Theorem 8.3 in Billingsley [2] and Lindvall [9], to show tightness of $\{X_T^\star(t), t \geq 0\}$, it suffices to check that

(i) for each η, there exists an a such that $\mathbb{P}(|X_T^\star(0)| > a) \leq \eta$ for all $T \geq 1$;

(ii) for each $\varepsilon, \eta > 0$, there exists $\delta \in (0, 1)$ and T_0 such that

$$\mathbb{P}\left(\sup_{t \leq s \leq t+\delta} |X_T^\star(s) - X_T^\star(t)| \geq \varepsilon \right) \leq \eta\delta \tag{5.12}$$

for all $T \geq T_0$ and $t \geq 0$.

Notice that (i) immediately follows from the convergence $\lim_{T\to\infty} X_T^\star(0) =_{\mathrm{d}} Y_H^\star(0)$, see Step 1. We therefore concentrate on (ii). We first observe that the stationarity of $\{X_T^\star(t), t \geq 0\}$ implies that

$$X_T^\star(s) - X_T^\star(t) =_{\mathrm{d}} X_T^\star(s - t) - X_T^\star(0);$$

hence it suffices to prove (5.12) for $t = 0$. Defining

$$\bar{Z}_T(u) := r_T Z_T^\star(u) - c, \quad \text{and} \quad W_T(s, t) := \int_s^t \bar{Z}_T(u)\mathrm{d}u,$$

with straightforward arguments we derive

$$
\begin{aligned}
\sup_{s\in[0,\delta]} |X_T^\star(s) - X_T^\star(0)| &= \sup_{s\in[0,\delta]} \left| \sup_{v\in(-\infty,s]} W_T(v,s) - \sup_{v\in(-\infty,0]} W_T(v,0) \right| \\
&= \sup_{s\in[0,\delta]} \left| \sup_{v\in(-\infty,s]} W_T(v,s) - \sup_{v\in(-\infty,0]} W_T(v,s) + W_T(0,s) \right| \\
&\le \sup_{s\in[0,\delta]} \left| \max\left\{ \sup_{v\in(-\infty,0]} W_T(v,s),\ \sup_{v\in(0,s]} W_T(v,s) \right\} - \sup_{v\in(-\infty,0]} W_T(v,s) \right| \\
&\qquad + \sup_{s\in[0,\delta]} |W_T(0,s)| \\
&= \sup_{s\in[0,\delta]} \left| \max\left\{ 0,\ \sup_{v\in(0,s]} W_T(v,s) - \sup_{v\in(-\infty,0]} W_T(v,0) - W_T(0,s) \right\} \right| \\
&\qquad + \sup_{s\in[0,\delta]} |W_T(0,s)| \\
&\le \sup_{s\in[0,\delta]} \left| \sup_{v\in(0,s]} W_T(v,s) - W_T(0,s) \right| + \sup_{s\in[0,\delta]} |W_T(0,s)| \qquad (5.13) \\
&= \sup_{s\in[0,\delta]} \left| - \sup_{v\in(0,s]} W_T(v,s) \right| + \sup_{s\in[0,\delta]} |W_T(0,s)| . \qquad (5.14)
\end{aligned}
$$

where (5.13) is a consequence of the fact that both

$$
\sup_{v\in(0,s]} \int_v^s \bar{Z}_T(u)\mathrm{d}u - \int_0^s \bar{Z}_T(u)\mathrm{d}u \quad \text{and} \quad \sup_{v\in(-\infty,0]} \int_v^0 \bar{Z}_T(u)\mathrm{d}u,
$$

are non negative a.s. Expression (5.14) is now majorized by

$$
\sup_{s\in[0,\delta]} \sup_{v\in(0,s]} \left| \int_0^v \bar{Z}_T(u)\mathrm{d}u \right| + \sup_{s\in[0,\delta]} \left| \int_0^s \bar{Z}_T(u)\mathrm{d}u \right| = 2 \sup_{s\in[0,\delta]} \left| \int_0^s \bar{Z}_T(u)\mathrm{d}u \right| . \qquad (5.15)
$$

The upper bound (5.15) implies that

$$
\mathbb{P}\left(\sup_{0\le s\le\delta} |X_T^\star(s) - X_T^\star(0)| \ge \varepsilon \right) \le \mathbb{P}\left(\sup_{s\in[0,\delta]} \left| \int_0^s (r_T Z_T^\star(u) - c)\mathrm{d}u \right| \ge \varepsilon/2 \right) .
$$

From Proposition 3.1 we infer that

$$
\lim_{T\to\infty} \mathbb{P}\left(\sup_{s\in[0,\delta]} \left| \int_0^s (r_T Z_T^\star(u) - c)\mathrm{d}u \right| \ge \varepsilon/2 \right) = \mathbb{P}\left(\sup_{s\in[0,\delta]} |B_H(s) - cs| \ge \varepsilon/2 \right)
$$

and hence it suffices to prove that

$$
\lim_{\delta\downarrow 0} \delta^{-1} \cdot \mathbb{P}\left(\sup_{s\in[0,\delta]} |B_H(s) - cs| \ge \varepsilon/2 \right) = 0, \qquad (5.16)
$$

but this statement immediately follows from

$$
\mathbb{P}\left(\sup_{s\in[0,\delta]} |B_H(s) - cs| \ge \varepsilon/2 \right) \quad \le \quad 2\mathbb{P}\left(\sup_{s\in[0,\delta]} B_H(s) \ge \varepsilon/2 - c\delta \right)
$$

$$= \quad 2\mathbb{P}\left(\sup_{s\in[0,1]} B_H(s) \geq \frac{\varepsilon/2 - c\delta}{\delta^H}\right) \tag{5.17}$$

$$= \quad 2\Psi\left(\frac{\varepsilon/2 - c\delta}{\delta^H}\right)(1 + o(1)) \tag{5.18}$$

$$= \quad o(\delta)$$

as $\delta \downarrow 0$, where $\Psi(\cdot)$ denotes the tail distribution of a standard normal random variable. Here (5.17) follows from the self-similarity of $B_H(t)$. Equation (5.18) is a consequence of the fact that $\lim_{x\to\infty} \mathbb{P}(\sup_{s\in[0,1]} B_H(s) \geq x)/\Psi(x) = 1$, see Theorem 1 of Berman [1]. This completes the proof. □

5.3 Proof of the small time-scale results

This section contains the proofs of Theorems 4.1 and 4.2.

Proof of Thm. 4.1. Proposition 2.1, part (i), takes care of the limit $N \to \infty$. Using the results in Lindvall [9], it suffices to show that

$$\lim_{T\downarrow 0}\{r_T A_T(0,t) - ct, t \in [0,S]\} =_{\mathrm{d}} \{B_1(t) - ct, t \in [0,S]\} \tag{5.19}$$

for each $S > 0$. The proof consists of two steps: convergence of finite-dimensional distributions and tightness.

Step 1: Convergence of finite-dimensional distributions. Recall that, for each $T > 0$, $A_T(0,t)$ is a Gaussian process with stationary increments. Hence the variance function $\mathbb{V}\mathrm{ar}\ (A_T(0,t))$ completely describes the covariance function of $A_T(0,t)$, and consequently also the finite-dimensional distributions. Thus, in order to complete this step of the proof, it suffices to prove the pointwise convergence of $\mathbb{V}\mathrm{ar}\ (r_T A_T(0,t))$ to $\mathbb{V}\mathrm{ar}\ B_1(t) = t^2$ as $T \downarrow 0$.

Let $t > 0$ be given. Recall that $\mathbb{V}\mathrm{ar}\ (r_T A_T(0,t)) = \sigma^2(tT)/\sigma^2(T)$. Now invoke part (i) of Lemma 5.1. We thus obtain the following limit

$$\lim_{T\downarrow 0} \mathbb{V}\mathrm{ar}\ (r_T A_T(0,t)) = t^2.$$

This completes the proof of the convergence of finite-dimensional distributions.

Step 2: Tightness. Let $S > 0$ be given. Due to Theorem 12.3 in [2] and the fact that $A_T(0,t)$ has stationary increments, it suffices to prove that

$$\mathbb{V}\mathrm{ar}\ (r_T A_T(0,t)) \leq K \cdot t^2$$

for a constant K, sufficiently small $T > 0$ and all $t \in [0,S]$.

Using (i) of Lemma 5.1 we infer that there exists a $T_0 > 0$ such that for each $v \in (0, ST_0]$

$$\sigma^2(v) \geq \frac{1}{2}R(0)v^2.$$

Moreover, $\sigma^2(v) \leq R(0)v^2$ for each $v > 0$. Hence

$$\mathbb{V}\mathrm{ar}\ (r_T A_T(0,t)) = \frac{\sigma^2(tT)}{\sigma^2(T)} \leq \frac{R(0)t^2T^2}{\frac{1}{2}R(0)T^2} = 2t^2$$

for each $T \leq T_0$ and $t \in [0, S]$. This completes the proof. □

Proof of Thm. 4.2. Due to Proposition 2.1 $\lim_{N\to\infty} X^\star_{T,N}(0) =_{\mathrm{d}} X^\star_T(0)$ for each $T > 0$. We analyze $X^\star_T(0)$ as $T \downarrow 0$. Straightforward manipulations yield

$$\begin{aligned}
\mathbb{P}(X^\star_T(0) > u) &= \mathbb{P}\left(\sup_{t\geq 0} r_T \int_0^t Z^\star_T(s)\mathrm{d}s - ct > u\right) \\
&= \mathbb{P}\left(\sup_{t\geq 0} \frac{T}{\sigma(T)} \int_0^t Z^\star(sT)\mathrm{d}s - ct > u\right) \\
&= \mathbb{P}\left(\sup_{t\geq 0} \frac{1}{\sigma(T)} \int_0^{tT} Z^\star(s)\mathrm{d}s - ct > u\right) \\
&= \mathbb{P}\left(\sup_{t\geq 0} \frac{1}{\sigma(T)} \int_0^{t} Z^\star(s)\mathrm{d}s - \frac{c}{T}t > u\right) \\
&= \mathbb{P}\left(\sup_{t\geq 0} \int_0^{t} Z^\star(s)\mathrm{d}s - \frac{c\sigma(T)}{T}t > \sigma(T)u\right)
\end{aligned}$$

for each $u \geq 0$. Invoking part (i) of Lemma 5.1, this immediately implies Remark 4.1:

$$\lim_{T\downarrow 0} \mathbb{P}(X^\star_T(0) > u) = \mathbb{P}\left(\sup_{t\geq 0} \int_0^t Z^\star(s)\mathrm{d}s - c\sqrt{R(0)}t > 0\right).$$

But

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\geq 0} \int_0^t Z^\star(s)\mathrm{d}s - c\sqrt{R(0)}t > 0\right) &= \mathbb{P}\left(\sup_{t\geq 0} \int_0^t (Z^\star(s) - c\sqrt{R(0)})\mathrm{d}s > 0\right) \\
&> \mathbb{P}(Z^\star(0) > c\sqrt{R(0)}) = \mathbb{P}(\mathcal{N} > c) \\
&= \mathbb{P}\left(\sup_{t\geq 0} B_1(t) - ct > u\right) = \mathbb{P}(Y^\star_1(0) > u)
\end{aligned}$$

for each $u \geq 0$, since $B_1(t) =_{\mathrm{d}} t\mathcal{N}$, where $\mathcal{N}$ denotes a standard normal random variable. This completes the proof. □

References

[1] Berman, S. (1985). An asymptotic formula for the distribution of the maximum of a Gaussian process with stationary increments. *J. Appl. Probab.* **22**, 454–460.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

[3] Bingham, N.H., Goldie, C.M. & Teugels, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

[4] Borovkov, A.A. (1976). *Stochastic Processes in Queueing Theory.* Springer, New York.

[5] Cao, J. & Ramanan, K. (2002). A Poisson limit for buffer overflow probabilities. *Proceedings INFOCOM 2002.*

[6] Dębicki, K. & Palmowski, Z. (1999). Heavy traffic asymptotics of on-off fluid model. *Queueing Systems* **33**, 327–338.

[7] Heath, D., Resnick, S. & Samorodnitsky, G. (1998). Heavy tails and long range dependence in ON/OFF processes and associated fluid models. *Math. Oper. Res.* **23**, 145–165.

[8] Kulkarni, V., & Rolski, T. (1994). Fluid model driven by an Ornstein-Uhlenbeck process. *Prob. Eng. Inf. Sci.* **8**, 403–417.

[9] Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty)$. *J. Appl. Probability* **10**, *109–121.*

[10] Mikosch, T., Resnick, S., Rootzén, H. & Stegeman, A. (2002). Is network traffic approximated by stable Lévy motion or fractional Brownian motion. *Ann. Appl. Prob.* **12**, 23–68.

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

[12] Piterbarg, V.I. (1996). *Asymptotic methods in the theory of Gaussian processes and fields.* Translations of Mathematical Monographs 148, AMS, Providence.

[13] Szczotka, W. (1980). Central Limit Theorem in $D[0,\infty)$ for breakdown processes. *Prob. Math. Stat.* **1**, 49–57.

[14] Taqqu, M.S., Willinger, W. & Sherman, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *Comp. Comm. Rev.* **27**, 5–23.

[15] Whitt, W. (2002). Stochastic-Process Limits. Springer, New York.