*Stochastic Systems*
arXiv: arXiv:0000.0000

# HEAVY-TRAFFIC LIMIT OF THE $GI/GI/1$ STATIONARY DEPARTURE PROCESS AND ITS VARIANCE FUNCTION

By Ward Whitt*, and Wei You*

Heavy-traffic limits are established for the stationary departure process from a $GI/GI/1$ queue and its variance function. The limit process is a function of the Brownian motion limits of the arrival and service processes plus the stationary reflected Brownian motion (RBM) limit of the queue-length process. An explicit expression is given for the variance function, which depends only on the first two moments of the interarrival times and service times plus the previously-determined correlation function of canonical (drift $-1$, diffusion coefficient 1) RBM. The limit for the variance function here is used to show that the approximation for the index of dispersion for counts of the departure process used in our new robust queueing network analyzer is asymptotically correct in the heavy-traffic limit.

**1. Introduction.** In this paper we establish a *heavy-traffic* (HT) limit for the stationary departure process from the stable $GI/GI/1$ queue and its variance function. In doing so, we are primarily motivated by our desire to develop a new *robust queueing network analyzer* (RQNA) for open networks of single-server queues exploiting the *index of dispersion for counts* (IDC) for all arrival processes, which we refer to as the RQNA-IDC. The new RQNA-IDC is a parametric-decomposition approximation (i.e., it treats the individual queues separately) like the *queueing network analyzer* (QNA) in [37]. Instead of approximately characterizing each flow by its rate and a single variability parameter, we approximately characterize the flow by its rate and the IDC, which is a real-valued function on the positive halfline, in particular the scaled variance function; see §6. The need for such an enhanced version of QNA has long been recognized, as can be seen from [32, 31, 38, 43].

This paper is a sequel to [40], which developed a *robust queueing* (RQ) approximation for the steady-state workload in a $G/G/1$ queue; see §6.1 here. That RQ algorithm extends an earlier RQ algorithm by Bandi, Bertsimas and Youssef [8]. Indeed, a full RQNA is developed in [8], but the RQ approximation in [8] is a parametric RQ formulation, based on a single variability parameter, as in [37]. In contrast, the new RQ algorithm in [40] involves

*Department of Industrial Engineering and Operations Research, Columbia University
*Keywords and phrases:* heavy traffic, departure processes, index of dispersion, queueing networks, robust queueing

a functional formulation that incorporates the variance of the input over time. The advantage of the new formulation is demonstrated by simulation comparisons in [40].

A full RQNA-IDC for open networks of $G/GI/1$ single-server queues is outlined in §6 of [40]. There it is noted that the main challenge in developing an effective RQNA that can better capture dependence in the arrival processes at the different queues is to be able to approximate the IDC of the stationary departure process from a $G/GI/1$ queue. Important theoretical support for the new RQNA-IDC developed and studied in [42] is provided by the present paper. Corollary 6.1 here shows that the proposed approximation for the IDC of the stationary departure process is asymptotically correct for the $GI/GI/1$ queue in the HT limit.

The HT limit here is also of considerable interest more generally, because the stationary departure process from a $GI/GI/1$ queue is remarkably complicated; e.g., it is only a stationary renewal process in the special case of an $M/M/1$ model, when it is Poisson, by Burke's [12] theorem; also see [33] and references therein. Indeed, explicit transform expressions for the variance function of the stationary departure process are evidently only available for the $M/GI/1$ and $GI/M/1$ models, due to Takacs [36] and Daley [16, 17]; see [9] and [27] for related results on $GI/GI/1$. We exploit the $GI/M/1$ and $M/GI/1/$ results here to directly establish HT limits for the variance function in §3 and §4.

The HT limit for the departure process starting empty in the $GI/GI/1$ model and more general multi-channel models is an old result, being contained in Theorem 2 of [29], but HT limits for associated stationary processes have proven far more difficult. To the best of our knowledge, we derive the first HT limits for the stationary departure process and its variance function for any $GI/GI/1$ model except $M/M/1$. The key to the process limit here is the recent HT limits for the stationary vector of queue lengths in a generalized Jackson network in Gamarnik and Zeevi [22] and Budhiraja and Lee [11]. The existence of the HT limit for the scaled variance function is a relatively direct consequence.

The most interesting and useful contribution here is the explicit form of the limiting variance function and its connection to basic functions of reflected Brownian motion (RBM) in [1, 2]. For this step, we exploit the transform limits in the $M/GI/1$ and $GI/M/1$ special cases. We apply time-space transformations of the underlying Brownian motions in the $GI/GI/1$ limit to show that the limit with adjusted parameters is the same as for the $M/GI/1$ or $GI/M/1$ special case. Thus, we can identify the explicit form of the limiting variance function for the $GI/GI/1$ model by exploiting the results for the special cases; see the proof of Theorem 5.3. This same formula serves as an approximation more generally; see §7.

Here is how this paper is organized: We start in §2 by providing a brief review of stationary point processes, focusing especially on the variance function. In §3 we use Laplace transforms (LT's) of the stationary departure process in the $GI/M/1$ queue derived by [16, 17] to derive the HT limit of its variance function. In §4 we use the HT limit for the Palm version of the mean function derived by [36] to derive the HT limit of the stationary variance function. (In §2.2 we review the application of the Palm-Khintchine equation to express the stationary variance in terms of the Palm mean function.) In §5 we establish the HT limit for the stationary departure process in the $GI/GI/1$ queue (Theorem 5.2) and its variance function (Theorem 5.3). In §6 we provide a brief overview of the application of Theorem 5.3 to support our RQNA-IDC developed in [42], which is briefly outlined in §6 of [40]. Finally, we discuss extensions in §7.

**2. Review of Stationary Point Processes.** In this section we review basic properties of stationary point processes; see [18] and [35] for more background. In §2.1 we review renewal processes and their Laplace transforms. In §2.2 we review the Palm-Khintchine equation and use it to express the variance function of a stationary point process in terms of the mean function of the Palm version.

2.1. *Renewal Processes and the Laplace Transform.* We start with a rate-$\lambda$ renewal process $N \equiv \{N(t) : t \geq 0\}$. Let $F$ be the *cumulative distribution function* (cdf) of the interval $U$ between points (the interarrival time in a $GI$ arrival process), having mean $E[U] = \lambda^{-1}$ and finite second moment. As a regularity condition for our queueing application, we also assume that $F$ has a *probability density function* (pdf) $f$, where $F(t) = \int_0^t f(u)\,du$, $t \geq 0$. Let $F_e$ be the cdf of the corresponding equilibrium distribution, which has pdf $f_e(t) = \lambda(1 - F(t))$. Let $E^e[\cdot]$ denote the expectation under the stationary distribution (with first interval distributed according to $F_e$) and let $E^0[\cdot]$ denote the expectation under *the Palm distribution* (with first interval distributed as $F$).

Conditioning on the first arrival, distributed as $F$ under the Palm distribution or as $F_e$ under stationary distribution, the renewal equations for the mean and second moment of $N(t)$, the number of points in an interval $[0, t]$, are:

$$
\begin{aligned}
m(t) &\equiv E^0[N(t)] = F(t) + \int_0^t m(t-s)dF(s) \\
m_e(t) &\equiv E^e[N(t)] = F_e(t) + \int_0^t m(t-s)dF_e(s) \\
\sigma(t) &\equiv E^0[N^2(t)] = F(t) + 2\int_0^t m(t-s)dF(s) + \int_0^t \sigma(t-s)dF(x) \\
\sigma_e(t) &\equiv E^e[N^2(t)] = F_e(t) + 2\int_0^t m(t-s)dF_e(s) + \int_0^t \sigma(t-s)dF_e(x).
\end{aligned}
$$

Throughout the paper, we use the Laplace Transform (LT) instead of the Laplace-Stieltjes Transform (LST). The LT of $f(t)$ and the LST of $F$, denoted by $\mathcal{L}(f)(s) \equiv \hat{f}(s)$, are

$$\hat{f}(s) \equiv \mathcal{L}(f)(s) \equiv \int_0^\infty e^{-st} f(t)dt = \int_0^\infty e^{-st} dF(t), \tag{2.1}$$

so that $f(t) = \mathcal{L}^{-1}(\hat{f})(t)$. Throughout the paper, we add a hat to either a LT or item that appears in LT. The LT of $f_e$ is then

$$\hat{f}_e(s) = \frac{\lambda(1-\hat{f}(s))}{s} \text{ and } \hat{F}_e(s) = \frac{\hat{f}_e(s)}{s},$$

where $\lambda^{-1} \equiv \int_0^\infty t f(t)\, dt$ is the mean. Applying the LT to the renewal equations, we obtain

$$\hat{m}(s) = \frac{\hat{f}(s)}{s(1-\hat{f}(s))} \tag{2.2}$$

$$\hat{m}_e(s) = \frac{\hat{f}_e(s)}{s(1-\hat{f}(s))} = \frac{\lambda}{s^2} \tag{2.3}$$

$$\hat{\sigma}(s) = \frac{\hat{f}(s) + 2s\hat{m}(s)\hat{f}(s)}{s(1-\hat{f}(s))} = \frac{\hat{f}(s)(1+\hat{f}(s))}{s(1-\hat{f}(s))^2} \tag{2.4}$$

$$\hat{\sigma}_e(s) = \frac{\lambda}{s^2} + \frac{2\lambda}{s}\hat{m}(s) = \frac{\lambda(1+\hat{f}(s))}{s^2(1-\hat{f}(s))} \tag{2.5}$$

From (2.3), we see that

$$E^e[N(t)] = \lambda t, \quad t \geq 0, \tag{2.6}$$

as must be true for any stationary point process.

Let $V(t) \equiv \mathrm{Var}^e(N(t))$ be the variance process of $N(t)$ under time-stationary distribution. (We omit the $e$ superscript on $V(t)$ because we will only discuss stationary variance functions.) Combining (2.5) and (2.6), we have

$$\hat{V}(s) = \frac{\lambda}{s^2} + \frac{2\lambda}{s}\hat{m}(s) - \frac{2\lambda^2}{s^3} = \frac{\lambda}{s^2} + \frac{2\lambda}{s}\frac{\hat{f}(s)}{s\left(1-\hat{f}(s)\right)} - \frac{2\lambda^2}{s^3}. \tag{2.7}$$

The variance function then can be obtained from the numerical inversion of the Laplace transform, e.g., see §13 of [3], [4] and [41]. Term by term inversion shows that we can express $V(t)$ in terms of the renewal function $m(t)$

$$V(t) = \lambda \int_0^t (1 + 2m(u) - 2\lambda u)du. \tag{2.8}$$

In §2.2 we show that the Palm-Khintchine equation can be used to derive a generalization of (2.8) for general stationary and ergodic point processes.

2.2. *The Palm-Khintchine Equation.* We now consider a continuous-time stationary point process, i.e., having stationary increments. The main idea is the Palm transformation relating continuous-time stationary processes to the associated discrete-time stationary processes. An important manifestation of that relation is the Palm-Khintchine equation; see Theorem 3.4.II. of [18]. It is important here because it can be applied to generalize the variance formula discussed in §2.1; see §2.4 of [15] and §3.4 of [18].

We focus on orderly stationary ergodic point processes with finite intensity. (Orderly means that the points occur one at a time.) Let $N(s,t]$ denote the number of events in interval $(s,t]$, and $N(t) \equiv N(0,t]$.

THEOREM 2.1. (*the Palm-Khintchine equation*) *For an orderly stationary point process of finite intensity $\lambda$ such that*

$$P^e(N(-\infty,0] = N(0,\infty) = \infty) = 1,$$

$$P^e(N(t) \le k) = 1 - \lambda \int_0^t q_k(u)du = \lambda \int_t^\infty q_k(u)du, \ \text{for } k = 0,1,2,\ldots, \tag{2.9}$$

*where $q_k(t)$ is the probability of exactly $k$ arrivals in $(0,t]$ under the Palm distribution, i.e.,*

$$q_k(t) = \lim_{h\downarrow 0} P(N(t) = k | N(-h,0] > 0)). \tag{2.10}$$

*Under ergodicity, the Palm distribution is equivalent to the event stationary distribution, so that* $q_k(t) = P^0(N(t) = k)$.

We now apply Theorem 2.1 to generalize (2.8) and (2.7) to the case of orderly stationary ergodic point process.

COROLLARY 2.1. (*the variance of a stationary ergodic point process*) *For a general stationary ergodic point process with rate* $\lambda$ *and finite second moment, the variance function is*

$$V(t) = \lambda \int_0^t (1 + 2m(u) - 2\lambda u) du, \quad t \geq 0, \tag{2.11}$$

*where*

$$m(t) \equiv E^0[N(t)] = \sum_{k=1}^{\infty} k q_k(t), \quad t \geq 0, \tag{2.12}$$

*and its LT is*

$$\hat{V}(s) = \frac{\lambda}{s^2} + \frac{2\lambda}{s} \hat{m}(s) - \frac{2\lambda^2}{s^3} \tag{2.13}$$

*where* $\hat{m}(s)$ *is the LT of* $m(t)$.

*Proof.* Let

$$p_k(t) = P^e(N(t) = k), \text{ for } k = 0, 1, 2, \ldots \tag{2.14}$$

so that $\sum_{i=1}^{k} p_k(t) = P^e(N(t) \leq k)$. With Theorem 2.1, we can write

$$\begin{aligned} V(t) = \sum_{k=1}^{\infty} k^2 p_k(t) - \lambda^2 t^2 &= \sum_{k=1}^{\infty} k^2 \lambda \int_0^t (q_{k-1}(u) - q_k(u)) du - \lambda^2 t^2 \\ &= \lambda \int_0^t (1 + 2m(u) - 2\lambda u) du, \end{aligned} \tag{2.15}$$

where $m(t) \equiv E^0[N(t)]$ as in (2.12) Taking the Laplace Transform, we obtain

$$\hat{V}(s) = \frac{\lambda}{s^2} + \frac{2\lambda}{s} \hat{m}(s) - \frac{2\lambda^2}{s^3}. \quad \blacksquare$$

**3. The Departure Variance in the $GI/M/1$ Queue.** Daley [16, 17] derived the LST of the variance $V_d(t)$ of the stationary departure process in a $GI/M/1$ queue. The associated LT of $V_d(t)$ is

(3.1)
$$\hat{V}_d(s) = \frac{\lambda}{s^2} + \frac{2\lambda}{s^3}\left(\mu\delta - \lambda + \frac{\mu^2(1-\delta)(1-\hat{\xi}(s))(\mu\delta(1-\hat{f}(s)) - s\hat{f}(s))}{(s+\mu(1-\hat{\xi}(s)))(s-\mu(1-\delta))(1-\hat{f}(s))}\right)$$

where $\lambda$ is the arrival rate, $\mu$ is the service rate (with $\lambda < \mu$), $\hat{f}(s) = E\left[e^{-sU}\right]$ is the LT of the interarrival-time pdf $f(t)$, $\hat{\xi}(s)$ is the root with the smallest absolute value in $z$ of the equation

$$z = \hat{f}(s + \mu(1-z)) \tag{3.2}$$

and $\delta = \hat{\xi}(0)$ is the unique root in $(0,1)$ of the equation

$$\delta = \hat{f}(\mu(1-\delta)), \tag{3.3}$$

which appears in the distribution of the stationary queue length in a $GI/M/1$ queue.

We now present a useful lemma on properties of $\hat{\xi}(s)$ and $\delta$; see p. 113 of [36] or Appendix 6 of [13]. (The notation here is slightly different.)

Lemma 3.1. (*root lemma from Takacs* [36]) *If $Re(s) \geq 0$, then the root $\hat{\xi}(s)$ of the equation*
$$z = \hat{f}(s + \mu(1-z))$$
*that has the smallest absolute value is*

$$\hat{\xi}(s) = \sum_{j=1}^{\infty} \frac{(-\mu)^{j-1}}{j!} \frac{d^{j-1}}{ds^{j-1}} \left(\hat{f}(\mu + s)\right)^j. \tag{3.4}$$

*This root $\hat{\xi}(s)$ is a continuous function of $s$ for $Re(s) \geq 0$. Furthermore, $z = \hat{\xi}(s)$ is the only root in the unit circle $z \leq 1$ if $Re(s) > 0$ or $Re(s) \geq 0$ and $\lambda/\mu < 1$. Specifically, $\delta = \hat{\xi}(0)$ is the smallest positive real root of the equation*
$$\delta = \hat{f}(\mu(1-\delta)).$$
*If $\lambda/\mu < 1$, then $\delta < 1$ and if $\lambda/\mu \geq 1$ then $\delta = 1$.*

We now establish a HT limit for the departure variance function in the $GI/M/1$ model. To do so, we consider a family of $GI/M/1$ models parameterized by $\rho$, where $\lambda \equiv E[U]$ and $\mu = \mu_\rho = E[V] \equiv \lambda(1 + (1-\rho)\gamma_\rho)$, where $\gamma_\rho$ are positive constants such that $\lim_{\rho\uparrow 1}\gamma_\rho = \gamma > 0$. Note that if $\gamma_\rho = 1/\rho$, then we come to the usual case of $\lambda/\mu = \rho$. We allow this general scaling so that we can gain insight into reflected Brownian motion (RBM) with non-unit drift. Let the HT-scaled variance function be

$$V^*_{d,\rho}(t) \equiv (1-\rho)^2 V_{d,\rho}\left((1-\rho)^{-2}t\right), \quad t \ge 0. \tag{3.5}$$

Throughout the paper, we use the asterisk (*) superscript with $\rho$ subscript to denote HT-scaled items in the queueing model, as in (3.5), and the asterisk without the $\rho$ subscript to denote the associated HT limit.

As should be expected from established HT limits, e.g., as in §5.7 and Chapter 9 in [39], the HT limit of the variance function $V^*_{d,\rho}(t)$ in (3.5) depends on properties of the normal distribution and RBM. Let $\tilde{\phi}(x)$ be the pdf and $\tilde{\Phi}(x)$ the cdf of the standard normal variable $N(0,1)$. (We use $\tilde{\phi}$ for the pdf, because we use $\phi$ for the reflection map; see (5.1).) Let $\tilde{\Phi}^c(x) \equiv 1 - \tilde{\Phi}(x)$ be the complementary cdf (ccdf). Let $R(t)$ be canonical RBM (having drift $-1$, diffusion coefficient 1) and let $R_e(t)$ be the stationary version, which has the exponential marginal distribution for each $t$ with mean $1/2$. Let $c^*(t)$ be the correlation function of $R_e$

$$\begin{aligned} c^*(t) &\equiv \frac{E[R_e(0)R_e(t)] - E[R_e(0)]E[R_e(t)]}{\mathrm{Var}(R_e(0))} \\ &= 2(1-2t-t^2)\tilde{\Phi}^c(\sqrt{t}) + 2\sqrt{t}\tilde{\phi}(\sqrt{t})(1+t), \\ &= 1 - H_2^*(t) \equiv 1 - \frac{E[R(t)^2|R(0)=0]}{E[R(\infty)^2]} \\ &= 1 - 2E[R(t)^2|R(0)=0], \quad t \ge 0; \end{aligned} \tag{3.6}$$

i.e., where $H_2^*(t)$ is the second-moment cdf of canonical RBM in [1], which has mean 1 and variance 2.5; see Corollaries 1.1.1 and 1.3.4 of [1] and Corollary 1 of [2]. The correlation function $c^*(t)$ has LT

$$\hat{c}^*(s) \equiv \frac{1}{s} - \frac{2}{s^2}\left(1 - \frac{\sqrt{1+2s}-1}{s}\right); \tag{3.7}$$

see (1.10) of [1]. Equivalently, the Gaussian terms in (3.6) can be reexpressed as $\tilde{\phi}(\sqrt{t}) = e^{-t/2}/\sqrt{2\pi}$ and $\tilde{\Phi}^c(\sqrt{t}) = (1 - \mathrm{erf}(\sqrt{t/2}))/2$, where erf is the error function.

By Corollary 1.3.5 of [1], the correlation function has tail asymptotics according to

$$c^*(t) = 1 - H_2^*(t) \sim \frac{16}{\sqrt{2\pi t^3}} e^{-(t/2)} \quad \text{as} \quad t \to \infty. \tag{3.8}$$

THEOREM 3.1. (*HT limit for the $GI/M/1$ departure variance*) *Consider the $GI/M/1$ model with $1/E[U] = \lambda$ and $1/E[V] = \mu_\rho \equiv \lambda(1 + (1-\rho)\gamma_\rho)$, where $\gamma_\rho$ are positive constants such that $\lim_{\rho\uparrow 1}\gamma_\rho = \gamma > 0$. Assume that $E[U^3] < \infty$ so that a two-term Taylor series expansion of the LT $\hat{f}(s)$ about the origin is valid with asymptotically negligible remainder. Then*

$$V^*_{d,\rho}(t) \to V^*_d(t) \quad as \quad \rho \uparrow 1 \tag{3.9}$$

*for $V^*_{d,\rho}(t)$ in (3.5), where the limit $V^*_d(t)$ is a finite function with LT*

$$\hat{V}^*_d(s) = \frac{\lambda}{s^2} + \left(\frac{2\lambda}{s^2}\right)\left(\frac{c_a^2 - 1}{c_a^2 + 1}\right)\left(\frac{\gamma}{\hat{\xi}^*(s)}\right) = \frac{\lambda c_a^2}{s^2} - (\lambda c_a^2 - \lambda)\hat{h}^*(s), \tag{3.10}$$

*where*

$$\hat{h}^*(s) \equiv \hat{h}^*_{\gamma,c_a^2}(s) \equiv \frac{1}{s^2}\left(1 - \left(\frac{2}{c_a^2+1}\right)\left(\frac{\gamma}{\hat{\xi}^*(s)}\right)\right) \tag{3.11}$$

*and $\hat{\xi}^*(s)$ is the unique root with non-negative real part of the quadratic equation*

$$\left(\frac{c_a^2+1}{2}\right)\hat{\xi}^*(s)^2 - \gamma\hat{\xi}^*(s) - \frac{s}{\lambda} = 0. \tag{3.12}$$

*Hence,*

$$h^*(t) = t\left(1 - w^*(\lambda\gamma^2 t/c_x^2)\right) = \frac{c_x^2}{2\lambda\gamma^2}\left(1 - c^*(\lambda\gamma^2 t/c_x^2)\right), \quad t \ge 0, \tag{3.13}$$

*and*

$$V^*_d(t) \equiv w^*\left(\lambda\gamma^2 t/c_x^2\right) c_a^2\lambda t + \left(1 - w^*\left(\lambda\gamma^2 t/c_x^2\right)\right) c_s^2 \lambda t, \quad t \ge 0, \tag{3.14}$$

*where $c_x^2 \equiv c_a^2 + c_s^2$ with $c_a^2 \equiv \mathrm{Var}(U)/E[U]^2$, $c_s^2 = 1$ and*

$$w^*(t) \equiv 1 - \frac{1 - c^*(t)}{2t}, \quad t \ge 0, \tag{3.15}$$

*with $c^*(t)$ being the correlation function of canonical RBM in (3.6). (As $t$ increases from $0$ to $\infty$, $w^*(t)$ increases from $0$ to $1$.) Equivalently,*

$$1 - w^*(t) = \frac{1 - c^*(t)}{2t} = \frac{H_2^*(t)}{2t}, \quad t \ge 0, \tag{3.16}$$

*for $H_2^*(t)$ the RBM second-moment cdf and*

$$\begin{aligned} V^*_d(t) &= c_a^2\lambda t + \left(1 - w^*\left(\lambda\gamma^2 t/c_x^2\right)\right)(c_s^2 - c_a^2)\lambda t \\ &= c_a^2\lambda t + \frac{(c_s^2 - c_a^2)c_x^2}{2\gamma^2} H_2^*(\lambda\gamma^2 t/c_x^2) \\ &= c_a^2\lambda t + \frac{(c_s^2 - c_a^2)c_x^2}{2\gamma^2} - \frac{(c_s^2 - c_a^2)c_x^2}{2\gamma^2} c^*(\lambda\gamma^2 t/c_x^2), \quad t \ge 0. \end{aligned} \tag{3.17}$$

*Proof.* We let $\rho \uparrow 1$ by decreasing the service rate, so that $1/E[U] = \lambda$ is fixed. To allow general drift in the Brownian HT limit, we let $1/E[V] = \mu_\rho \equiv \lambda + (1-\rho)\lambda\gamma_\rho$ in system $\rho$, for positive constants $\gamma_\rho \to \gamma$. Under this setting, we have $(\lambda - \mu_\rho)/(1-\rho) \to -\lambda\gamma$ as $\rho \uparrow 1$. By (3.1) and (3.5), we have

$$\begin{aligned}\hat{V}^*_{d,\rho}(s) &= \mathcal{L}\left((1-\rho)^2 V_{d,\rho}\left((1-\rho)^{-2}t\right)\right) = (1-\rho)^4 \hat{V}_{d,\rho}\left((1-\rho)^2 s\right)\\ &= \frac{\lambda}{s^2} + \frac{\lambda}{s^2}\hat{W}(s)\end{aligned}$$

where

$$\hat{W}(s) \equiv \frac{2}{(1-\rho)^2 s}\left(\mu_\rho(\delta - \frac{\lambda}{\mu_\rho}) + \frac{\mu_\rho \delta \frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} - \hat{f}\left((1-\rho)^2 s\right)}{\frac{(1-\rho)^2 s + \mu_\rho(1-\hat{\xi}((1-\rho)^2 s))}{\mu_\rho(1-\xi((1-\rho)^2 s))} \cdot \frac{(1-\rho)^2 s - \mu_\rho(1-\delta)}{\mu_\rho(1-\delta)} \cdot \frac{1-\hat{f}((1-\rho)^2 s)}{(1-\rho)^2 s}}\right).$$

Then, we write

$$\begin{aligned}\hat{W}(s) &= \frac{2\mu_\rho}{(1-\rho)^2 s}\frac{\left(\delta - \frac{\lambda}{\mu_\rho}\right)\hat{H}_\rho(s)\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} + \delta\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} - \frac{1}{\mu_\rho}\hat{f}\left((1-\rho)^2 s\right)}{\hat{H}_\rho(s)\frac{1-\hat{f}((1-\rho)^2 s)}{(1-\rho)^2 s}}\\
&= \frac{2\mu_\rho}{(1-\rho)^2 s}\frac{\delta\left(\hat{H}_\rho(s)+1\right)\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} - \frac{\lambda}{\mu_\rho}\hat{H}_\rho(s)\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} - \frac{1}{\mu_\rho}\hat{f}\left((1-\rho)^2 s\right)}{\hat{H}_\rho(s)\frac{1-\hat{f}((1-\rho)^2 s)}{(1-\rho)^2 s}}\\
&= \frac{1}{\hat{H}_\rho(s)\frac{1-\hat{f}((1-\rho)^2 s)}{(1-\rho)^2 s}}\frac{2\mu_\rho}{(1-\rho)^2 s}\left((\delta - \frac{\lambda}{\mu_\rho})\left(\hat{H}_\rho(s)+1\right)\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s}\right.\\
&\qquad\left. + \frac{\lambda}{\mu_\rho}\left(\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} - \frac{1}{\lambda}\right) + \frac{1}{\mu_\rho}\left(1 - \hat{f}\left((1-\rho)^2 s\right)\right)\right)\\
&= \frac{2\mu_\rho}{\hat{H}_\rho(s)\frac{1-\hat{f}((1-\rho)^2 s)}{(1-\rho)^2 s}}\left(\frac{\delta - \frac{\lambda}{\mu_\rho}}{1-\rho}\frac{\hat{H}_\rho(s)+1}{(1-\rho)s}\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s}\right.\\
&\qquad\left. + \frac{\lambda\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} - 1}{\mu_\rho(1-\rho)^2 s} + \frac{1-\hat{f}\left((1-\rho)^2 s\right)}{\mu_\rho(1-\rho)^2 s}\right)\end{aligned}$$

where

$$\hat{H}_\rho(s) \equiv \left(\frac{1}{\mu_\rho}\frac{(1-\rho)^2 s}{1-\hat{\xi}\left((1-\rho)^2 s\right)} + 1\right)\left(\frac{1}{\mu_\rho}\frac{(1-\rho)^2 s}{1-\delta} - 1\right).$$

By Lemma 3.1, we know that $\delta$ is positive and real, and $\delta < 1$ if $\rho < 1$ while $\delta = 1$ if $\rho = 1$. Hence, we may restrict the function $\hat{f}$ to the real axis. Then, expanding $\hat{f}$ in a Taylor series about 0, yields

$$\begin{aligned}
\delta = \hat{f}(\mu_\rho(1-\delta)) \Rightarrow \delta &= \hat{f}(0) + \hat{f}'(0)\mu_\rho(1-\delta) + \left(\frac{1}{2}\hat{f}''(0)\mu_\rho^2 + o(1)\right)(1-\delta)^2 \\
&\Rightarrow 0 = 1 - \delta - \frac{\mu_\rho}{\lambda}(1-\delta) + \left(\frac{1}{2}\hat{f}''(0)\mu_\rho^2 + o(1)\right)(1-\delta)^2 \\
&\Rightarrow 0 = \frac{1-\frac{\mu_\rho}{\lambda}}{1-\rho} + \left(\frac{c_a^2+1}{2}\frac{\mu_\rho^2}{\lambda^2} + o(1)\right)\frac{1-\delta}{1-\rho} \\
&\Rightarrow \frac{1-\delta}{1-\rho} = \gamma_\rho\left(\frac{c_a^2+1}{2\rho} + o(1)\right)^{-1} \qquad (3.18)
\end{aligned}$$

This implies that the following limit exist

$$\delta^* \equiv \lim_{\rho\uparrow 1}\frac{1-\delta}{1-\rho} = \frac{2\gamma}{c_a^2+1}. \qquad (3.19)$$

Now, let $\hat{\xi}_{\rho,s} \equiv \hat{\xi}\left((1-\rho)^2 s\right) = \hat{f}\left((1-\rho)^2 s + \mu_\rho(1-\hat{\xi}_{\rho,s})\right)$, then similarly we have

$$0 = \gamma_\rho\frac{1-\hat{\xi}_{\rho,s}}{1-\rho} + \frac{s}{\lambda} - \left(\frac{c_a^2+1}{2\lambda^2} + o(1)\right)\left((1-\rho)s + \mu_\rho\frac{1-\hat{\xi}_{\rho,s}}{1-\rho}\right)^2. \qquad (3.20)$$

Then (3.20) implies that the following limit exist

$$\hat{\xi}^*(s) \equiv \lim_{\rho\uparrow 1}\frac{1-\hat{\xi}_{\rho,s}}{1-\rho}, \qquad (3.21)$$

and

$$\frac{c_a^2+1}{2}\left(\hat{\xi}^*(s)\right)^2 - \gamma\hat{\xi}^*(s) - \frac{s}{\lambda} = 0. \qquad (3.22)$$

Recall that $\hat{\xi}_{\rho,s}$ is defined to be the root of $z = \hat{f}((1-\rho)^2 s + \mu_\rho(1-z))$ with smallest absolute value. By Lemma 3.1, this root is unique and lies in the unit circle unless $s = 0$ and $\rho = 1$, in which case $\hat{\xi}(0) = 1$. Furthermore, it can be proved by Weierstrass Preparation Theorem that $\hat{\xi}_{\rho,s}$ is continuous in $(\rho, s)$. Hence, we have

$$Re\left(\frac{1-\hat{\xi}_{\rho,s}}{1-\rho}\right) > 0, \text{ for all } \rho < 1 \text{ and } s > 0.$$

By taking limit $\rho \uparrow 1$, we have $Re(\hat{\xi}^*(s)) \geq 0$ for all $s > 0$.

As a consequence, we pick the root of (3.22) with non-negative real part. In particular, for real $s$, we have

$$\hat{\xi}^*(s) = \frac{\gamma + \sqrt{\gamma^2 + 2(c_a^2+1)s/\lambda}}{c_a^2+1}. \tag{3.23}$$

For complex $s$, the square root in (3.23) correspond to two complex roots, which are also the roots of $\left(\gamma - \sqrt{\gamma^2 + 2(c_a^2+1)s/\lambda}\right)/(c_a^2+1)$, since the polynomial in (3.22) is of order 2. Hence, we may use the same expression (3.23) as in the real case, as long as we pick the one with non-negative real part.

Combining (3.19) and (3.21), we obtain

$$\lim_{\rho\uparrow 1} \hat{H}_\rho(s) = -1,$$

and

$$\hat{H}^*(s) \equiv \lim_{\rho\uparrow 1} \frac{\hat{H}_\rho(s)+1}{(1-\rho)s} = \lim_{\rho\uparrow 1} \left( \frac{1}{\mu_\rho}\frac{1-\rho}{1-\delta} - \frac{1}{\mu_\rho}\frac{1-\rho}{1-\hat{\xi}\left((1-\rho)^2 s)\right)} + O(1-\rho) \right)$$

$$= \frac{1}{\lambda}\left( \frac{c_a^2+1}{2\gamma} - \frac{1}{\hat{\xi}^*(s)} \right) < \infty, \tag{3.24}$$

where $\hat{\xi}^*$ is defined in (3.21), so that

$$\hat{H}^*(s) = \frac{c_a^2+1}{2\lambda\gamma} s^2 \hat{h}^*(s) \tag{3.25}$$

for $\hat{h}^*(s)$ in (3.11). Moreover, we have

$$\lim_{\rho\uparrow 1} \frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} = -\hat{f}'(0) = E[U] = 1/\lambda.$$

and

$$\lim_{\rho\uparrow 1} \frac{\frac{1-\hat{f}\left((1-\rho)^2 s\right)}{(1-\rho)^2 s} - \frac{1}{\lambda}}{(1-\rho)^2 s} = -\frac{\hat{f}''(0)}{2} = -\frac{E[U^2]}{2} = -\frac{c_a^2+1}{2\lambda^2}.$$

Combining everything into the Laplace Transform of $(1-\rho)^2 V_{d,\rho}\left((1-\rho)^{-2}t\right)$, we have

$$\hat{V}_d^*(s) \equiv \lim_{\rho\uparrow 1} \hat{V}_{d,\rho}^*(s) = \frac{\lambda}{s^2} - \frac{\lambda(c_a^2-1)}{s^2}\left(\frac{2\gamma\lambda}{c_a^2+1}\hat{H}^*(s) - 1\right) \tag{3.26}$$

$$= \frac{\lambda}{s^2} - \frac{\lambda(c_a^2-1)}{s^2}\left(\frac{2\gamma\lambda}{c_a^2+1}\hat{H}^*(s) - 1\right) \tag{3.27}$$

$$= \frac{\lambda}{s^2} + \frac{2\lambda}{s^2}\frac{c_a^2-1}{c_a^2+1}\frac{\gamma}{\hat{\xi}^*(s)}. \tag{3.28}$$

Plugging in (3.23), we obtain

$$\begin{aligned}\hat{V}_d^*(s) &= \frac{\lambda}{s^2} + \frac{2\lambda}{s^2}\frac{c_a^2-1}{c_a^2+1}\frac{\gamma}{\hat{\xi}^*(s)} \\ &= \frac{\lambda}{s^2} + \frac{\lambda}{s^2}\frac{c_a^2-1}{c_a^2+1}\frac{\sqrt{1+2(c_a^2+1)s/(\lambda\gamma^2)}-1}{s/(\lambda\gamma^2)},\end{aligned}$$

where we pick the root such that $\left(\sqrt{1+2(c_a^2+1)s/(\lambda\gamma^2)}-1\right)/(s/(\lambda\gamma^2))$ has non-negative real part. We used the fact that $Re(z) \geq 0$ if and only if $Re(1/z) \geq 0$ for $z \neq 0$.

For the explicit inversion, one can exploit the LT of the correclation function in (3.7) and note that $\mathcal{L}(f(at))(s)$, for any constant $a \neq 0$ and any function $f$ with LT $\hat{f}$. For our case here, we use $a = \lambda\gamma^2/(c_a^2+1)$. ▪

Combining (3.8) and (3.17), we obtain the asymptotic behavior of the departure variance function.

Corollary 3.1. (*asymptotic behavior of the departure variance function*) *Under the assumptions in Theorem 3.1,*

$$\begin{aligned} V_d^*(t) &= c_a^2\lambda t + \frac{(c_s^2-c_a^2)c_x^2}{2\gamma^2} - \frac{(c_s^2-c_a^2)c_x^2}{2\gamma^2}c^*(\lambda\gamma^2 t/c_x^2) \\ &\sim c_a^2\lambda t + \frac{(c_s^2-c_a^2)c_x^2}{2\gamma^2} - \frac{8(c_s^2-c_a^2)c_x^5}{\gamma^5}\frac{1}{\sqrt{2\pi\lambda^3 t^3}}e^{-\frac{\lambda\gamma^2 t}{2c_x^2}} \quad \text{as} \quad t\to\infty, \end{aligned} \tag{3.29}$$

*where here* $c_s^2 = 1$.

Figure 1 (left) reports $V^*(t) - \lambda c_a^2 t$ for four sets of parameters such that the limiting constant $(1-c_a^4)/2\gamma^2$ in Corollary 3.1 will be $1.5, 0.375, -1.5$ and $-1.5$, respectively. Figure 1 (right) confirms Theorem 3.1 by comparing simulation estimates of the HT-scaled departure variance function $V_{d,\rho}^*(t) \equiv (1-\rho)^2 V((1-\rho)^{-2}t) - \lambda c_a^2 t$ for $\rho = 0.8$ and $0.9$ from simulation with the theoretical limit $V^*(t) - \lambda c_a^2 t$ for the $E_2/M/1$ model, where $c_a^2 = 0.5$, showing that the theoretical limit in (3.10) serves as a good approximation of the HT-scaled variance function.

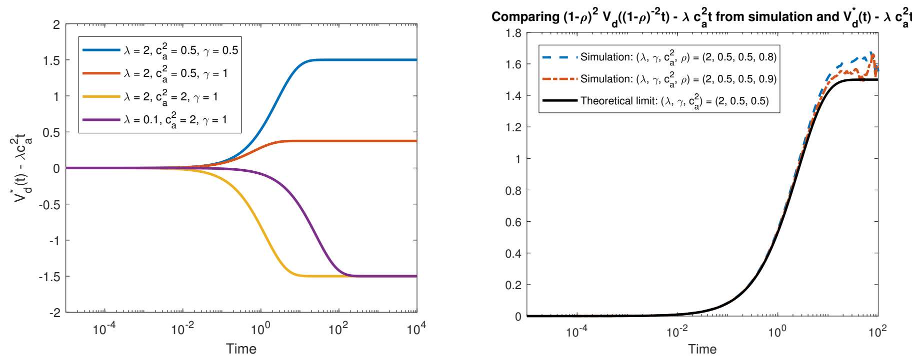


FIG 1. *On the left is $V^*(t) - c_a^2 t$ for four sets of parameters calculated from numerically inverting (3.10). On the right is the the HT-scaled variance $(1-\rho)^2 V((1-\rho)^{-2}t) - c_a^2 t$ for $\rho = 0.8$ and $0.9$ in the $E_2/M/1$ model, estimated by simulation, compared with the theoretical limit $V^*(t) - \lambda c_a^2 t$.*

**4. The Departure Variance in the $M/GI/1$ Queue.** In this section, we prove that the HT limit for the stationary departure variance in (3.14) also holds true for the $M/G/1$ model. Of course, here we restrict attention to $c_a^2 = 1$ instead of $c_s^2 = 1$ before. Theorem 5.3 will show that the same formula is valid for $GI/GI/1$ with general $c_a^2$ and $c_s^2$ (where both are not 0).

Recall from (2.13) that the Laplace Transform of the variance function of a general stationary and ergordic point process is

$$\hat{V}(s) = \frac{\lambda}{s^2} + \frac{2\lambda}{s}\hat{m}(s) - \frac{2\lambda^2}{s^3}.$$

In the case of the $M/GI/1$ model, [36] (on p. 78) derived an expression for $\hat{m}_d(s)$.

THEOREM 4.1. (*Laplace transform of the Palm mean function*) *For the departure process from a $M/G/1$ queue,*

$$\hat{m}_d(s) \equiv \int_0^\infty e^{-st} m_d(t)dt = \frac{\hat{g}(s)}{s(1-\hat{g}(s))}\left(1 - \frac{s\Pi(\hat{\nu}(s))}{s + \lambda(1-\hat{\nu}(s))}\right), \tag{4.1}$$

*where* $\hat{g}(s) = E\left[e^{-sV}\right]$ *is the Laplace Transform of the service pdf* $g(t)$, $\hat{\nu}(s)$ *is the root with the smallest absolute value in* $z$ *of the equation*

$$z = \hat{g}(s + \lambda(1 - z)) \tag{4.2}$$

*and*

$$\Pi(z) \equiv E\left[z^Q\right] = \frac{(1 - \lambda/\mu)(1 - z)\hat{g}(\lambda(1 - z))}{\hat{g}(\lambda(1 - z)) - z}. \tag{4.3}$$

*is the probability generating function of the distribution of the stationary queue length* $Q$.

Note from (2.2) that the first part in (4.1), i.e.

$$\frac{\hat{g}(s)}{s(1 - \hat{g}(s))},$$

is exactly the Laplace Transform of the mean process of the service renewal process.

Now, we state the HT limit in terms of the HT-scaled variance function defined in (3.5).

THEOREM 4.2. (*HT limit for the* $M/GI/1$ *departure variance*) *Consider an* $M/G/1$ *model with* $1/E[V] = \mu$ *and* $E[U] = \lambda_\rho \equiv \mu(1 - (1 - \rho)\gamma_\rho)$, *where* $\gamma_\rho$ *are positive constants such that* $\lim_{\rho \uparrow 1} \gamma_\rho = \gamma > 0$. *Assume that* $E[V^3] < \infty$ *so that a two-term Taylor series expansion of the LT* $\hat{g}(s)$ *about the origin is valid with asymptotically negligible remainder. Then*

$$V^*_{d,\rho}(t) \to V^*_d(t) \quad as \quad \rho \uparrow 1 \tag{4.4}$$

*for* $V^*_{d,\rho}(t)$ *in* (3.5), *where the limit* $V^*_d(t)$ *is a finite function with LT*

$$\hat{V}^*_d(s) = \frac{\mu c_s^2}{s^2} + \frac{\gamma\mu^2(1 - c_s^2)}{s^3}\hat{\nu}^*(s), \tag{4.5}$$

*where* $\hat{\nu}^*(s)$ *is the unique root with positive real part of the equation*

$$\frac{1 + c_s^2}{2}(\hat{\nu}^*(s))^2 + \gamma\hat{\nu}^*(s) - \frac{s}{\mu} = 0. \tag{4.6}$$

*Hence,* $V^*_d(t)$ *is again given by* (3.14), *with* $\mu$ *replacing* $\lambda$, *i.e.,*

$$V^*_d(t) \equiv w^*(\mu\gamma^2 t/c_x^2)c_a^2\mu t + \left(1 - w^*(\mu\gamma^2 t/c_x^2)\right) c_s^2\mu t, t \geq 0, \tag{4.7}$$

*with* $w^*(t)$ *in* (3.15).

*Proof of Theorem 4.2.* First, we derive the HT limit for the service variance function. Let $\mu = 1/E[V]$ be the service rate, then

$$\begin{aligned}
\hat{V}_s^*(s) = \lim_{\rho\uparrow 1} \hat{V}_{s,\rho}^*(s) &\equiv \lim_{\rho\uparrow 1} \mathcal{L}\left((1-\rho)^2 V_{s,\rho}\left((1-\rho)^{-2}t\right)\right) \\
&= \lim_{\rho\uparrow 1} (1-\rho)^4 \hat{V}_{s,\rho}\left((1-\rho)^2 t\right) \\
&= \lim_{\rho\uparrow 1} \left( \frac{\mu}{s^2} + \frac{2\mu}{s^2} \frac{\hat{g}\left((1-\rho)^2 t\right)}{1-\hat{g}\left((1-\rho)^2 t\right)} - \frac{2\mu^2}{(1-\rho)^2 s^3} \right) \\
&= \frac{\mu}{s^2} + \frac{2\mu^2}{s^2} \lim_{\rho\uparrow 1} \frac{1}{(1-\rho)^2 s} \left( \frac{\hat{g}\left((1-\rho)^2 t\right)}{\mu \frac{1-\hat{g}((1-\rho)^2 t)}{(1-\rho)^2 s}} - 1 \right) \\
&= \frac{\mu c_s^2}{s^2}.
\end{aligned}$$

Hence, we have

$$V_s^*(t) \equiv \lim_{\rho\uparrow 1} (1-\rho)^2 V_{s,\rho}\left((1-\rho)^{-2} t\right) = \mu c_s^2 t. \tag{4.8}$$

Now, we turn our focus to the departure process. To simplify the proof, we consider the HT-scaled difference between departure variance function and service variance function. Let $1/E[V] = \mu$ and $1/E[U] = \lambda_\rho \equiv \mu(1 - (1-\rho)\gamma_\rho)$, where $\gamma_\rho$ are positive constants such that $\lim_{\rho\uparrow 1} \gamma_\rho = \gamma > 0$. Under this setting, we have $(\lambda_\rho - \mu)/(1-\rho) \to -\mu\gamma$ as $\rho \uparrow 1$. Let $\hat{V}_{d,\rho}^*(s)$

and $\hat{V}^*_{s,\rho}(s)$ be the LT of $V^*_{d,\rho}(s)$ and $V^*_{s,\rho}(s)$, respectively. By (4.1), we have

$$\begin{aligned}
\hat{V}^*_d(s) - \hat{V}^*_s(s) &= \lim_{\rho\uparrow 1}\left(\hat{V}^*_{d,\rho}(s) - \hat{V}^*_{s,\rho}(s)\right) \\
&= \lim_{\rho\uparrow 1}(1-\rho)^4\left(\hat{V}_{d,\rho}\left((1-\rho)^2 s\right) - \hat{V}_{s,\rho}\left((1-\rho)^2 s\right)\right) \\
&= \lim_{\rho\uparrow 1}\left(\frac{\lambda_\rho-\mu}{s^2} + \frac{2(\lambda_\rho-\mu)}{s^2}\frac{\hat{g}\left((1-\rho)^2 s\right)}{1-\hat{g}\left((1-\rho)^2 s\right)} - \frac{2(\lambda_\rho^2-\mu^2)}{(1-\rho)^2 s^3}\right) \\
&\qquad - \lim_{\rho\uparrow 1}\left(\frac{2\lambda_\rho}{s^2}\frac{\hat{g}\left((1-\rho)^2 s\right)}{1-\hat{g}\left((1-\rho)^2 s\right)}\frac{(1-\rho)^2 s\Pi\left(\hat{\nu}\left((1-\rho)^2 s\right)\right)}{(1-\rho)^2 s + \lambda_\rho\left(1-\hat{\nu}\left((1-\rho)^2 s\right)\right)}\right) \\
&= \lim_{\rho\uparrow 1}\frac{\lambda_\rho-\mu}{s^2}\left(1 - 2(\lambda_\rho+\mu)\frac{1}{(1-\rho)^2 s}\left(1 - \frac{\hat{g}\left((1-\rho)^2 s\right)}{\mu\frac{1-\hat{g}((1-\rho)^2 s)}{(1-\rho)^2 s}}\right)\right) \\
&\qquad + \lim_{\rho\uparrow 1}\frac{2\lambda_\rho}{s^2}\frac{\hat{g}\left((1-\rho)^2 s\right)}{\frac{1-\hat{g}((1-\rho)^2 s)}{(1-\rho)^2 s}}\cdot\frac{1}{1-\rho}\left(\frac{\gamma_\rho}{s} - \frac{\Pi\left(\hat{\nu}\left((1-\rho)^2 s\right)\right)}{(1-\rho)s + \lambda_\rho\frac{1-\hat{\nu}((1-\rho)^2 s)}{1-\rho}}\right) \\
&= \hat{F}^{(1)}_\rho(s) + \frac{2\lambda_\rho\mu}{s^2}\frac{\hat{g}\left((1-\rho)^2 s\right)}{\mu\frac{1-\hat{g}((1-\rho)^2 s)}{(1-\rho)^2 s}}\frac{1}{(1-\rho)s + \lambda_\rho\frac{1-\hat{\nu}((1-\rho)^2 s)}{1-\rho}}\cdot\hat{F}^{(2)}_\rho(s)
\end{aligned}$$

where

$$\hat{F}^{(1)}_\rho(s) \equiv \frac{\lambda_\rho-\mu}{s^2}\left(1 - 2(\lambda_\rho+\mu)\frac{1}{(1-\rho)^2 s}\left(1 - \frac{\hat{g}\left((1-\rho)^2 s\right)}{\mu\frac{1-\hat{g}((1-\rho)^2 s)}{(1-\rho)^2 s}}\right)\right)$$

and

$$\hat{F}^{(2)}_\rho(s) \equiv \frac{\gamma_\rho}{1-\rho}\left(1-\rho + \frac{\lambda_\rho}{s}\frac{1-\hat{\nu}\left((1-\rho)^2 s\right)}{1-\rho} - \frac{1}{\gamma_\rho}\Pi\left(\hat{\nu}\left((1-\rho)^2 s\right)\right)\right).$$

One can easily show that $\hat{F}^{(1)}_\rho(s)$ converges to 0 as $\rho\uparrow 1$. Note also that $\hat{g}(0) = 1$ and $\hat{g}'(0) = -E[V] = -1/\mu$, then

$$\lim_{\rho\uparrow 1}\frac{\hat{g}\left((1-\rho)^2 s\right)}{\mu\frac{1-\hat{g}((1-\rho)^2 s)}{(1-\rho)^2 s}} = 1.$$

Furthermore, a Taylor series expansion around $s = 0$ yields

$$\begin{aligned}\frac{\hat{\nu}\left((1-\rho)^2 s\right) - 1}{1-\rho} &= \frac{\hat{g}\left((1-\rho)^2 s + \lambda_\rho\left(1 - \hat{\nu}\left((1-\rho)^2 s\right)\right)\right) - 1}{1-\rho} \\ &= -\frac{1-\rho}{\mu} s + \frac{\lambda_\rho}{\mu} \frac{\hat{\nu}\left((1-\rho)^2 s\right) - 1}{1-\rho} \\ &\quad + \frac{\hat{g}''(0) + o(1)}{2(1-\rho)} \left((1-\rho)^2 s + \lambda_\rho\left(1 - \hat{\nu}\left((1-\rho)^2 s\right)\right)\right)^2,\end{aligned}$$

which implies that

$$\lim_{\rho \uparrow 1} \hat{\nu}\left((1-\rho)^2 s\right) = 1 \tag{4.9}$$

and

$$\begin{aligned}0 &= -\frac{s}{\mu} + \frac{1 - \frac{\lambda_\rho}{\mu}}{1-\rho} \frac{1 - \hat{\nu}\left((1-\rho)^2 s\right)}{1-\rho} + \frac{\hat{g}''(0) + o(1)}{2(1-\rho)^2} \left((1-\rho)^2 s + \lambda_\rho\left(1 - \hat{\nu}((1-\rho)^2 s)\right)\right)^2 \\ &= -\frac{s}{\mu} + \gamma_\rho \frac{1 - \hat{\nu}\left((1-\rho)^2 s\right)}{1-\rho} + \frac{\hat{g}''(0) + o(1)}{2} \left((1-\rho) s + \lambda_\rho \frac{1 - \hat{\nu}((1-\rho)^2 s)}{1-\rho}\right)^2 \\ &= -\frac{s}{\mu} + \gamma_\rho \frac{1 - \hat{\nu}\left((1-\rho)^2 s\right)}{1-\rho} + \frac{\lambda_\rho^2}{\mu^2} \frac{c_s^2 + 1}{2} \left(\frac{1 - \hat{\nu}((1-\rho)^2 s)}{1-\rho}\right)^2 + o(1),\end{aligned}$$

where we used the fact that $\hat{g}''(0) = E[V^2] = (c_s^2 + 1)/\mu^2$. Hence,

$$\lim_{\rho \uparrow 1} \frac{1 - \hat{\nu}((1-\rho)^2 s)}{1-\rho} = \nu^*(s),$$

where

$$\frac{1 + c_s^2}{2} (\hat{\nu}^*(s))^2 + \gamma \hat{\nu}^*(s) - \frac{s}{\mu} = 0. \tag{4.10}$$

With essentially the same argument as in the proof of Theorem 3.1, one can also show that $\nu^*(s)$ is the only root of (4.10) with positive real part, furthermore

$$\nu^*(s) = \frac{-\gamma + \sqrt{\gamma^2 + 2(1 + c_s^2)s/\mu}}{1 + c_s^2}. \tag{4.11}$$

It remains to show that $\hat{F}_\rho^{(2)}(s)$ converges (pointwise) to a proper limit.

To this end, we write

$$\hat{F}_\rho^{(2)}(s) = \frac{\gamma_\rho}{1-\rho}\left(1-\rho+\frac{\lambda_\rho}{s}\frac{1-\hat{\nu}\left((1-\rho)^2 s\right)}{1-\rho} - \frac{1}{\gamma_\rho}\Pi\left(\hat{\nu}\left((1-\rho)^2 s\right)\right)\right)$$
$$= \gamma_\rho + \gamma_\rho \frac{1-\hat{\nu}\left((1-\rho)^2 s\right)}{1-\rho}\frac{1}{1-\rho}\left(\frac{\lambda_\rho}{s} - \frac{1}{\gamma_\rho}\frac{(1-\lambda_\rho/\mu)(1-\rho)\hat{g}\left(\lambda_\rho\left(1-\hat{\nu}\left((1-\rho)^2 s\right)\right)\right)}{\hat{g}(\lambda_\rho(1-\hat{\nu}\left((1-\rho)^2 s\right)) - \hat{\nu}\left((1-\rho)^2 s\right)}\right)$$
$$= \gamma_\rho + \gamma_\rho \frac{1-\hat{\nu}\left((1-\rho)^2 s\right)}{1-\rho}\frac{1}{1-\rho}\left(\frac{\lambda_\rho}{s} - \frac{(1-\rho)^2\hat{g}\left(\lambda_\rho\left(1-\hat{\nu}\left((1-\rho)^2 s\right)\right)\right)}{\hat{g}(\lambda_\rho(1-\hat{\nu}\left((1-\rho)^2 s\right)) - \hat{\nu}\left((1-\rho)^2 s\right)}\right)$$

Note that

$$\hat{g}\left(\lambda_\rho(1-\hat{\nu}\left((1-\rho)^2 s\right)\right) - \hat{\nu}\left((1-\rho)^2 s\right)$$
$$= \hat{g}\left(\lambda_\rho(1-\hat{\nu}\left((1-\rho)^2 s\right)\right) - \hat{g}\left((1-\rho)^2 s + \lambda_\rho(1-\hat{\nu}\left((1-\rho)^2 s\right)\right)$$
$$= (1-\rho)^2\frac{s}{\mu} - \hat{g}''(0)(1-\rho)^2 s\lambda_\rho\left(1-\hat{\nu}\left((1-\rho)^2 s\right)\right) + O((1-\rho)^4),$$

one can easily show that

$$\lim_{\rho\uparrow 1}\hat{F}_\rho^{(2)}(s) = \gamma - \gamma\hat{\nu}^*(s)\frac{\mu}{s}\left(1+c_s^2\hat{\nu}^*(s)\right). \tag{4.12}$$

Plugging everything into the Laplace Transform of the heavy-traffic scaled difference of the variance functions, we have

$$\hat{V}_d^*(s) = \hat{V}_s^*(s) + \frac{2\mu^2}{s^2}\frac{1}{\mu\hat{\nu}^*(s)}\left(\gamma-\gamma\hat{\nu}^*(s)\frac{\mu}{s}\left(1+c_s^2\hat{\nu}^*(s)\right)\right)$$
$$= \frac{\mu c_s^2}{s^2} + \frac{\gamma\mu^2(1-c_s^2)}{s^3}\hat{\nu}^*(s) \tag{4.13}$$

where we apply (4.10) to obtain the simplified expression in (4.13).

To obtain the explicit inversion, we write

$$\hat{V}_d^*(s) = \frac{\mu c_s^2}{s^2} + \frac{\gamma\mu^2(1-c_s^2)}{s^3}\frac{-\gamma+\sqrt{\gamma^2+2(1+c_s^2)s/\mu}}{1+c_s^2}.$$

Then, one exploit the LT of the correclation function in (3.7) and note that $\mathcal{L}(f(at))(s) = \hat{f}(t/a)/a$, for any constant $a \neq 0$ and any function $f$ with LT $\hat{f}$. For our case here, we use $a = \mu\gamma^2/(1+c_s^2)$. ∎

With the same technique as in Corollary 3.1, one can prove the following corollary, which yields exactly the same asymptotic behavior.

Corollary 4.1. (*asymptotic behavior of the departure variance curve*) *Under the assumptions in Theorem 4.2, we have the limit in* (3.29), *except now* $c_a^2 = 1$ *and* $c_s^2$ *is general.*

**5. Heavy-Traffic Limit for the Stationary Departure Process.** In this section, we establish an HT limit for the stationary departure process in a $GI/GI/1$ queue. To do so, we apply the recent HT results for the stationary queue length (number in system) in [22] and [11] together with the HT limits for the general single-server queue in §9.3 of [39] and the general reflection mapping with non-zero initial conditions in §13.5 of [39]. As in [39], a major component of the proof is the continuous mapping theorem.

The corresponding limit starting out empty is contained in Theorem 2 of [29]. There has since been a substantial literature on that case; see [23, 30, 39]. As can be seen from §9.3 and §13.5 of [39], for the queue length, the key map is the reflection map $\phi$ applied to a potential net-input function $x$,

$$\phi(x)(t) \equiv x(t) - \zeta(x)(t), \quad t \geq 0, \tag{5.1}$$

where

$$\zeta(x) \equiv \inf\{x(s) : 0 \leq s \leq t\} \wedge 0, \quad t \geq 0, \tag{5.2}$$

with $a \wedge b \equiv \min\{a, b\}$, so that $\zeta(x) \leq 0$ and $\phi(x)(t) \geq x(t)$ for all $t \geq 0$. The key point is that we now allow $x(0) \neq 0$,

5.1. *A General Heavy-Traffic Limit for the $G/G/1$ Model.* For the general $G/G/1$ single-server queue with unlimited waiting space and service provided in order of arrival, we consider a family of processes indexed by the traffic intensity $\rho$, where $\rho \uparrow 1$. Let $Q_\rho(t)$ be the number of customers in the system at time $t$; let $A_\rho(t)$ count the number of arrivals in the interval $[0, t]$; let $S_\rho(t)$ be a corresponding counting process for the successive service times, applied after time 0, to be applied to the initial $Q_\rho(0)$ customers and to all new arrivals; let $B_\rho(t)$ be the cumulative time that the server is busy in the interval $[0, t]$. Then the queue-length process can be expressed as

$$Q_\rho(t) \equiv Q_\rho(0) + A_\rho(t) - S_\rho(B_\rho(t)), \quad t \geq 0, \tag{5.3}$$

where the three components are typically dependent. (For simplicity, we assume that $A_\rho(0) = S_\rho(0) = B_\rho(0) = 0$ w.p.1.)

We have in mind that the system is starting in steady-state. Thus the triple $(Q_\rho(0), A_\rho(\cdot), S_\rho(\cdot))$ is in general quite complicated for each $\rho$. Even in the relatively tractable $GI/GI/1$ cases, which we shall primarily treat, the residual interarrival time and service time at time 0 will be complicated, depending on $\rho$ and $Q_\rho(0)$. We will need to make assumptions ensuring that these are uniformly asymptotically negligible in the HT limit.

By flow conservation, the departure (counting) process can be represented as

$$D_\rho(t) \equiv A_\rho(t) - Q_\rho(t) + Q_\rho(0), \quad t \geq 0. \tag{5.4}$$

Directly, or by combining (5.3) and (5.4),

$$D_\rho(t) \equiv S_\rho(B_\rho(t)), \quad t \geq 0. \tag{5.5}$$

Let

$$X_\rho(t) \equiv Q_\rho(0) + A_\rho(t) - S_\rho(t), \quad t \geq 0, \tag{5.6}$$

be a net-input process, acting as if the server is busy all the time, and thus allowing $X_\rho(t)$ to assume negative values. As a consequence of the assumptions above, $X_\rho(0) = Q_\rho(0)$. Roughly,

$$Q_\rho(t) \approx \phi(X_\rho)(t), \quad t \geq 0, \tag{5.7}$$

for $\phi$ in (5.1), but the exact relation breaks down because the service process shuts down when the system becomes idle, so that a new service time does not start until after the next arrival. While (5.7) does not hold exactly for each $\rho$, it holds in the HT limit, as shown in Theorem 9.3.4 of [39]. It would hold exactly if we used the modified system in which we let the continuous-time service process run continuously, so that equation (5.7) holds as an equality, as done by [10] and then again in §2 of [28]. Because the modified system has been shown to be asymptotically equivalent to the original system for these HT limits in [10] and [28], that is an alternate approach.

We now introduce HT-scaled versions of these processes, for that purpose, let

$$\begin{aligned}
X^*_\rho(t) &\equiv (1-\rho)X_\rho((1-\rho)^{-2}t),\\
Q^*_\rho(t) &\equiv (1-\rho)Q_\rho((1-\rho)^{-2}t),\\
A^*_\rho(t) &\equiv (1-\rho)[A_\rho((1-\rho)^{-2}t) - (1-\rho)^{-2}t],\\
S^*_\rho(t) &\equiv (1-\rho)[S_\rho((1-\rho)^{-2}t) - (1-\rho)^{-2}t/\rho],\\
B^*_\rho(t) &\equiv (1-\rho)[B_\rho((1-\rho)^{-2}t) - (1-\rho)^{-2}t],\\
D^*_\rho(t) &\equiv (1-\rho)[D_\rho((1-\rho)^{-2}t) - (1-\rho)^{-2}t],
\end{aligned} \tag{5.8}$$

Let $\mathcal{D}$ be the function space of all right-continuous real-valued functions on $[0,\infty)$ with left limits, with the usual $J_1$ mode of convergence, which reduces to uniform convergence over all bounded intervals for continuous limit functions. Let $\mathcal{D}^k$ be the $k$-fold product space, using the product topology on all product spaces. Let $\Rightarrow$ denote convergence in distribution. Let $e$ be the identity function in $\mathcal{D}$, i.e., $e(t) \equiv t$, $t \geq 0$.

THEOREM 5.1. *If*

$$(Q_\rho^*(0), A_\rho^*, S_\rho^*) \Rightarrow (Q^*(0), A^*, S^*) \quad in \quad \mathbb{R} \times \mathcal{D}^2 \quad as \quad \rho \uparrow 1, \tag{5.9}$$

*where $A^*$ and $S^*$ have continuous sample paths with $A^*(0) = S^*(0) = 0$ w.p.1., then*

$$(A_\rho^*, S_\rho^*, B_\rho^*, X_\rho^*, Q_\rho^*, D_\rho^*) \Rightarrow (A^*, S^*, B^*, X^*, Q^*, D^*) \quad in \quad \mathcal{D}^6 \quad as \quad \rho \uparrow 1, \tag{5.10}$$

*where*

$$\begin{aligned} B^* &\equiv \zeta(X^*) < 0, \quad X^* \equiv Q^*(0) + A^* - S^* - e, \\ Q^* &\equiv \phi(X^*) = X^* - \zeta(X^*) \quad and \\ D^* &\equiv X^* + e + \zeta(X^*) \\ &= Q^*(0) + A^* - Q^*(t) = Q^*(0) + A^* - \phi(X^*). \end{aligned} \tag{5.11}$$

*for $\phi$ and $\zeta$ in (5.1) and (5.2).*

*Proof.* First, note that

$$X_\rho^*(t) = Q_\rho^*(0) + A_\rho^*(t) - S_\rho^*(t) - t/\rho, \quad t \geq 0, \tag{5.12}$$

because $A_\rho^*$ and $S_\rho^*$ have different translation terms in (5.8), ensuring that the potential rate out is $1/\rho$, which exceeds the rate in of 1, consistent with a stable model for each $\rho$, $0 < \rho < 1$. Hence, under the assumption, $X_\rho^* \Rightarrow X^* = Q^*(0) + A^* - S^* - e$ in $\mathcal{D}$. The limits for $Q_\rho^*$ and $D_\rho^*$ then follow from the continuous mapping theorem after carefully accounting for the busy and idle time of the server; see the proof of Theorem 9.3.4 and preceding material in [39]. ∎

5.2. *A Heavy-Traffic Limit for the Stationary Processes.* Theorem 5.1 is not easy to apply to establish HT limits for stationary processes because condition (5.9) is not easy to check and the limit in (5.10) and (5.11) is not easy to evaluate.

In order to establish a tractable HT limit for the stationary departure process, we apply the recent HT limits for the stationary queue length in [22] and [11]. Their HT limits are for generalized open Jackson networks of queues, which for the single queue we consider reduce to the $GI/GI/1$ model. Following [11], we assume that the interarrival times and service times come from independent sequences of i.i.d. random variables with uniformly bounded third moments ($2 + \epsilon$ would do).

THEOREM 5.2. *For the $GI/GI/1$ model indexed by $\rho$, assume that (i) the interarrival-time cdf has a pdf as in* §2.1 *and (ii) the interarrival times and service times have means* 1 *and* $1/\rho$*, scv's* $c_a^2$ *and* $c_s^2$*, without both being* 0*, and uniformly bounded third moments. Then, for each* $\rho$*,* $0 < \rho < 1$*, the process* $Q_\rho^*$ *can be regarded as a stationary process, the process* $D_\rho^*$ *can be regarded as a stationary point process (with stationary increments), and condition* (5.9) *holds with*

$$A^* \equiv c_a B_a \quad \text{and} \quad S^* \equiv c_s B_s, \tag{5.13}$$

*where* $B_a$ *and* $B_s$ *are independent standard (mean 0, variance 1) Brownian motions (BM's) that are independent of* $Q^*(0)$*, which is distributed as* $R_e(0)$ *with* $R_e$ *being a stationary RBM with drift* $-1$ *and variance* $c_x^2 \equiv c_a^2 + c_s^2$*, and so an exponential marginal distribution, i.e.,*

$$P(Q^*(0) > x) = e^{-2x/c_x^2}, \quad x \geq 0. \tag{5.14}$$

*As a consequence, the limits in Theorem 5.1 hold, where*

$$X^*(t) \equiv Q^*(0) + c_a B_a(t) - c_s B_s(t) - t, \quad t \geq 0, \tag{5.15}$$

*with* $Q^*(0)$*,* $B_a$ *and* $B_s$ *being mutually independent, so that*

$$\begin{aligned} D^* &\equiv S^* + e + \zeta(X^*) \\ &= c_s B_s + e + \zeta(Q^*(0) + c_a B_a - c_s B_s - e) \end{aligned} \tag{5.16}$$

*for* $\zeta$ *in* (5.2) *or*

$$\begin{aligned} D^*(t) &= c_a B_a(t) + Q^*(0) - Q^*(t) \\ &= c_a B_a(t) + Q^*(0) - \phi(Q^*(0) + c_a B_a - c_s B_s - e)(t), \quad t \geq 0. \end{aligned} \tag{5.17}$$

*for* $\phi$ *in* (5.1)*.*

*Proof.* First, recall that the HT limit as $\rho \to 1$ starting empty is the RBM which converges as $t \to \infty$ to the exponential distribution in (5.14). We will be applying [22] and [11] to show that the two iterated limits involving $\rho \to 1$ and $t \to \infty$ are equal. Toward that end, we observe that, by §§X.3-X.4 of [7], the queue-length process has a proper steady-state distribution for each $\rho$. As in [22] and [11], we add the residual interarrival times and service times to the state description for $Q_\rho(t)$ to make it a Markov process that has a unique steady-state distribution for each $\rho$. These residual interarrival and service times are asymptotically negligible in the HT limit. The associated departure process $D_\rho(t)$ then necessarily is a stationary point process for each $\rho$. We then can apply [22] and [11] to have a limit for the scaled stationary distributions, so that condition (5.9) holds with (5.13). Hence, we can apply Theorem 5.1 with these special initial distributions to get the associated process limits in the space $\mathcal{D}$. ∎

We now establish an HT limit for the variance of the stationary departure process. The form of that limit is already given in Theorem 3.1.

THEOREM 5.3. (*limiting variance*) *Under the conditions of Theorem 5.2 plus the usual uniform integrability conditions, for which it suffices for the interarrival times and service times to have uniformly bounded fourth moments,*

$$\begin{aligned} V_{d,\rho}^*(t) &\equiv \mathrm{Var}(D_\rho^*(t)) = E[D_\rho^*(t)^2] \\ &\to E[D^*(t)^2] = \mathrm{Var}(D^*(t)) \equiv V_d^*(t) \quad \textit{as} \quad \rho \uparrow 1, \end{aligned} \tag{5.18}$$

*where*

$$\begin{aligned} V_d^*(t) &= c_a^2 t + \frac{c_x^4}{2}\left(1 - c^*(t/c_x^2)\right) - 2\mathrm{Cov}(c_a B_a(t), Q^*(t)), \\ &= w^*(t/c_x^2) c_a^2 t + (1 - w^*(t/c_x^2)) c_s^2 t, \end{aligned} \tag{5.19}$$

*with* $c_x^2 = c_a^2 + c_s^2$, $c^*(t)$ *is the correlation function in* (3.6) *and* $w^*(t)$ *is the weight function in* (3.15)*; i.e.,* $V_d^*(t)$ *is given in* (3.14) *with* $\lambda = \gamma = 1$*, but allowing general* $c_s^2$*. Moreover, we have the covariance formulas:*

$$\mathrm{Cov}(c_a B_a(t), Q^*(t)) = \frac{c_a^2 c_x^2}{2}\left(1 - c^*(t/c_x^2)\right),$$

$$\mathrm{Cov}(c_s B_s(t), Q^*(t)) = \frac{-c_s^2 c_x^2}{2}\left(1 - c^*(t/c_x^2)\right)$$

*and*

$$\mathrm{Cov}(X^*, Q^*(t)) = \frac{c_x^4}{4}\left(2 - c^*(t/c_x^2)\right) = 2\mathrm{Var}(Q^*(0)) - \mathrm{Cov}(Q^*(0), Q^*(t)), \quad t \ge 0.$$

*Proof.* By combining Theorems 2.1 and 4.2 in Chapter X of [7], we deduce that the $k^{th}$ moment of the steady-state queue length in finite if the $(k+1)^{st}$ moment of the service time are finite. We add the extra uniformly bounded fourth moment to provide the uniform integrability needed to get convergence of the moments in the HT limit. We use (5.4) to obtain the corresponding result for the departure process.

To get (5.19), combine (5.18) and (5.16). Note that

$$\mathrm{Var}(Q^*(t)) = \mathrm{Var}(Q^*(0)) = c_x^2/4,$$

so that

$$\text{(5.20)}\quad \mathrm{Var}(D^*(t)) = c_a^2 t + \frac{c_x^4}{2} - 2\mathrm{Cov}(Q^*(0), Q^*(t)) - 2\mathrm{Cov}(c_a B_a(t), Q^*(t)),$$

where

$$\text{(5.21)}\qquad \mathrm{Cov}(Q^*(0), Q^*(t)) = \frac{c_x^4}{4} c^*(t/c_x^2), \quad t \ge 0;$$

see §2 of [2] or Theorem 5.7.11 of [39]. Inserting (5.21) into (5.20) yields the first line in (5.19) above. To establish the second limit, we do a space-time transformation of the limit, so that the limit is the same as one of the models analyzed directly.

Let us rescale space and time so that the general result is in terms on $B_a$ instead of $c_a B_a$ (assuming that $c_a > 0$), so that we can apply the established result for the $M/GI/1$ model. (Essentially the same argument works for $GI/M/1$.) The first step is to observe that the HT limit for the departure process $\{D^*(t) : t \ge 0\}$ can be written as a function $\Psi : \mathbb{R} \times \mathcal{D}^2 \to \mathcal{D}$ of the vector process $\{(Q^*(0), c_a B_a(t), c_s B_s(t), -t) : t \ge 0\}$; i.e.,

$$\text{(5.22)}\qquad D^* = \Psi((Q^*(0), c_a B_a, c_s B_s, -e))$$

or, by (5.16),

$$\begin{aligned} \{D^*(t) : t \ge 0\} &= \{\Psi((Q^*(0), c_a B_a, c_s B_s, -e))(t) : t \ge 0\} \\ \text{(5.23)}\qquad &= \{Q^*(0) + c_a B_a(t) - \phi(Q^*(0) + c_a B_a - c_s B_s - e)(t) : t \ge 0\}. \end{aligned}$$

If we replace the basic vector process $(Q^*(0), c_a B_a, c_s B_s, -e)$ by another that has the same distribution as a process, then the distribution of $D^* \equiv \{D^*(t) : t \ge 0\}$ will be unchanged.

By the basic time and space scaling of BM, for $c_a > 0$, the stochastic processes have equivalent distributions as follows

$$\begin{aligned} \{Q^*(0), c_a B_a(t), c_s B_s(t), -t\} &\stackrel{\mathrm{d}}{=} c_a^2 \left\{ \frac{Q^*(0)}{c_a^2}, B_a(t/c_a^2), \frac{c_s}{c_a} B_s(t/c_a^2), -\frac{t}{c_a^2} \right\} \\ \text{(5.24)}\qquad &\equiv c_a^2 \left\{ \frac{Q^*(0)}{c_a^2}, B_a(u), \frac{c_s}{c_a} B_s(u), -u \right\}, \end{aligned}$$

where $u = t/c_a^2$. After this transformation, to describe the system at time $u = c_a^2 t$, the associated RBM has drift $-1$ and variance coefficient $1 + (c_s^2/c_a^2) = c_x^2/c_a^2$. Note that the mean of the steady-state distribution associated with the new RBM is the diffusion coefficient divided by twice the absolute value of the drift, which is $c_x^2/(2c_a^2)$. As a result, $Q^*(0)/c_a^2$ is exactly the steady-state distribution needed for the new RBM. From above, we see that

$$\begin{aligned} D^*(t) &\stackrel{\mathrm{d}}{=} \Psi\left(c_a^2\left\{Q^*(0)/c_a^2, B_a(u), (c_s/c_a)B_s(u), -u\right\}\right), \quad \text{for} \quad u = t/c_a^2 \\ &= c_a^2\Psi\left(\left\{Q^*(0)/c_a^2, B_a(u), (c_s/c_a)B_s(u), -u\right\}\right), \\ &\equiv c_a^2\tilde{D}^*(u) = c_a^2\tilde{D}^*(t/c_a^2). \end{aligned}$$

where $\tilde{D}^*(u) \equiv \Psi\left(\left\{Q^*(0)/c_a^2, B_a(u), (c_s/c_a)B_s(u), -u\right\}\right)$, corresponding to the $M/GI/1$ model with service scv $c_s^2/c_a^2$. Now, let $\tilde{w}^*(t)$ denote the associated weight function in (4.7) with $(\mu, \gamma, \tilde{c}_x^2) = (1, 1, c_x^2/c_a^2)$, so that

$$\tilde{w}^*(t) = w^{**}(c_a^2 t/c_x^2).$$

We now turn to the variance. By applying (4.7), we obtain

$$\begin{aligned} V_d^*(t) = c_a^4\tilde{V}_d^*(u) &= c_a^4\tilde{V}^*(t/c_a^2) \\ &= c_a^4\left(\tilde{w}^*(t/c_a^2)\frac{t}{c_a^2} + (1-\tilde{w}^*(t/c_a^2))\frac{c_s^2}{c_a^2}\frac{t}{c_a^2}\right) \\ &= w^{**}(t/c_x^2)c_a^2 t + (1-w^{**}(t/c_a^2))c_s^2 t. \end{aligned}$$

which agrees with the $GI/GI/1$ formula in (5.19). Thus, we have proved the variance formula for $GI/GI/1$.

Finally, it remains to establish the covariance formulas. First, by comparing the two lines in (5.19), we must have

$$\mathrm{Cov}(c_a B_a(t), Q^*(t)) = \frac{c_a^2 c_x^2}{2}\left(1 - c^*(t/c_x^2)\right).$$

Let $\tilde{B}_s(t) = -B_s(t)$, then we have

$$(Q^*(0), c_a B_a, c_s B_s) \stackrel{\mathrm{d}}{=} (Q^*(0), c_a B_a, c_s \tilde{B}_s),$$

so

$$\begin{aligned} \mathrm{Cov}(c_a B_a(t), Q^*(t)) &= \mathrm{Cov}(c_a B_a(t), \phi(Q^*(0) + c_a B_a + c_s\tilde{B}_s - e)(t)) \\ &= \mathrm{Cov}(c_a B_a(t), \phi(Q^*(0) + c_a B_a + c_s B_s - e)(t)), \end{aligned}$$

and

$$\begin{aligned} \mathrm{Cov}(c_s B_s(t), Q^*(t)) &= \mathrm{Cov}(-c_s\tilde{B}_s(t), \phi(Q^*(0) + c_a B_a + c_s\tilde{B}_s - e)(t)) \\ &= -\mathrm{Cov}(c_s B_s(t), \phi(Q^*(0) + c_a B_a + c_s B_s - e)(t)) \end{aligned}$$

By symmetry, we thus have

$$\mathrm{Cov}(c_s B_s(t), Q^*(t)) = -\frac{c_s^2 c_x^2}{2}(1 - c^*(t/c_x^2)).$$

Then

$$\begin{aligned}\mathrm{Cov}(X^*(t), \phi(X^*)(t)) &= \mathrm{Cov}(X^*(t), Q^*(t)) \\ &= \mathrm{Cov}(Q^*(0) + c_a B_a(t) - c_s B_s(t), Q^*(t)) \\ &= \frac{c_x^4}{4} c^*(t/c_x^2) + \frac{c_a^2 c_x^2}{2}(1 - c^*(t/c_x^2)) + \frac{c_s^2 c_x^2}{2}(1 - c^*(t/c_x^2)) \\ &= \frac{c_x^4}{4} + \frac{c_x^4}{4}(1 - c^*(t/c_x^2)). \quad \blacksquare\end{aligned}$$

REMARK 5.1. (*the quasireversible case*) The limit process $(A^*, S^*, X^*, Q^*, D^*)$, where

$$(A^*, S^*, X^*) = (c_a B_a, c_s B_s, Q^*(0) + c_a B_a - c_s B_s - \eta e),$$

as in Theorem 5.2, can be called *the Brownian queue*; see [24, 25, 26, 33]. The Brownian queue is known to be quasireversible if and only if $c_a^2 = c_s^2$. In that case, the stationary departure process is a BM and the departures in the past are independent of the steady-state content. Consistent with that theory, $V_d^*(t) = c_a^2 t, t \geq 0$ in (5.19) if and only if $c_a^2 = c_s^2$.

**6. Application to a Robust Queueing Network Analyzer.** We conclude by explaining the important role that Theorem 5.3 plays in our *Robust Queueing Network Analyzer* (RQNA) based on the *index of dispersion for counts* (IDC), which we refer to as RQNA-IDC. In §6.1 we briefly review the *robust queueing* (RQ) approximation for the mean steady-state workload at a $G/GI/1$ queue developed in [40], which requires the IDC of the arrival process as model data. Then, in §6.2 we review the approximation for the IDC of the departure process that we propose in [42], which is supported by this paper.

6.1. *The Mean Steady-State Workload at a $G/GI/1$ Queue.* In this section we review the RQ approximation for the mean steady-state workload at a $G/G/1$ queue developed in [40]. We start with a rate-1 stationary arrival process $A \equiv \{A(t) : t \geq 0\}$, having stationary increments. Thus, for a renewal arrival process, we work with the associated equilibrium renewal process. We also start with an stationary and ergodic sequence of mean-1 service times $\{V_k : k \geq 1\}$, possibly correlated with the arrival process. Each service time is distributed as a random variable $V$ with cdf $G$ having mean $E[V] \equiv \mu^{-1}$ and finite scv $c_s^2$. Let $Y(t)$ be the associated rate-1 total input process, defined by

$$Y(t) \equiv \sum_{k=1}^{A(t)} V_k, \quad t \geq 0, \tag{6.1}$$

which also has stationary increments. To specify associated queueing processes, we introduce the traffic intensity $\rho$, $0 < \rho < 1$. We keep the service rate at 1, but we make the arrival rate equal to $\rho$ in model $\rho$ by letting

$$A_\rho(t) \equiv A(\rho t), t \geq 0, \quad \text{so that} \quad Y_\rho(t) \equiv Y(\rho t), \quad t \geq 0. \tag{6.2}$$

Let $Z_\rho(t)$ be the workload at time $t$ in model $\rho$ with arrival rate $\rho$ and service rate 1, then

$$\begin{aligned} Z_\rho(t) = \phi(Y_\rho - e)(t) &= Y_\rho(t) - t - \inf_{s \leq t}\{Y_\rho(s) - s\} \\ &= \sup_{s \leq t}\{Y_\rho(t) - Y_\rho(s) - (t - s)\}. \end{aligned}$$

As in §3.1 of [40], we can apply a reverse-time construction to write the steady-state workload $Z_\rho$ as a simple supremum.

For the RQ approximation, we assume that the rate-1 arrival process $A$ is partially characterized by its *index of dispersion for counts* (IDC) $I_a(t)$, where

$$I_a(t) \equiv \frac{\mathrm{Var}(A(t))}{E[A(t)]}, \quad t \geq 0; \tag{6.3}$$

see §4.5 of [14]. Hence,

$$I_{a,\rho}(t) = I_a(\rho t), \quad t \geq 0. \tag{6.4}$$

For the RQ approximation of the mean steady-state workload, we use the associated *index of dispersion for work* (IDW) $I_w \equiv \{I_w(x) : x \geq 0\}$ introduced by [19], which is defined by

$$I_w(t) \equiv \frac{\mathrm{Var}(Y(t))}{E[V]E[Y(t)]} = \frac{\mathrm{Var}(Y(t))}{t}, \quad t \geq 0; \tag{6.5}$$

see §4.3 of [40] for key properties.

Let $Z_\rho \equiv Z(\rho, I_a, c_s^2)$ denote the steady-state workload in the $G/GI/1$ model when the arrival rate is $\rho$, the (rate-1) arrival process IDC is $I_a$ as in (6.3) and service times are i.i.d. with mean 1 and scv $c_s^2$. In this context,

$$I_w(t) = I_a(t) + c_s^2, \quad t \geq 0, \tag{6.6}$$

as noted in §4.3.1 of [40], and the RQ approximation (based on this partial model characterization) is

$$E[Z_\rho] \equiv E[Z(\rho, I_a, c_s^2)] \approx Z_\rho^* \equiv \sup_{x\geq 0}\{-(1-\rho)x/\rho + \sqrt{2xI_w(x)}\}, \tag{6.7}$$

for the IDW in (6.6). The approximation (6.7) comes from (28) of [40], assuming that we set the parameter $b_f = \sqrt{2}$, which makes the approximation asymptotically correct for the $GI/GI/1$ model in both the heavy-traffic and light-traffic limits; see Theorem 5 of [40]. Notice that the approximation in (6.7) is directly a supremum of a real-valued function, and so can be computed quite easily for any given triple $(\rho, I_a, c_s^2)$.

6.2. *The IDC of a Stationary Departure Process.* The main challenge in developing a full RQNA-IDC involving a decomposition approximation is calculating or approximating the required IDC for the arrival process at each queue. For a renewal arrival process, the IDC $I_a$ can be computed by inverting the LT $\hat{V}(s)$ in (2.7) using the LT $\hat{m}(s)$ of the renewal function $m(t)$ in (2.2), which only requires the LT $\hat{f}(s)$ of the interarrival-time pdf in (2.1). Numerical algorithms for calculating and simulations algorithms for estimating the IDC are discussed in [41].

As discussed in §6 of [40], this approximation is not difficult for the network operations of superposition and independent splitting. Hence, the main challenge becomes approximating the IDC of a departure process from a $G/GI/1$ queue, partially specified by the triple $(\rho, I_a, c_s^2)$. As in §6 of [40], we propose approximating the IDC of the departure process, $I_{d,\rho}(t)$ by the weighted average of the IDC's of the arrival and service processes, i.e.,

$$I_{d,\rho}(t) \approx w_\rho(t)I_a(t) + (1 - w_\rho(t))I_s(t), \quad t \geq 0, \tag{6.8}$$

where we assume that $I_a$ has been determined. However, we now need to know the IDC of the service process in addition to the triple $(\rho, I_a, c_s^2)$. Thus, for the $G/GI/1$ model we assume that we have the full-service-time cdf $G$, and consider the model data to $(\rho, I_a, G)$, from the cdf $G$ we can then calculate the IDC of the associated renewal process generated by the service times to get $I_s$. For a more general $G/G/1$ queue, we would assume that the service process is independent of the arrival process and that the IDC $I_s$ were given directly.

Based on initial study, a candidate weight function $w_\rho(t)$ was suggested in (43) of [40], but based on further study we propose the new weight function

$$w_\rho(t) \equiv w^*((1-\rho)^2 t/c_x^2), t \geq 0, \tag{6.9}$$

where $c_x^2 \equiv c_a^2 + c_s^2$ and $w^*$ is given in (3.15) with $c^*$ in (3.6); i.e., $w^*$ is obtained from the heavy-traffic limit of the HT-scaled variance of the stationary departure process from the $GI/GI/1$ queue in Theorem 5.3.

Thus to support the new RQNA-IDC, we can apply the following corollary to Theorem 5.3.

COROLLARY 6.1. (*asymptotically correct in HT*) *The IDC approximation for $I_d(t)$ in (6.8) with the weight function $w_\rho(t)$ in (6.9) is asymptotically correct as $\rho \uparrow 1$; i.e.,*

$$I_{d,\rho}((1-\rho)^{-2}t) \to w^*(t/c_x^2)c_a^2 + (1 - w^*(t/c_x^2))c_s^2 \quad \textit{as} \quad \rho \uparrow 1, \tag{6.10}$$

*consistent with Theorem 5.3.*

*Proof.* By the IDC definitions in (6.3) and (6.4),

$$\begin{aligned} I_{d,\rho}((1-\rho)^{-2}t) &= \frac{V_{d,\rho}((1-\rho)^{-2}t)}{(1-\rho)^{-2}t} \\ &= \frac{V_{d,\rho}^*(t)}{t} \to \frac{V_d^*(t)}{t} \\ &= w^*(t/c_x^2)c_a^2 + (1 - w^*(t/c_x^2))c_s^2 \quad \text{as} \quad \rho \uparrow 1. \end{aligned} \tag{6.11}$$

applying Theorem 5.3 in the second line. ∎

Hence, for $\rho$ not too small,

$$I_{d,\rho}((1-\rho)^{-2}t) \approx w^*(t/c_x^2)c_a^2 + (1 - w^*(t/c_x^2))c_s^2 \quad \text{for all} \quad t \tag{6.12}$$

and

$$I_{d,\rho}(t) \approx w^*((1-\rho)^2 t/c_x^2)c_a^2 + (1 - w^*((1-\rho)^2 t/c_x^2))c_s^2 \quad \text{for all} \quad t. \tag{6.13}$$

Results of simulation experiments to evaluate the approximation are reported in [42].

**7. Extensions.** The approximation for the departure IDC $I_d(t)$ in (6.8) and (6.9) should be good for much more general models than $GI/GI/1$, with the independence conditions relaxed and more than 1 server. We also conjecture that the HT limit of the variance function in Theorem 5.3 extends to a larger class of models as well. Indeed, we conjecture that the limits established for $GI/GI/1$ extend in that way. First, Theorem 5.1 extends quite directly by exploiting [28, 29]. For the extension of Theorem 5.2, there is a large class of models for which the HT-scaled arrival and service processes have the limits

$$A^* \equiv c_a B_a \quad \text{and} \quad S^* \equiv c_s B_s, \tag{7.1}$$

where $B_a$ and $B_s$ are independent standard (mean 0, variance 1) Brownian motions (BM's) that are independent of the initial queue length. What is needed is the extension of [22] and [11] to more general models. We conjecture that can be done for $GI/GI/s$ and other models with regenerative structure in the arrival and service processes. For $GI/GI/s$ the queue-length process again becomes a Markov process if we append the $s$ elapsed service times as well as the elapsed interarrival time, but it remains to do the hard technical analysis leading to an appropriate Lyapunov function.

It is also of interest to establish related results for departure processes in models with non-renewal arrival processes, as in [20] and references therein. It also remains to establish new HT limits for stationary departure processes from a queue within a network, obeying the HT FCLT in [34].

The relevant approximation for the stationary departure process from a many-server $GI/GI/s$ queue evidently is quite different, being more like the service process than the arrival process. We conjecture that the relevant many-server heavy-traffic limit for the stationary departure process is a Gaussian process with the covariance function of the stationary renewal processes associated with the service times, as in the CLT for renewal processes in Theorems 7.2.1 and 7.2.4 of [39]. Partial support comes from [5], Appendix F of [6] and [21].

**Acknowledgements.** This research was supported by NSF grant CMMI 1634133.

## REFERENCES

[1] J. Abate and W. Whitt. Transient behavior of regulated Brownian motion I: Starting at the origin. *Advances in Applied Probability*, 19(3):560–598, 1987.

[2] J. Abate and W. Whitt. The correlation functions of RBM and $M/M/1$. *Stochastic Models*, 4(2):315–359, 1988.

[3] J. Abate and W. Whitt. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10:5–88, 1992.

[4] J. Abate and W. Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, 7:36–43, 1995.

[5] A. K. Aras, C. Chen, and Y. Liu. Many-server Gaussian limits for overloaded queues with customer abandonment and nonexponential service. North Carolina State University, https://yunanliu.wordpress.ncsu.edu/, 2017.

[6] K. Aras, Y. Liu, and W. Whitt. Heavy-traffic limit for the initial content process,. *Stochastic Systems*, 7(1):95–142, 2017.

[7] S. Asmussen. *Applied Probability and Queues*. Springer, New York, second edition, 2003.

[8] C. Bandi, D. Bertsimas, and N. Youssef. Robust queueing theory. *Operations Research*, 63(3):676–700, 2015.

[9] D. Bertsimas and D. Nakazato. The departure process from a $GI/G/1$ queue and its applications to the analysis of tandem queues. Operations Research Center Report OR 245-91, MIT¡ Cambridge, MA, 1990.

[10] A. A. Borovkov. Some limit theorems in the theory of mass service, II. *Theor. Probability Appl.*, 10:375–400, 1965.

[11] A. Budhiraja and C. Lee. Stationary distribution convergence for generalized Jackson networks in heavy traffic. *Mathematics of Operations Research*, 34(1):45–56, 2009.

[12] P. J. Burke. The output of a queueing system. *Operations Research*, 4(6):699–704, 1956.

[13] J. W. Cohen. *The Single Server Queue*. North-Holland, Amsterdam, second edition, 1982.

[14] D. R. Cox and P. A. W. Lewis. *The Statistical Analysis of Series of Events*. Methuen, London, 1966.

[15] D. J. Daley. Weakly stationary point processes and random measures. *Journal of the Royal Statistical Society*, 33(3):406–428, 1971.

[16] D. J. Daley. Further second-order properties of certain single-server queueing systems. *Stoch. Proc. Appl.*, 3(2):185–191, 1975.

[17] D. J. Daley. Queueing output processes. *Adv. Appl. Prob.*, 8(2):395–415, 1976.

[18] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer, Oxford, U. K., second edition, 2008.

[19] K. W. Fendick and W. Whitt. Measurements and approximations to describe the offered traffic and predict the average workload in a single-server queue. *Proceedings of the IEEE*, 71(1):171–194, 1989.

[20] H.-W. Ferng and J.-F. Chang. Departure processes of $BMAP/G/1$ queues. *Queueing Systems*, 39(1):109–135, 2001.

[21] D. Gamarnik and D. A. Goldberg. Steady-state $GI/GI/n$ queue in the Halfin-Whitt regime. *Ann. Appl. Probability*, 23(6):2382–2419, 2013.

[22] D. Gamarnik and A. Zeevi. Validity of heavy traffic steady-state approximations in generalized Jackson networks. *Advances in Applied Probability*, 16(1):56–90, 2006.

[23] A. Al Hanbali, M. Mandjes, Y. Nazarathy, and W. Whitt. The asymptotic variance of departures in critically loaded queues. *Adv. Appl. Prob.*, 43(1):243–263, 2011.

[24] J. M. Harrison. *Brownian Motion and Stochastic Flow Systems*. Wiley, New York, 1985.

[25] J. M. Harrison and R. J. Williams. On the quasireversibility of a multiclass Brownian service station. *Annals of Probability*, 18(3):1249–1268, 1990.

[26] J. M. Harrison and R. J. Williams. Brownian models of feedforward queueing networks: quasireversibility and product-form solutions. *Ann. Appl. Prob.*, 2(2):263–293, 1992.

[27] J.Q. Hu. The departure process of the $GI/G/1$ queue and its Maclaurin series. *Operations Research*, 44(5):810–815, 1996.

[28] D. L. Iglehart and W. Whitt. Multiple channel queues in heavy traffic, I. *Advances in Applied Probability*, 2(1):150–177, 1970.

[29] D. L. Iglehart and W. Whitt. Multiple channel queues in heavy traffic, II: Sequences, networks and batches. *Advances in Applied Probability*, 2(2):355–369, 1970.

[30] F. I. Karpelevich and A. Ya. Kreinin. *Heavy-Traffic Limits for Multiphase Queues*, volume 137. American Mathematical Society, Providence, RI, 1994.

[31] S. Kim. Modeling cross correlation in three-moment four-parameter decomposition approximation of queueing networks. *Operations Research*, 59(2):480–497, 2011.

[32] S. Kim. The two-moment three-parameter decomposition approximation of queueing networks with exponential residual renewal processes. *Queueing Systems*, 68:193–216, 2011.

[33] N. O'Connell and M. Yor. Brownian analogues of Burke's theorem. *Stoch. Proc. Appl.*, 96:285–304, 2001.

[34] M. I. Reiman. Open queueing networks in heavy traffic. *Math. Oper. Res.*, 9(3):441–458, 1984.

[35] K. Sigman. *Stationary Marked Point Processes: An Intuitive Approach.* Chapman and Hall/CRC, New York, 1995.

[36] L. Takacs. *Introduction to the Theory of Queues.* Oxford University Press, 1962.

[37] W. Whitt. The queueing network analyzer. *Bell Laboratories Technical Journal*, 62(9):2779–2815, 1983.

[38] W. Whitt. Variability functions for parametric-decomposition approximations of queueing networks. *Management Science*, 41(10):1704–1715, 1995.

[39] W. Whitt. *Stochastic-Process Limits.* Springer, New York, 2002.

[40] W. Whitt and W. You. Using robust queueing to expose the impact of dependence in single-server queues. Operations Research, published online, July 31, 2017, doi:10.1287/opre.2017.1649, 2017.

[41] W. Whitt and W. You. Algorithms to compute the index of dispersion of a stationary point process. in preparation, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html, 2018.

[42] W. Whitt and W. You. A robust queueing network analyzer based on indices of dispersion. in preparation, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html, 2018.

[43] Q. Zhang, A. Heindle, and E. Smirni. Characterizing the $BMAP/MAP/1$ departure process via the ETAQA truncation. *Stochastic Models*, 21(2-3):821–846, 2005.

Department of IEOR,
S. W. Mudd Building,
500 West 120th Street,
New York, NY 10027-6699
E-mail: ww2040@columbia.edu
wy2225@columbia.edu