Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 121 (2011) 1524–1545
stochastic processes and their applications
www.elsevier.com/locate/spa

Transient behavior of the Halfin–Whitt diffusion

Johan S.H. van Leeuwaarden[a,*], Charles Knessl[b]

[a] Eindhoven University of Technology and EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[b] University of Illinois at Chicago, Department of Mathematics, Statistics and Computer Science, 815 South Morgan Street, Chicago, IL 60607-7045, USA

Received 13 October 2010; received in revised form 1 March 2011; accepted 9 March 2011
Available online 21 March 2011

Abstract

We consider the heavy-traffic approximation to the $GI/M/s$ queueing system in the Halfin–Whitt regime, where both the number of servers s and the arrival rate λ grow large (taking the service rate as unity), with $\lambda = s - \beta\sqrt{s}$ and β some constant. In this asymptotic regime, the queue length process can be approximated by a diffusion process that behaves like a Brownian motion with drift above zero and like an Ornstein–Uhlenbeck process below zero. We analyze the transient behavior of this hybrid diffusion process, including the transient density, approach to equilibrium, and spectral properties. The transient behavior is shown to depend on whether β is smaller or larger than the critical value $\beta_* \approx 1.85722$, which confirms the recent result of Gamarnik and Goldberg (2008) [8].

MSC: 60K25; 60J60; 60J70; 34E05

Keywords: $GI/M/s$ queue; $M/M/s$ queue; Halfin–Whitt regime; Queues in heavy traffic; Diffusion; Asymptotic analysis

1. Introduction

Halfin and Whitt [13] introduced in their 1981 paper a new heavy-traffic limit theorem for the $GI/M/s$ system. They demonstrated how under certain conditions a sequence of normalized queue-length processes converges to a process that behaves like a Brownian motion with drift

* Corresponding author.
E-mail addresses: j.s.h.v.leeuwaarden@tue.nl (J.S.H. van Leeuwaarden), knessl@uic.edu (C. Knessl).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2011.03.007

above zero and like an Ornstein–Uhlenbeck process below zero. We refer to this hybrid diffusion process as the *Halfin–Whitt diffusion*. Our concern is with the transient behavior of this diffusion.

What is nowadays known as the Halfin–Whitt regime refers to the scaling of the arrival rate λ and the numbers of servers s such that, while both λ and s increase toward infinity, the traffic intensity $\rho = \lambda/s$ approaches one and

$$(1-\rho)\sqrt{s} \to \beta, \quad \beta \in (-\infty, \infty). \tag{1.1}$$

This type of scaling was already proposed by Erlang (see [4]) for the $M/M/s/s$ system, and by Pollaczek [21], p. 28, for the $M/D/s$ system. Halfin and Whitt [13] presented a formal limit theorem for the $GI/M/s$ system. Then, some two decades later, the regime got immensely popular due to its application to call centers (see [3,10,14]). The scaling (1.1) combines large capacity with high utilization such that the probability of delay converges to a non-degenerate limit away from both zero and one; cf. (2.28). Limit theorems for other, more general systems were obtained in [9,11,15,19,20,22]. For delay systems like $M/D/s$ and $GI/M/s$ one should impose $\beta \in (0, \infty)$ to guarantee stability.

In [13] it is established that by setting the traffic intensity $\rho = 1 - \beta/\sqrt{s}$, $\beta \in (0, \infty)$, the number of customers in the $M/M/s$ system can be roughly expressed as $s + \sqrt{s}X(t)$ for s sufficiently large and $(X(t))_{t\geq 0}$ the Halfin–Whitt diffusion. It is further shown that properties of the limiting diffusion process for the $GI/M/s$ system can be obtained from $(X(t))_{t\geq 0}$ as well. The boundary between the Brownian motion and the Ornstein–Uhlenbeck process can be thought of as the number of servers, and $(X(t))_{t\geq 0}$ will keep fluctuating between these two regions. The process mimics a single server queue above zero, and an infinite server queue below zero, for which Brownian motion and the Ornstein–Uhlenbeck process are indeed the respective heavy-traffic limits. As β increases, capacity grows and the Halfin–Whitt diffusion will spend more time below zero.

The diffusion process $(X(t))_{t\geq 0}$ can thus be employed to obtain simple approximations for the system's behavior. The steady-state properties of the diffusion are well-studied, but less is known about the transient behavior. Transient results enhance our understanding of how the $GI/M/s$ system behaves over various time and space scales. Results for the mean hitting time were presented in Maglaras and Zeevi [19]. We shall derive explicit results for the transient density of the diffusion, both exact and asymptotic.

We first derive the Laplace transform over time, which leads to a representation of the density as a contour integral, from which a spectral expansion may be obtained by analyzing the complex singularities of the integrand. The spectral expansion can be interpreted as a large-time expansion in which the first term, corresponding to the singularity at zero, gives the steady-state density (which exists if $\beta > 0$). The other singularities of the Laplace transform provide finite-time corrections to the steady-state density. This facilitates us to study how, and in what time (relaxation time), the process converges to its steady state.

The approach to equilibrium is governed by the singularity in the left half-plane with the largest real part. This dominant singularity turns out to be either a branch point or a pole, depending on whether β is smaller or larger than the critical value $\beta_* \approx 1.85722$. This confirms the recent result of Gamarnik and Goldberg [8] who identified β_* using the framework of Karlin and McGregor [16] for birth–death processes, and the result of van Doorn [24] on the spectral gap of the $M/M/s$ queue. We shall also show how the branch point and the pole each give rise to different large-time asymptotics for the density. The main results are presented in Section 2 and the proofs are presented in Section 3.

2. Main results

Halfin and Whitt [13] considered the $M/M/s$ queue with customers arriving according to a Poisson process and requiring service at rate μ. Letting the arrival rate λ^s increase with s so that $\rho^s = \lambda^s/(s\mu)$, they obtained the scaling limit of the sequence of normalized processes $X^s(t) = (Q^s(t) - s)/\sqrt{s}$ with $Q^s(t)$ the number of customers in the system at time t. Let $X^s = (X^s(t))_{t\geq 0}$ and "$\Rightarrow$" denote weak convergence in the space $D[0,\infty)$.

Theorem 1 (*Halfin and Whitt [13]*)**.** *If* $\sqrt{s}(1-\rho^s) \to \beta \in (0,\infty)$*, and* $X^s(0) \Rightarrow c \in \mathbb{R}$*, then* $X^s \to X$*, where the limit* X *is the diffusion process with infinitesimal drift function*

$$C(x) = \begin{cases} -\mu\beta, & x > 0, \\ -\mu x - \mu\beta, & x < 0, \end{cases} \tag{2.1}$$

and infinitesimal variance $B(x) = 2\mu$.

We set $\mu = 1$ without loss of generality. The Halfin–Whitt diffusion $X = (X(t))_{t\geq 0}$ is thus a Markov process on the real line with continuous paths and density $p = p(x,t)$ that satisfies the forward Kolmogorov equation

$$\frac{\partial}{\partial t} p(x,t) = -\frac{\partial}{\partial x}[C(x)p(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[B(x)p(x,t)]. \tag{2.2}$$

There is the initial condition $p(x,0) = \delta(x - x_0)$ (the Dirac function), the interface conditions $p(0+,t) = p(0-,t)$ and $p_x(0+,t) = p_x(0-,t)$, and the boundary conditions $p(\infty,t) = p(-\infty,t) = 0$.

This diffusion process applies directly to the $M/M/s$ system. For the $GI/M/s$ system we would need to first take the diffusion coefficient $B(x) = (1+c^2)$, with $c^2 > 0$, and scale x so as to make $B(x) = 2$, and then scale β by the same factor as x (see [13], Theorem 4).

2.1. Laplace transforms

Define the Laplace transform over time $\hat{p}$ by

$$\hat{p}(x;\theta) = \int_0^\infty e^{-\theta t} p(x,t)\mathrm{d}t, \quad \Re(\theta) > 0. \tag{2.3}$$

Let

$$R_\beta(\theta) = \frac{D'_{-\theta}(-\beta)}{D_{-\theta}(-\beta)} \tag{2.4}$$

with $D_\nu(z)$ the parabolic cylinder function with index ν and argument z, which may be defined, for example, by the integrals

$$D_\nu(z) = \frac{e^{-z^2/4}}{\Gamma(-\nu)} \int_0^\infty e^{-zu} e^{-u^2/2} u^{-\nu-1}\mathrm{d}u, \quad \Re(\nu) < 0, \tag{2.5}$$

$$D_\nu(z) = \frac{e^{z^2/4}}{i\sqrt{2\pi}} \int_{\mathcal{C}} u^\nu e^{u^2/2} e^{-uz}\mathrm{d}u. \tag{2.6}$$

Here, $\Gamma(\cdot)$ is the Gamma function, and the contour $\mathcal{C}$ in the second integral is a vertical Bromwich contour in the half-plane $\Re(u) > 0$. It is well known that $D_\nu(z)$ is an entire function of both index ν and argument z, and various properties of $D_\nu(z)$ are given in [1], Chapter 19, and [12], p. 1092–1095.

Below we give expressions for $\hat{p}$, where we must distinguish the cases $x_0 > 0$ and $x_0 < 0$. These are derived in Section 3.1.

Theorem 2. *Consider* $x_0 > 0$.

(i) *For* $x > 0$,

$$\hat{p}(x;\theta) = \frac{e^{\frac{1}{2}\beta(x_0-x)}}{\sqrt{\beta^2+4\theta}}\left(e^{-|x-x_0|\sqrt{\theta+\beta^2/4}} - e^{-(x+x_0)\sqrt{\theta+\beta^2/4}}\right) + \frac{e^{\frac{1}{2}\beta(x_0-x)}e^{-(x+x_0)\sqrt{\theta+\beta^2/4}}}{\sqrt{\theta+\beta^2/4}-R_\beta(\theta)}. \tag{2.7}$$

(ii) *For* $x < 0$,

$$\hat{p}(x;\theta) = e^{-\frac{1}{4}x^2}e^{-\frac{1}{2}\beta x}\frac{D_{-\theta}(-\beta-x)}{D_{-\theta}(-\beta)}\frac{e^{\frac{1}{2}x_0\beta - x_0\sqrt{\theta+\beta^2/4}}}{\sqrt{\theta+\beta^2/4}-R_\beta(\theta)}. \tag{2.8}$$

Theorem 3. *Consider* $x_0 < 0$.

(i) *For* $x > 0$,

$$\hat{p}(x;\theta) = e^{\frac{1}{4}x_0^2}e^{\frac{1}{2}\beta x_0}\frac{D_{-\theta}(-\beta-x_0)}{D_{-\theta}(-\beta)}\frac{e^{-\frac{1}{2}x\beta - x\sqrt{\theta+\beta^2/4}}}{\sqrt{\theta+\beta^2/4}-R_\beta(\theta)}. \tag{2.9}$$

(ii) *For* $x < 0$,

$$\begin{aligned}\hat{p}(x;\theta) = {} & A(\theta)e^{\frac{1}{4}(x_0^2-x^2)}e^{\frac{1}{2}\beta(x_0-x)}D_{-\theta}(-\beta-x) \\ & + \mathbf{1}\{x_0 < x < 0\}e^{\frac{1}{4}(x_0^2-x^2)}e^{\frac{1}{2}\beta(x_0-x)}\frac{\Gamma(\theta)}{\sqrt{2\pi}} \\ & \times [D_{-\theta}(-\beta-x_0)D_{-\theta}(\beta+x) - D_{-\theta}(\beta+x_0)D_{-\theta}(-\beta-x)],\end{aligned} \tag{2.10}$$

where

$$A(\theta) = \frac{\Gamma(\theta)}{\sqrt{2\pi}} \times \left[D_{-\theta}(\beta+x_0) - \frac{D_{-\theta}(\beta)D_{-\theta}(-\beta-x_0)}{D_{-\theta}(-\beta)}\frac{\sqrt{\theta+\beta^2/4}+R_{-\beta}(\theta)}{\sqrt{\theta+\beta^2/4}-R_\beta(\theta)}\right]. \tag{2.11}$$

In (2.10) $\mathbf{1}\{\cdot\}$ is the indicator function. Note that if $x_0 \to 0^+$ the first term in the right-hand side in (2.7) disappears, and $D_{-\theta}(-\beta-x_0)/D_{-\theta}(-\beta) \to 1$ as $x_0 \to 0^-$ in (2.9). Thus (2.7) and (2.9) give the same result if $x_0 \to 0$. If $x < 0$ we have $\mathbf{1}\{x_0 < x < 0\} \to 0$ as $x_0 \to 0^-$, so the second term in the right-hand side of (2.10) disappears. Also as $x_0 \to 0^-$, (2.11) shows that

$$\begin{aligned} A(\theta) &\to \frac{\Gamma(\theta)}{\sqrt{2\pi}} D_{-\theta}(\beta)\left[1 - \frac{\sqrt{\theta+\beta^2/4} + R_{-\beta}(\theta)}{\sqrt{\theta+\beta^2/4} - R_{\beta}(\theta)}\right] \\ &= \frac{\Gamma(\theta)}{\sqrt{2\pi}} \frac{D_{-\theta}(\beta)}{\sqrt{\theta+\beta^2/4} - R_{\beta}(\theta)}[-R_{-\beta}(\theta) - R_{-\beta}(\theta)] \\ &= \frac{1}{D_{-\theta}(-\beta)} \frac{1}{\sqrt{\theta+\beta^2/4} - R_{\beta}(\theta)}, \end{aligned} \tag{2.12}$$

where we used (2.4) and (2.14) below. Thus (2.8) and (2.10) also agree if $x_0 \to 0$.

We can rewrite (2.10) in the following alternate form:

$$\begin{aligned} \hat{p}(x;\theta) = {} & \frac{\Gamma(\theta)}{\sqrt{2\pi}} \mathrm{e}^{\frac{1}{4}(x_0^2-x^2)} \mathrm{e}^{\frac{1}{2}\beta(x_0-x)} \Bigg[D_{-\theta}(\beta+x_>) D_{-\theta}(-\beta-x_<) \\ & - \frac{D_{-\theta}(\beta)}{D_{-\theta}(-\beta)} D_{-\theta}(-\beta-x) D_{-\theta}(-\beta-x_0) \Bigg] \\ & + \mathrm{e}^{\frac{1}{4}(x_0^2-x^2)} \mathrm{e}^{\frac{1}{2}\beta(x_0-x)} \frac{D_{-\theta}(-\beta-x) D_{-\theta}(-\beta-x_0)}{D_{-\theta}^2(-\beta)[\sqrt{\theta+\beta^2/4} - R_{\beta}(\theta)]}, \end{aligned} \tag{2.13}$$

where $x_> = \max\{x, x_0\}$ and $x_< = \min\{x, x_0\}$. The equivalence of (2.10) and (2.13) follows from the Wronskian identity (this may be shown directly using (2.5) and (2.6))

$$-\frac{\sqrt{2\pi}}{\Gamma(\theta)} = D_{-\theta}(z) D'_{-\theta}(-z) + D_{-\theta}(-z) D'_{-\theta}(z), \tag{2.14}$$

which is independent of z.

While it does not seem possible to invert the Laplace transforms in Theorems 2 and 3 to get the density $p(x,t)$ explicitly, parts of $\hat{p}$ can be inverted. For $x_0 > 0$ we note that the part of $\hat{p}$ that is in the first line in the right-hand side of (2.7) inverts to

$$\frac{1}{2\sqrt{\pi t}} \mathrm{e}^{-\frac{1}{4}\beta^2 t} \mathrm{e}^{\frac{1}{2}\beta(x_0-x)} (\mathrm{e}^{-\frac{1}{4}(x-x_0)^2/t} - \mathrm{e}^{-\frac{1}{4}(x+x_0)^2/t}), \tag{2.15}$$

which is similar to the density for a Brownian motion with absorption at $x = 0$. The inversion of the second part of $\hat{p}$ in (2.7) seems less straightforward. Of course, the full process is not absorbed at $x = 0$, but rather crosses this interface with a certain reflection law.

For $x_0 < 0$ we can invert the term in the right-hand side of (2.13) that is proportional to $D_{-\theta}(\beta + x_>) D_{-\theta}(-\beta - x_<)$. Since $\Gamma(\theta)$ has simple poles at $\theta = -n$, $n = 0, 1, 2, \ldots$, with residues $(-1)^n/n!$, and $D_{-\theta}(\cdot)$ is an entire function of θ, the first term inverts to

$$\mathrm{e}^{\frac{1}{4}(x_0^2-x^2)} \mathrm{e}^{\frac{1}{2}\beta(x_0-x)} \sum_{n=0}^{\infty} D_n(-\beta-x_0) D_n(\beta+x) \frac{(-1)^n \mathrm{e}^{-nt}}{n!\sqrt{2\pi}}. \tag{2.16}$$

This corresponds to the transient solution of the standard free space Ornstein–Uhlenbeck process, starting at x_0 at time $t = 0$ (see Section 2.5). The remaining two terms in (2.13) represent the effects of the interface at $x = 0$, where the form of the drift changes. As $t \to \infty$ (2.16) approaches $\exp(-(x+\beta)^2/2)/\sqrt{2\pi}$, as only the term $n = 0$ remains, and $D_0(z) = \mathrm{e}^{-z^2/4}$.

2.2. Relaxation time

In queueing theory, the *relaxation time* is a notion that measures the time it takes for the system to approach its steady-state behavior. There are various ways to define relaxation time, but we use the definition

$$\tau = \inf\{T : p(x,t) - p(x,\infty) = O(\mathrm{e}^{-t/T})\}, \tag{2.17}$$

in the spirit of [2,5,18]. The Laplace transform $\hat{p}$ is analytic in the entire θ-plane, except for singularities in the range $\Re(\theta) \le 0$. Hence, the asymptotic behavior of $p(x,t)$ (for large t) is determined by the singularity $\hat{\theta}$ closest to the imaginary axis. In fact, from (2.17) it follows that

$$\tau^{-1} = -\Re(\hat{\theta}). \tag{2.18}$$

The dominant singularity $\hat{\theta}$ will either be the branch point $\theta_B = -\frac{1}{4}\beta^2$ or the largest negative solution θ_P to

$$\varphi_\beta(\theta) := \sqrt{\theta + \beta^2/4} - R_\beta(\theta) = 0. \tag{2.19}$$

We shall study later in more detail all solutions to (2.19) (see Sections 2.5 and 3.3). In particular we will enumerate the number of solutions in the range $\Re(\theta) \in (-\beta^2/4, 0)$ for any $\beta > 0$. We have the following result.

Theorem 4. *Let β_* represent the smallest positive real solution to*

$$D'_{\beta^2/4}(-\beta) = 0. \tag{2.20}$$

The dominant singularity $\hat{\theta}$ of the Laplace transform $\hat{p}(x;\theta)$ is then given by

$$\hat{\theta} = \begin{cases} \theta_B = -\dfrac{1}{4}\beta^2, & 0 < \beta \le \beta_*, \\ \theta_P, & \beta \ge \beta_*. \end{cases} \tag{2.21}$$

The numerical value of β_* is 1.85722 . . ..

This completely determines the relaxation time as defined in (2.17). More detailed information on the distance to steady state can be obtained from investigating $\hat{p}$ in the vicinity of the dominant singularity; see Theorems 5 and 6. When $\beta \le 0$ the process is transient and the large-time behavior is still determined by θ_B.

Using asymptotic results in [1], Chapter 19, we have the estimate, for $\beta \to \infty$,

$$D'_{\beta^2/4}(-\beta) \sim \left(\frac{3}{2}\right)^{1/6} \frac{1}{\sqrt{\pi}} \Gamma(2/3) \left(\frac{\beta}{2}\right)^{\beta^2/4} \mathrm{e}^{-\beta^2/8} \beta^{2/3} \sin(\pi\beta^2/4 + \pi/6). \tag{2.22}$$

This expression shows that $D'_{\beta^2/4}(-\beta)$ has an infinite number of sign changes as β increases toward infinity, and thus (2.20) has infinitely many roots, which we denote by $\beta_{*,N}$, $N = 1, 2, 3, \ldots$ with $\beta_{*,1} = \beta_*$.

Using the recurrence relation (see [12], p. 1094)

$$D'_p(z) = -\frac{1}{2} z D_p(z) + p D_{p-1}(z) \tag{2.23}$$

with $p = \beta^2/4$ and $z = -\beta$ it follows that (2.20) is equivalent to

$$\frac{-2D_{\beta^2/4}(-\beta)}{\beta D_{\beta^2/4-1}(-\beta)} = 1. \tag{2.24}$$

Using the integral representation in [12], p. 1094, this equation is equivalent to

$$\frac{2}{\beta}\frac{\int_0^\infty x^{1-\beta^2/4}\mathrm{e}^{-(\beta-x)^2/2}\mathrm{d}x}{\int_0^\infty x^{-\beta^2/4}\mathrm{e}^{-(\beta-x)^2/2}\mathrm{d}x} = 1, \quad \beta^2 < 4, \tag{2.25}$$

which is the expression derived by Gamarnik and Goldberg [8]. To see the equivalence we rewrite the integral in the numerator of (2.25) and integrate by parts, which yields

$$\begin{aligned}\int_0^\infty x^{-\beta^2/4}(x-\beta+\beta)\mathrm{e}^{-(\beta-x)^2/2}\mathrm{d}x &= \beta\int_0^\infty x^{-\beta^2/4}\mathrm{e}^{-(\beta-x)^2/2}\mathrm{d}x \\ &\quad - \int_0^\infty x^{-\beta^2/4}\mathrm{d}[\mathrm{e}^{-(\beta-x)^2/2}] \\ &= \int_0^\infty (\beta - \beta^2 x^{-1}/4)x^{-\beta^2/4}\mathrm{e}^{-(\beta-x)^2/2}\mathrm{d}x \\ &= \beta\Gamma(1-\beta^2/4)\mathrm{e}^{-\beta^2/4}D_{\beta^2/4-1}(-\beta) \\ &\quad - \frac{\beta^2}{4}\Gamma(-\beta^2/4)\mathrm{e}^{-\beta^2/4}D_{\beta^2/4}(-\beta).\end{aligned} \tag{2.26}$$

Here we used (2.5) to express the last integrals in terms of the parabolic cylinder functions.

By (2.25), (2.26) should equal

$$\frac{\beta}{2}\int_0^\infty x^{-\beta^2/4}\mathrm{e}^{-(\beta-x)^2/2}\mathrm{d}x = \frac{\beta}{2}\Gamma(1-\beta^2/4)\mathrm{e}^{-\beta^2/4}D_{\beta^2/4-1}(-\beta). \tag{2.27}$$

Using the fact that $w\Gamma(w) = \Gamma(w+1)$ with $w = -\beta^2/4$ and then canceling the common factors $\Gamma(1-\beta^2/4)\mathrm{e}^{-\beta^2/4}$, (2.26) equalling (2.27) implies that $\frac{1}{2}\beta D_{\beta^2/4-1}(-\beta) + D_{\beta^2/4}(-\beta) = 0$, which is (2.24).

2.3. Limiting density

Let $\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$ and $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \mathrm{e}^{-u^2/2}\mathrm{d}u$, be the density and the distribution function of a standard normal random variable. Then we define

$$C_0(\beta) = \left[1 + \frac{\beta\Phi(\beta)}{\phi(\beta)}\right]^{-1}, \tag{2.28}$$

which is the non-degenerate limit of the delay probability. Defining $p(x,\infty) = \lim_{t\to\infty} p(x,t)$, the limiting distribution of the diffusion process is given by (see [13])

$$p(x,\infty) = \begin{cases} C_0(\beta)\beta\mathrm{e}^{-\beta x}, & x > 0, \\ C_0(\beta)\beta\mathrm{e}^{-\frac{1}{2}x^2}\mathrm{e}^{-\beta x}, & x < 0. \end{cases} \tag{2.29}$$

This also follows from our expression for the Laplace transform $\hat{p}$. Since $D_0(\beta) = \mathrm{e}^{-\beta^2/4}$, we have $R_\beta(0) = \frac{1}{2}\beta$, and the function $\hat{p}$ has a pole at $\theta = 0$ if $\beta > 0$ (the stable case). Calculating the residue yields

$$p(x,\infty) = \frac{1}{1-\beta R_\beta'(0)} \begin{cases} \beta \mathrm{e}^{-\beta x}, & x > 0, \\ \beta \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{e}^{-\beta x}, & x < 0, \end{cases} \tag{2.30}$$

and some further algebra shows that indeed $R_\beta'(0) = -\Phi(\beta)/\phi(\beta)$.

2.4. Large-time asymptotics

We give the approach to equilibrium, distinguishing the cases x, x_0 positive or negative. We note that $p(x,\infty) = 0$ if $\beta \le 0$. Here $a(t) \sim b(t)$ means that $\lim_{t\to\infty}[a(t)/b(t)] = 1$. The proof of the following theorem is sketched in Section 3.2.

Theorem 5. *Consider $x_0 > 0$.*

(i) *For $x > 0$, $\beta < \beta_*$, and $\beta \ne 0$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{2\sqrt{\pi} t^{3/2}} \mathrm{e}^{-\frac{1}{4}\beta^2 t} \mathrm{e}^{\frac{1}{2}\beta(x_0 - x)} \times \left[x x_0 - \frac{x + x_0}{R_\beta(-\beta^2/4)} + \frac{1}{R_\beta^2(-\beta^2/4)} \right]. \tag{2.31}$$

(ii) *For $x < 0$, $\beta < \beta_*$, and $\beta \ne 0$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{2\sqrt{\pi} t^{3/2}} \mathrm{e}^{-\frac{1}{4}\beta^2 t} \mathrm{e}^{-\frac{1}{4}x^2} \mathrm{e}^{\frac{1}{2}\beta(x_0 - x)} \times \frac{[1 - x_0 R_\beta(-\beta^2/4)]}{R_\beta^2(-\beta^2/4)} \frac{D_{\beta^2/4}(-\beta - x)}{D_{\beta^2/4}(-\beta)}. \tag{2.32}$$

(iii) *For $x > 0$ and $\beta > \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \mathrm{e}^{\theta_P t} \frac{\mathrm{e}^{\frac{1}{2}\beta(x_0 - x)} \mathrm{e}^{-(x + x_0)\sqrt{\theta_P + \beta^2/4}}}{\varphi_\beta'(\theta_P)}. \tag{2.33}$$

(iv) *For $x < 0$ and $\beta > \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \mathrm{e}^{\theta_P t} \mathrm{e}^{-\frac{1}{4}x^2} \mathrm{e}^{-\frac{1}{2}\beta x} \frac{D_{-\theta_P}(-\beta - x)}{D_{-\theta_P}(-\beta)} \frac{\mathrm{e}^{\frac{1}{2}x_0\beta - x_0\sqrt{\theta_P + \beta^2/4}}}{\varphi_\beta'(\theta_P)}. \tag{2.34}$$

(v) *For $x > 0$ and $\beta = \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{\sqrt{\pi t}} \mathrm{e}^{-\frac{1}{4}\beta_*^2 t} \mathrm{e}^{\frac{1}{2}\beta_*(x_0 - x)}. \tag{2.35}$$

(vi) *For $x < 0$ and $\beta = \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{\sqrt{\pi t}} \mathrm{e}^{-\frac{1}{4}\beta_*^2 t} \mathrm{e}^{-\frac{1}{4}x^2} \mathrm{e}^{\frac{1}{2}\beta_*(x_0 - x)} \frac{D_{\beta_*^2/4}(-\beta_* - x)}{D_{\beta_*^2/4}(-\beta_*)}. \tag{2.36}$$

Theorem 6. *Consider $x_0 < 0$.*

(i) *For $x > 0$, $\beta < \beta_*$, and $\beta \neq 0$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{2\sqrt{\pi} t^{3/2}} \mathrm{e}^{-\frac{1}{4}\beta^2 t} \mathrm{e}^{\frac{1}{2}\beta(x_0 - x)} \mathrm{e}^{\frac{1}{4}x_0^2} \times \frac{D_{\beta^2/4}(-\beta - x_0)}{D_{\beta^2/4}(-\beta)} \left[\frac{x}{R_\beta(-\beta^2/4)} - \frac{1}{R_\beta^2(-\beta^2/4)} \right]. \tag{2.37}$$

(ii) *For $x < 0$, $\beta < \beta_*$, and $\beta \neq 0$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{2\sqrt{\pi} t^{3/2}} \mathrm{e}^{-\frac{1}{4}\beta^2 t} \mathrm{e}^{\frac{1}{2}\beta(x_0 - x)} \mathrm{e}^{\frac{1}{4}(x_0^2 - x^2)} \times \frac{1}{R_\beta^2(-\beta^2/4)} \frac{D_{\beta^2/4}(-\beta - x) D_{\beta^2/4}(-\beta - x_0)}{D^2_{\beta^2/4}(-\beta)}. \tag{2.38}$$

(iii) *For $x > 0$ and $\beta > \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \mathrm{e}^{\theta_P t} \mathrm{e}^{\frac{1}{4}x_0^2} \mathrm{e}^{\frac{1}{2}\beta x_0} \frac{D_{-\theta_P}(-\beta - x_0)}{D_{-\theta_P}(-\beta)} \frac{\mathrm{e}^{-\frac{1}{2}x\beta - x\sqrt{\theta_P + \beta^2/4}}}{\varphi'_\beta(\theta_P)}. \tag{2.39}$$

(iv) *For $x < 0$ and $\beta > \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \mathrm{e}^{\theta_P t} \mathrm{e}^{\frac{1}{4}(x_0^2 - x^2)} \mathrm{e}^{\frac{1}{2}\beta(x_0 - x)} \frac{D_{-\theta_P}(-\beta - x_0) D_{-\theta_P}(-\beta - x)}{D^2_{-\theta_P}(-\beta)\varphi'_\beta(\theta_P)}. \tag{2.40}$$

(v) *For $x > 0$ and $\beta = \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{\sqrt{\pi t}} \mathrm{e}^{-\frac{1}{4}\beta_*^2 t} \mathrm{e}^{\frac{1}{4}x_0^2} \mathrm{e}^{\frac{1}{2}\beta_*(x_0 - x)} \frac{D_{\beta_*^2/4}(-\beta_* - x_0)}{D_{\beta_*^2/4}(-\beta_*)}. \tag{2.41}$$

(vi) *For $x < 0$ and $\beta = \beta_*$,*

$$p(x,t) - p(x,\infty) \sim \frac{1}{\sqrt{\pi t}} \mathrm{e}^{-\frac{1}{4}\beta_*^2 t} \mathrm{e}^{\frac{1}{4}(x_0^2 - x^2)} \mathrm{e}^{\frac{1}{2}\beta_*(x_0 - x)} \times \frac{D_{\beta_*^2/4}(-\beta_* - x) D_{\beta_*^2/4}(-\beta_* - x_0)}{D^2_{\beta_*^2/4}(-\beta_*)}. \tag{2.42}$$

Here $\varphi'_\beta(\theta_P) = (4\theta_P + \beta^2)^{-1/2} - R'_\beta(\theta_P)$, as in (2.19). When $\beta = 0$ the result is independent of x_0 and we have

$$p(x,t) \sim \frac{1}{\sqrt{\pi t}} \begin{cases} 1, & x > 0, \\ \mathrm{e}^{-x^2/2}, & x < 0. \end{cases} \tag{2.43}$$

2.5. Spectral properties

We first note that in the limit $\beta \to \infty$ with $x, x_0 \to -\infty$, and with $x + \beta \equiv y$ and $x_0 + \beta \equiv y_0$ fixed, the expression(s) in Theorem 3 approach the Laplace transform of the standard Ornstein–Uhlenbeck process. We use the alternate form (2.13) for $\hat{p}(x;\theta)$, which applies for

$x, x_0 < 0$. For $\beta \to +\infty$ ($-\beta \to -\infty$) we use the asymptotic results ([12], p. 1094)

$$D_{-\theta}(-\beta) \sim \frac{\sqrt{2\pi}}{\Gamma(\theta)} \beta^{\theta-1} \mathrm{e}^{\beta^2/4}, \tag{2.44}$$

$$D_{-\theta}(\beta) \sim \beta^{-\theta} \mathrm{e}^{-\beta^2/4}, \tag{2.45}$$

with which $D_{-\theta}^{-2}(-\beta)$ and $D_{-\theta}(\beta)/D_{-\theta}(-\beta)$ decay roughly like $\mathrm{e}^{-\beta^2/2}$ as $\beta \to \infty$. Thus letting $\beta \to \infty$ in (2.13) (with y, y_0 fixed) and noting that $\exp[\frac{1}{4}(x_0^2 - x^2) + \frac{1}{2}\beta(x_0 - x)] = \exp[\frac{1}{4}(y_0^2 - y^2)]$ we find that

$$\hat{p}(x; \theta) \to \frac{\Gamma(\theta)}{\sqrt{2\pi}} \mathrm{e}^{\frac{1}{4}(y_0^2 - y^2)} D_{-\theta}(y_>) D_{-\theta}(-y_<), \tag{2.46}$$

where $y_> = \max\{y, y_0\}$ and $y_< = \min\{y, y_0\}$. The right side of (2.46) is precisely the Laplace transform of the density of the standard Ornstein–Uhlenbeck process, which inverts to (see [17])

$$q(y, t) = \frac{\mathrm{e}^{\frac{1}{4}(y_0^2 - y^2)}}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{D_n(y_0) D_n(y)}{n!} \mathrm{e}^{-nt}, \quad y \in \mathbb{R}. \tag{2.47}$$

Here $D_n(y) = \mathrm{e}^{-y^2/4} 2^{-n/2} H_n(y/\sqrt{2})$ where $H_n(\cdot)$ is the nth Hermite polynomial. This series may be explicitly summed to give

$$q(y, t) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{1}{1 - \mathrm{e}^{-2t}}} \exp\left[-\frac{(y - y_0 \mathrm{e}^{-t})^2}{2(1 - \mathrm{e}^{-2t})}\right], \quad y \in \mathbb{R}. \tag{2.48}$$

The singularities in (2.46) are those of the factor $\Gamma(\theta)$, which are simple poles at $\theta = -N$; $N = 0, 1, 2, \ldots$, with residues $(-1)^N/N!$. We now examine how the spectrum of the Halfin–Whitt diffusion approaches that of the Ornstein–Uhlenbeck process, for $\beta \to \infty$. We recall that the former spectrum consists of the continuous spectrum where $\Re(\theta) \in (-\infty, -\beta^2/4)$ and a discrete spectrum which consists of $\theta = 0$ (for any $\beta > 0$) and the solutions to $\varphi_\beta(\theta) = 0$ in (2.19). Finding the eigenvalues is equivalent to solving

$$D_p(-\beta)\sqrt{\beta^2/4 - p} = D_p'(-\beta). \tag{2.49}$$

Here, we set $p = -\theta$ and multiplied (2.19) by $D_p(-\beta)$. Note that the roots of $D_p(-\beta) = 0$ would correspond to zeros, and not poles, of $\hat{p}(x; \theta)$, so that (2.49) and (2.19) are equivalent.

In Section 3.3 we show that all eigenvalues must be real and we give a topological argument that establishes the following result.

Theorem 7. *Denoting the solutions to (2.20) by $0 < \beta_* < \beta_{*,2} < \beta_{*,3} < \cdots$, for $\beta_{*,N} < \beta < \beta_{*,N+1}$ (2.49) has exactly N solutions (at least for $1 \le N \le 100$) in the interval $p \in (0, \beta^2/4)$. We call these solutions $p_N(\beta)$ and they satisfy the asymptotic properties*

$$p_N(\beta_{*,N}) = \frac{1}{4}\beta_{*,N}^2 \sim N - \frac{1}{6}, \quad N \to \infty \tag{2.50}$$

and (for N fixed)

$$p_N(\beta) - N \sim -\frac{1}{(N-1)!} \frac{\mathrm{e}^{-\beta^2/2}}{\sqrt{2\pi}} \beta^{2N-3}, \quad \beta \to \infty. \tag{2.51}$$

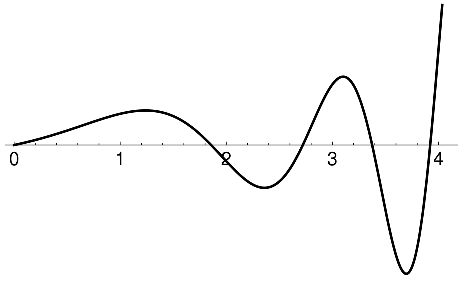

Fig. 1. The function $D'_{\beta^2/4}(-\beta)$ for $\beta \in [0, 4]$.

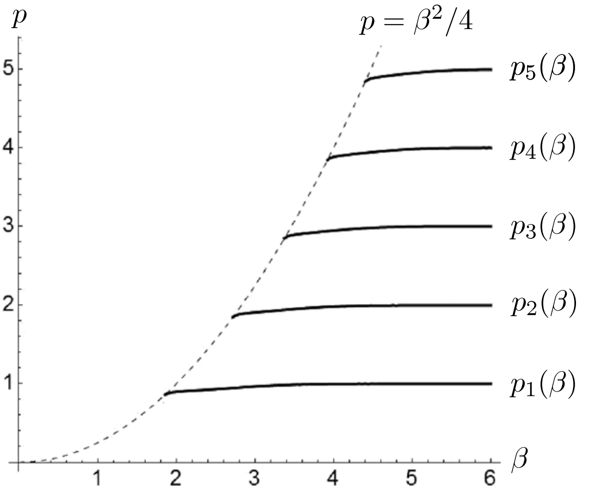

Fig. 2. Solutions to $\sqrt{\beta^2/4 - p}\, D_p(-\beta) = D'_p(-\beta)$.

The critical values $\beta_{*,N}$ are asymptotically given by

$$\beta_{*,N} \sim 2\sqrt{N - 1/6}, \quad N \to \infty. \tag{2.52}$$

We establish this theorem in Section 3.3. The argument relies on the positivity of a sequence Ω_N in (3.42). We established this numerically for $N \leq 100$ and asymptotically for $N \to \infty$, but not yet for all N, thus the qualification above. The expressions in (2.52), and hence (2.50), follow immediately from (2.22), by setting $\sin(\pi\beta^2/4 + \pi/6) = 0$. While these apply for $N \to \infty$ they work well also for moderate N. For example, the exact value of β_* ($N = 1$) is $1.85722\ldots$, while (2.52) gives the approximation $\sqrt{10/3} = 1.82574\ldots$. The first few solutions of (2.20) are given by $\beta_{*,2} \approx 2.72133$, $\beta_{*,3} \approx 3.37465$, and $\beta_{*,4} \approx 3.92155$. In Fig. 1 we plot $D'_{\beta^2/4}(-\beta)$ versus β for $\beta \in [0, 4]$, which illustrates the first four roots. Fig. 1 also illustrates the growth and rapid oscillations of $D'_{\beta^2/4}(-\beta)$, as predicted by the asymptotic formula (2.22).

To see the evolution of the discrete part of the spectrum with increasing β, we plot in Fig. 2 the solutions $p = p(\beta)$ to the equation in (2.49). This clearly shows the branches $p_N(\beta)$ being "born" along the parabola $p = \beta^2/4$ at the points $(\beta_{*,N}, \frac{1}{4}\beta_{*,N}^2)$. Once a branch is born it remains in the range $0 < p < \beta^2/4$ and approaches N as $\beta \to \infty$. We see that each $p_N(\beta)$ increases with β, but only slightly. Indeed Theorem 7 shows that the total movement of the N-th branch satisfies

$$p_N(\infty) - p_N(\beta_{*,N}) \to \frac{1}{6} = 0.1666\ldots, \quad N \to \infty. \tag{2.53}$$

When $N = 1$ (resp., $N = 2$) the exact values of the left side of (2.53) are 0.1377 (resp., 0.1486). Fig. 2 thus illustrates how the discrete Ornstein–Uhlenbeck spectrum develops from the present model as $\beta \uparrow \infty$. The expression in (2.51) quantifies the "flatness" of the curves in Fig. 2 for large values of β, and gives the gap(s) between the discrete spectra of the Ornstein–Uhlenbeck process and the Halfin–Whitt diffusion. Establishing (2.51) requires more work, and this is done in Section 3.3, along with the first part of Theorem 7.

3. Proofs

3.1. Derivation of the Laplace transforms

We now present the proofs of Theorems 2 and 3. If p satisfies (2.2) its Laplace transform satisfies

$$\theta \hat{p}(x;\theta) - \delta(x - x_0) = -\frac{\mathrm{d}}{\mathrm{d}x}[A(x)\hat{p}(x;\theta)] + \frac{\mathrm{d}^2 \hat{p}(x;\theta)}{\mathrm{d}x^2}, \tag{3.1}$$

where

$$-\frac{\mathrm{d}}{\mathrm{d}x}[A(x)\hat{p}(x;\theta)] = \begin{cases} \beta \dfrac{\mathrm{d}}{\mathrm{d}x}\hat{p}(x;\theta), & x > 0, \\ (x+\beta)\dfrac{\mathrm{d}}{\mathrm{d}x}\hat{p}(x;\theta) + \hat{p}(x;\theta), & x < 0. \end{cases} \tag{3.2}$$

First we take $x_0 > 0$ so that $\delta(x - x_0) = 0$ in the range $x < 0$. For $x < 0$ we write $\hat{p} = \mathrm{e}^{-x^2/4}\mathrm{e}^{-\beta x/2} v$ and then (3.1) reduces to the differential equation ([6], p. 116)

$$v'' + \left[\frac{1}{2} - \theta - \frac{1}{4}(x+\beta)^2\right] v = 0, \tag{3.3}$$

whose solution is $v(x;\theta) = \alpha_1(\theta) D_{-\theta}(-\beta - x)$, where α_1 is still to be determined. Note that $D_{-\theta}(-z)$ has Gaussian decay as $z \to -\infty$, while $D_{-\theta}(z)$, which is a second solution to the parabolic cylinder equation (3.3), grows like $O(\mathrm{e}^{z^2/4})$ as $z \to -\infty$.

For $x > 0$ the function $\hat{p} = \mathrm{e}^{wx}$ satisfies the homogeneous version of (3.1) if

$$w^2 + \beta w - \theta = 0, \tag{3.4}$$

with solutions $w = \frac{1}{2}[-\beta - \sqrt{\beta^2 + 4\theta}]$ and $w_* = \frac{1}{2}[-\beta + \sqrt{\beta^2 + 4\theta}]$. It thus follows that

$$\hat{p}(x;\theta) = \begin{cases} \alpha_2(\theta)\mathrm{e}^{wx} + \alpha_3(\theta)\mathrm{e}^{w_* x}, & 0 < x < x_0 \\ \alpha_4(\theta)\mathrm{e}^{wx}, & x > x_0, \end{cases} \tag{3.5}$$

where α_2, α_3 and α_4 still need to be determined. Continuity at $x = x_0$ yields $\hat{p}(x_0^+;\theta) = \hat{p}(x_0^-;\theta)$ and the derivative has a jump at x_0, with

$$\hat{p}_x(x_0^+;\theta) - \hat{p}_x(x_0^-;\theta) = -\int_{x_0^-}^{x_0^+} \delta(x - x_0)\mathrm{d}x = -1, \tag{3.6}$$

which translates into

$$\alpha_2 e^{wx_0} + \alpha_3 e^{w_* x_0} = \alpha_4 e^{wx_0}, \tag{3.7}$$

$$w\alpha_4 e^{wx_0} - w\alpha_2 e^{wx_0} - w_*\alpha_3 e^{w_* x_0} = -1. \tag{3.8}$$

The continuity of p and p_x at $x = 0$ implies the continuity of $\hat{p}$ and $\hat{p}_x = \partial \hat{p}/\partial x$, and this yields the additional relations

$$\alpha_2 + \alpha_3 = \alpha_1 D_{-\theta}(-\beta), \tag{3.9}$$

$$w\alpha_2 + w_*\alpha_3 = -\alpha_1 \left[D'_{-\theta}(-\beta) + \frac{1}{2}\beta D_{-\theta}(-\beta) \right]. \tag{3.10}$$

(3.7)–(3.10) give four equations for the four unknowns α_1, α_2, α_3 and α_4. Some further algebra and the definition $R_\beta(\theta) = D'_{-\theta}(-\beta)/D_{-\theta}(-\beta)$ yields

$$\alpha_1(\theta) = -\frac{1}{D_{-\theta}(-\beta)} \frac{\alpha_3(\theta)\sqrt{\beta^2 + 4\theta}}{R_\beta(\theta) + w + \beta/2}, \tag{3.11}$$

$$\alpha_2(\theta) = -\alpha_3(\theta) - \frac{\alpha_3(\theta)\sqrt{\beta^2 + 4\theta}}{R_\beta(\theta) + w + \beta/2}, \tag{3.12}$$

$$\alpha_3(\theta) = \frac{1}{\sqrt{\beta^2 + 4\theta}} e^{-x_0 w_*}, \tag{3.13}$$

$$\alpha_4(\theta) = \alpha_2(\theta) + \alpha_3(\theta) e^{x_0(w_* - w)}. \tag{3.14}$$

We thus obtain Theorem 2. Using the absolute value $|x - x_0|$ allows us to write the solution as a single formula that applies for all $x > 0$ (cf. (2.7)).

To establish Theorem 3 we note that now $\delta(x - x_0) = 0$ in the range $x > 0$. Thus we write

$$\hat{p}(x; \theta) = \gamma_4(\theta) e^{wx}, \quad x > 0, \tag{3.15}$$

and we need $\hat{p}$ to decay for $x \to -\infty$ so we write

$$\hat{p}(x; \theta) = \gamma_1(\theta) e^{-\frac{1}{4}x^2} e^{-\frac{1}{2}\beta x} D_{-\theta}(-\beta - x), \quad x < x_0 < 0. \tag{3.16}$$

But in the range $x_0 < x < 0$ the solution will involve both of the parabolic cylinder functions $D_{-\theta}(-\beta - x)$ and $D_\theta(\beta + x)$, hence

$$\hat{p}(x; \theta) = e^{-\frac{1}{4}x^2} e^{-\frac{1}{2}\beta x} [\gamma_2(\theta) D_{-\theta}(-\beta - x) + \gamma_3(\theta) D_{-\theta}(\beta + x)]. \tag{3.17}$$

The functions $\gamma_j(\theta)$ are determined by continuity of $\hat{p}$ and $\frac{d}{dx}\hat{p}$ at $x = 0$, which leads to

$$\gamma_4 = \gamma_2 D_{-\theta}(-\beta) + \gamma_3 D_{-\theta}(\beta), \tag{3.18}$$

$$w\gamma_3 = -\frac{1}{2}\beta\gamma_4 - \gamma_2 D'_{-\theta}(-\beta) + \gamma_3 D'_{-\theta}(\beta), \tag{3.19}$$

continuity of $\hat{p}$ at $x = x_0$,

$$\gamma_1 D_{-\theta}(-\beta - x_0) = \gamma_2 D_{-\theta}(-\beta - x_0) + \gamma_3 D_{-\theta}(\beta + x_0), \tag{3.20}$$

and the jump condition of $\frac{\mathrm{d}}{\mathrm{d}x}\hat{p}$ at $x = x_0$

$$-1 = \mathrm{e}^{-\frac{1}{4}x_0^2}\mathrm{e}^{-\frac{1}{2}\beta x_0}[-\gamma_2 D'_{-\theta}(-\beta - x_0) + \gamma_3 D'_{-\theta}(\beta + x_0) + \gamma_1 D'_{-\theta}(-\beta - x_0)]. \quad (3.21)$$

Eqs. (3.18)–(3.21) give a 4×4 linear system whose solution leads to Theorem 3. The Wronskian identity (2.14) allows us to simplify some of the final expressions. In Theorem 3, $A(\theta)$ is the same as $\gamma_1(\theta)\mathrm{e}^{-x_0^2/4}\mathrm{e}^{-\beta x_0/2}$.

3.2. Sketch of the derivation of the asymptotic results

We now briefly derive the asymptotic results that appear in Theorems 5 and 6. We refer the reader to [7] for a discussion of the asymptotic evaluation of an inverse Laplace transform for $t \to \infty$. This involves locating the singularity with the largest real part (cf. Theorem 4) and expanding the integrand near this dominant singularity. Below we sketch some of the details, paying particular attention to the case where the pole at θ_P is close to the branch point at $-\beta^2/4$. Theorems 5 and 6 include the critical case $\beta = \beta_*$, which corresponds to there being an algebraic factor $t^{-1/2}$ in the relaxation asymptotics (rather than $t^{-3/2}$ ($\beta < \beta_*$) or t^0 ($\beta > \beta_*$)). The discussion below also shows how to treat cases where $\beta \approx \beta_*$, which we did not give in Theorems 5 and 6.

Consider a contour integral

$$I(t) = \frac{1}{2\pi\mathrm{i}}\int_{\mathrm{Br}} \frac{g(z)}{\sqrt{z} + f(z)}\mathrm{e}^{zt}\mathrm{d}z. \quad (3.22)$$

Here Br is a vertical Bromwich contour in the z-plane, with the integrand analytic to the right of Br. First we assume that f and g are analytic functions of z in the half-plane $\Re(z) < -\varepsilon_0$ for some $\varepsilon_0 > 0$ with $g(0) \neq 0$ and $f(0) \neq 0$. Then the asymptotics as $t \to \infty$ are governed by the branch point at $z = 0$, if $\sqrt{z} + f(z) = 0$ has no solutions in the range $\Re(z) > 0$. Under these assumptions we can obtain the asymptotics of (3.22) simply by expanding the analytic functions f and g about $z = 0$:

$$\begin{aligned} I(t) &= \frac{1}{2\pi\mathrm{i}}\int_{\mathrm{Br}} \frac{g(0)}{f(0)}\left[1 - \frac{\sqrt{z}}{f(0)} + O(z, z^{3/2})\right]\mathrm{e}^{zt}\mathrm{d}z \\ &= \frac{g(0)}{f(0)}\frac{\mathrm{d}}{\mathrm{d}t}\left\{\frac{1}{2\pi\mathrm{i}}\int_{\mathrm{Br}}\left[1 - \frac{\sqrt{z}}{f(0)} + O(z, z^{3/2})\right]\frac{\mathrm{e}^{zt}}{z}\mathrm{d}z\right\} \\ &\sim -\frac{g(0)}{f^2(0)}\frac{\mathrm{d}}{\mathrm{d}t}[\mathcal{L}^{-1}(z^{-1/2})(t)] \\ &= -\frac{g(0)}{f^2(0)}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{\sqrt{\pi t}}\right) = \frac{1}{2\sqrt{\pi}}\frac{g(0)}{f^2(0)}t^{-3/2}. \end{aligned} \quad (3.23)$$

Here $\mathcal{L}^{-1}(F(z))$ is the inverse Laplace transform of $F(z)$. We note that error terms involving integer powers of z in the Taylor expansion can be interpreted as the distributions $\delta(t)$, $\delta'(t)$, etc., and these vanish for $t > 0$. The fractional power error terms, such as $z^{3/2}$, will invert to a term proportional to $t^{-5/2}$, which is smaller than the leading term.

If $g(0) \neq 0$ but $f(0) = 0$ then again expanding about $z = 0$ leads to

$$I(t) = \frac{1}{2\pi\mathrm{i}}\int_{\mathrm{Br}} g(0)\left[\frac{1}{\sqrt{z}} + O(\sqrt{z})\right]\mathrm{e}^{zt}\mathrm{d}z \sim \frac{g(0)}{\sqrt{\pi t}}. \quad (3.24)$$

If $f(z)+\sqrt{z}=0$ has a solution at $z=z_*$ in the range $\Re(z)>0$, with $f'(z_*)+\frac{1}{2}z_*^{-1/2}\neq 0$ then the simple pole at z_* determines the behavior of $I(t)$ and we obtain

$$I(t)\sim\frac{g(z_*)}{f'(z_*)+\frac{1}{2}z_*^{-1/2}}\mathrm{e}^{z_* t}. \tag{3.25}$$

We can also consider the case where the branch point and pole are close to each other. Then $f(0)$ would be small so we set $f(0)=\varepsilon$. By expanding the integrand about $z=0$ and introducing the (large) time scale $t=\varepsilon^{-2}T$ we have

$$\begin{aligned}I(t)&\sim\frac{1}{2\pi\mathrm{i}}\int_{\mathrm{Br}}\frac{g(0)}{\sqrt{z}+\varepsilon}\mathrm{e}^{zt}\mathrm{d}z\\&=g(0)\left\{\frac{|\varepsilon|\mathrm{sgn}(\varepsilon)}{\sqrt{\pi T}}-\frac{2\varepsilon}{\sqrt{\pi}}\mathrm{e}^{T}\int_{\sqrt{T}\mathrm{sgn}(\varepsilon)}^{\infty}\mathrm{e}^{-u^2}\mathrm{d}u\right\}.\end{aligned} \tag{3.26}$$

For $\varepsilon>0$ and $T\to\infty$ we recover the behavior in (3.23), as the right-hand side of (3.26) becomes $O(T^{-3/2})$. For $\varepsilon<0$ and $T\to\infty$ (3.26) behaves as an exponential, as in (3.25). Finally, if $\varepsilon=0$ (3.26) becomes $g(0)/\sqrt{\pi t}$, so that (3.24) is recovered as a special case.

Since $D_{-\theta}(\cdot)$ is an entire function of θ, we immediately obtain Theorems 5 and 6. When $\beta=0$ or $\beta=\beta_*$ the asymptotics follow from (3.24), when $\beta>\beta_*$ (3.25) applies, while for $\beta<\beta_*$ (with $\beta\neq 0$) (3.23) holds. We must simply identify $f(z)$ and $g(z)$ from Theorems 2 and 3, which necessitates that we distinguish between x, x_0 positive and negative.

3.3. Derivation of the spectral properties

We establish Theorem 7, and hence Theorem 4 about the relaxation time. We first show that any eigenvalue λ must be real. Let $p(x,t)=\mathrm{e}^{\lambda t}q(x)$ be a solution to (2.1) and (2.2). Then setting

$$q(x)=\begin{cases}\mathrm{e}^{-\beta x/2}R(x), & x>0,\\ \mathrm{e}^{-x^2/4}\mathrm{e}^{-\beta x/2}R(x), & x<0,\end{cases} \tag{3.27}$$

we obtain from (2.1) and (2.2)

$$\lambda R=\begin{cases}R_{xx}-\dfrac{1}{4}\beta^2 R, & x>0,\\[2mm] R_{xx}+\left[\dfrac{1}{2}-\dfrac{1}{4}(x+\beta)^2\right]R, & x<0.\end{cases} \tag{3.28}$$

We take the complex conjugates of (3.28) and note that β is real. Multiplying (3.28) by $\bar{R}$ (the complex conjugate of R), multiplying the complex conjugates of (3.28) by R, and subtracting the results, yields

$$(\lambda-\bar{\lambda})R\bar{R}=\bar{R}R_{xx}-R\bar{R}_{xx} \tag{3.29}$$

and this holds for both $x>0$ and $x<0$. By integrating (3.29) over all x and using the fact that R, $\bar{R}$, R_x and $\bar{R}_x$ are continuous at $x=0$ we obtain

$$(\lambda-\bar{\lambda})\int_{-\infty}^{\infty}|R(x)|^2\mathrm{d}x=\int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x}[\bar{R}R_x-\bar{R}_x R]\mathrm{d}x=0. \tag{3.30}$$

Hence $\lambda = \bar{\lambda}$ and any eigenvalue (hence solution of (2.49)) must be real. By multiplying (3.28) by R and integrating by parts we then obtain

$$\lambda \int_{-\infty}^{\infty} R^2 \mathrm{d}x - \frac{1}{2}\int_{-\infty}^{0} R^2 \mathrm{d}x = -\frac{\beta^2}{4}\int_0^{\infty} R^2 \mathrm{d}x - \int_{-\infty}^{0} \frac{(\beta+x)^2}{4} R^2 \mathrm{d}x - \int_{-\infty}^{\infty} R_x^2 \mathrm{d}x. \tag{3.31}$$

From (3.31) we conclude that $\lambda < \frac{1}{2}$, but this is not very sharp since of course $\lambda \le 0$. Next we turn to equation (2.49). To show that for $\beta \in (\beta_{*,N}, \beta_{*,N+1})$, (2.49) has exactly N roots we use a topological argument, showing that a new root ($p = -\theta$) enters the interval $(0, \beta^2/4)$ at the critical values $\beta = \beta_{*,N}$, and that once a root enters this interval it cannot leave. This behavior is also clearly demonstrated numerically by Fig. 2. We recall that $D_p(-\beta)$ is an entire function of both p and β, and then any solution of (2.49), call it $p_N(\beta)$, must be a smooth function of β. This can be obtained by differentiating (2.49) (with $-\theta$ replaced by $p_N(\beta)$) with respect to β and using the analyticity of $D_p(-\beta)$. Thus a root cannot disappear in a discontinuous manner, and if it leaves the interval $p \in (0, \beta^2/4)$ it must do so at one of the endpoints.

For $\beta \in (\beta_*, \beta_{*,2}) = (1.85722\ldots, 2.72133\ldots)$ we can show that (2.49) has a root simply by examining the signs of the left and right sides of (2.49), at $p = 0^+$ and at $p = \beta^2/4$. Consider $D'_{\beta^2/4}(-\beta)$. When $\beta = 0$ this function is zero and it becomes positive for β small and positive (see expression (3.32) below with $p = \beta^2/4$, and also Fig. 1). Then by the definitions of β_* and $\beta_{*,2}$ (as the first two positive roots) it follows that $D'_{\beta^2/4}(-\beta) < 0$ for $\beta \in (\beta_*, \beta_{*,2})$. Thus for β in this range the right side of (2.49) is negative at $p = \beta^2/4$ while the left side vanishes.

Now consider the behavior of (2.49) for $p \to 0^+$ and β fixed. By a Taylor series we have, using (2.6),

$$D'_p(-\beta) = \frac{\beta}{2}\mathrm{e}^{-\beta^2/4} + p\mathcal{A}(\beta) + O(p^2), \tag{3.32}$$

where

$$\begin{aligned}\mathcal{A}(\beta) &= -\frac{\mathrm{d}}{\mathrm{d}\beta}\left[\frac{\mathrm{e}^{\beta^2/4}}{\mathrm{i}\sqrt{2\pi}}\int_C (\log u)\mathrm{e}^{u^2/2}\mathrm{e}^{\beta u}\mathrm{d}u\right] \\ &= \frac{1}{\mathrm{i}\sqrt{2\pi}}\frac{\beta}{2}\mathrm{e}^{\beta^2/4}\int_C (\log u)\mathrm{e}^{u^2/2}\mathrm{e}^{\beta u}\mathrm{d}u + \frac{\mathrm{e}^{\beta^2/4}}{\mathrm{i}\sqrt{2\pi}}\int_C \frac{1}{u}\mathrm{e}^{u^2/2}\mathrm{e}^{\beta u}\mathrm{d}u.\end{aligned} \tag{3.33}$$

Here C is a vertical contour with $\Re(u) > 0$ and we integrated by parts to get the second expression in (3.33). By similarly expanding $D_p(-\beta)$ for $p \to 0$ we have

$$D_p(-\beta) = \mathrm{e}^{-\beta^2/4} + p\mathcal{C}(\beta) + O(p^2), \tag{3.34}$$

where

$$\mathcal{C}(\beta) = \frac{\mathrm{e}^{\beta^2/4}}{\mathrm{i}\sqrt{2\pi}}\int_C (\log u)\mathrm{e}^{u^2/2}\mathrm{e}^{\beta u}\mathrm{d}u. \tag{3.35}$$

Using (3.32)–(3.34) and expanding (2.49) for small p leads to

$$\begin{aligned}\frac{\beta}{2}\sqrt{1-4p/\beta^2}[\mathrm{e}^{-\beta^2/4}+p\mathcal{C}(\beta)+O(p^2)] &= \frac{\beta}{2}\mathrm{e}^{-\beta^2/4}+p\left[\frac{\beta}{2}\mathcal{C}(\beta)-\frac{1}{\beta}\mathrm{e}^{-\beta^2/4}\right] \\ &\quad + O(p^2) \\ &= \frac{\beta}{2}\mathrm{e}^{-\beta^2/4}+p\mathcal{A}(\beta)+O(p^2). \qquad (3.36)\end{aligned}$$

Thus the right side of (3.36) (and hence (2.49)) will exceed the left side if

$$\mathcal{A}(\beta)-\frac{\beta}{2}\mathcal{C}(\beta) > -\frac{1}{\beta}\mathrm{e}^{-\beta^2/4}. \qquad (3.37)$$

But from (3.33) and (3.35) we have

$$\begin{aligned}\mathcal{A}(\beta)-\frac{\beta}{2}\mathcal{C}(\beta) &= \frac{\mathrm{e}^{\beta^2/4}}{\mathrm{i}\sqrt{2\pi}}\int_C u^{-1}\mathrm{e}^{u^2/2}\mathrm{e}^{\beta u}\mathrm{d}u \\ &= \mathrm{e}^{\beta^2/4}\int_{-\infty}^{\beta}\mathrm{e}^{-v^2/2}\mathrm{d}v. \qquad (3.38)\end{aligned}$$

The last equality follows from recognizing that a parabolic cylinder function of order -1 can be expressed in terms of the standard error function.

Thus for $\beta > 0$ (3.37) is clearly satisfied. Hence at $p = 0^+$ the right side of (2.49) exceeds the left side, and this shows that (2.49) has at least one root for $\beta \in (\beta_*, \beta_{*,2})$. Actually, the same argument holds for any interval $(\beta_{*,2M-1}, \beta_{*,2M})$ for $M = 1, 2, \ldots$. We have also shown that $p = 0$ is a root of (2.49) for all $\beta > 0$, but in view of (3.37) $p = 0$ can never be a double root. This shows that a solution $p_N(\beta)$ to (2.49) cannot leave the interval $(0, \beta^2/4)$ at $p = 0$.

We next examine (2.49) near the endpoint $p = \beta^2/4$. We can clearly only have solutions if $p < \beta^2/4$, so we set

$$\beta = \beta_{*,N} + \xi \qquad (3.39)$$

$$p = \frac{1}{4}\beta^2 - q = \frac{1}{4}\beta_{*,N}^2 + \frac{1}{2}\beta_{*,N}\xi + \frac{1}{4}\xi^2 - q. \qquad (3.40)$$

Again using the analyticity of $D_p(-\beta)$ in both p and β we expand (2.49) in Taylor series about $(p, \beta) = (\frac{1}{4}\beta_{*,N}^2, \beta_{*,N})$ to obtain

$$\sqrt{q} = \frac{\xi}{2}[1 + \beta_{*,N}\Omega_N] + O(\xi^2), \qquad (3.41)$$

where

$$\Omega_N = \left.\frac{\frac{\partial}{\partial p}D'_p(-\beta)}{D_p(-\beta)}\right|_{\beta=\beta_{*,N};\ p=\frac{1}{4}\beta_{*,N}^2}. \qquad (3.42)$$

To obtain (3.41) from (2.49) we also divided by $D_p(-\beta)$, expanded this function, and eliminated the second derivative $D''_p(-\beta)$ by using the parabolic cylinder equation $-D''_p(-\beta) = (p + \frac{1}{2} - \frac{1}{4}\beta^2)D_p(-\beta)$. Thus (3.41) is an algebraic curve (in fact a half-parabola) that approximates (2.49) near the birth points of $p_N(\beta)$ (see Fig. 2). If $1 + \beta_{*,N}\Omega_N > 0$ then this curve exists for $\xi > 0$ and not for $\xi < 0$, while $\xi = p = 0$ is a solution for all N. This shows that the curve $p_N(\beta)$ enters the interval $(0, \beta^2/4)$ for $\beta > \beta_{*,N}$ $(\xi > 0)$ rather than exits it. Such a curve can only

Table 1
Some values of Ω_N in (3.42) and the asymptotic expression (3.45).

N	Ω_N	(3.45)
1	2.4108	2.5654
2	2.7927	2.9256
3	3.0305	3.1458
4	3.2054	3.3083
5	3.3449	3.4387
10	3.8020	3.8708
20	4.3013	4.3509
30	4.6165	4.6573
40	4.8517	4.8872
50	5.0412	5.0730

enter the interval at $p = \beta^2/4$ at a root of (2.20). We have thus shown that a new root enters the interval $(0, \beta^2/4)$ at each critical value $\beta = \beta_{*,N}$, and once a new root enters it cannot leave across either endpoint. It remains only to show that $1 + \beta_{*,N}\Omega_N$ is positive for all $N \geq 1$, which will certainly be true if $\Omega_N > 0$.

We have verified numerically that $\Omega_N > 0$ for all $1 \leq N \leq 100$. Furthermore, we have (from [1], Chapter 19) the following asymptotic formulas for $\beta \to \infty$, which are analogous to (2.22)

$$D'_{\beta^2/4}(-\beta) \sim \left(\frac{2}{3}\right)^{1/6} \frac{1}{\sqrt{\pi}} \Gamma(1/3) \left(\frac{\beta}{2}\right)^{\beta^2/4} \mathrm{e}^{-\beta^2/8} \beta^{1/3} \cos(\pi\beta^2/4 + \pi/3) \tag{3.43}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p} D'_p(-\beta)\bigg|_{p=\beta^2/4} \sim {} & \left(\frac{3}{2}\right)^{1/6} \frac{1}{\sqrt{\pi}} \Gamma(2/3) \left(\frac{\beta}{2}\right)^{\beta^2/4} \mathrm{e}^{-\beta^2/8} \beta^{2/3} \\ & \times [\log(\beta/2) \sin(\pi\beta^2/4 + \pi/6) + \pi \cos(\pi\beta^2/4 + \pi/6)].\end{aligned} \tag{3.44}$$

Using (3.43) and (3.44) with $\beta = \beta_{*,N}$ leads to

$$\Omega_N \sim 3^{1/3} \Gamma^2(2/3)(N - 1/6)^{1/6}, \quad N \to \infty. \tag{3.45}$$

Thus Ω_N is positive for N sufficiently large. In Table 1 we give some exact values of Ω_N and compare these to the asymptotic formula in (3.45). The agreement is very good and Table 1 gives strong evidence for the positivity of the sequence Ω_N.

We have thus established the first part of Theorem 7, concerning the roots of (2.19) or (2.49), up to the positivity of the Ω_N, for which we provide strong numerical evidence. Certainly we established Theorem 7 for N up to 100.

To derive (2.51) in Theorem 7 we study asymptotically, as $\beta \to \infty$, the equation

$$\frac{D'_{-\theta}(-\beta)}{D_{-\theta}(-\beta)} = \sqrt{\theta + \beta^2/4}. \tag{3.46}$$

For $\beta \to \infty$ the right-hand side becomes

$$\frac{\beta}{2}\left[1 + \frac{2\theta}{\beta^2} - \frac{2\theta^2}{\beta^4} + O(\beta^{-6})\right]. \tag{3.47}$$

In this limit the parabolic cylinder functions have the expansion

$$D_{-\theta}(-\beta) = (-\beta)^{-\theta} e^{-\beta^2/4} \left[1 - \frac{\theta(\theta+1)}{2\beta^2} + O(\beta^{-4})\right] + \frac{\sqrt{2\pi}}{\Gamma(\theta)} \beta^{\theta-1} e^{\beta^2/4} [1 + O(\beta^{-2})]. \tag{3.48}$$

The second term is exponentially large ($O(e^{\beta^2/4})$) while the first term is exponentially small ($O(e^{-\beta^2/4})$), unless $\theta = 0, -1, -2, \ldots$. In that case $1/\Gamma(\theta)$ vanishes and then $D_n(-\beta)$ is exponentially small, and proportional to the nth Hermite polynomial. Our analysis of (3.46) will show that θ must be very close to a negative integer if (3.46) holds. If this were not the case then the second term in (3.48) would dominate and $D'_{-\theta}(-\beta)/D_{-\theta}(-\beta) \sim -\beta/2$ which could not equal (3.47) for $\beta \to \infty$.

For $\theta \to -N$ we have

$$\Gamma(\theta) = \frac{(-1)^N}{N!} \frac{1}{\theta + N} + O(1), \tag{3.49}$$

which is just the Laurent expansion of $\Gamma(\theta)$ near a pole. To balance the two parts of the right-hand side of (3.48) we need to scale $\theta + N$ to be roughly $O(e^{-\beta^2/2})$, so we define ω_N by

$$\theta + N = \omega_N e^{-\beta^2/2}. \tag{3.50}$$

Then (3.48) becomes

$$D_{-\theta}(-\beta) = e^{-\beta^2/4} \left\{(-\beta)^N \left[1 - \frac{N(N-1)}{2\beta^2} + O(\beta^{-4})\right] + (-1)^N N! \omega_N \beta^{-N-1} [1 + O(\beta^{-2})]\right\}. \tag{3.51}$$

To obtain (3.51) we replaced θ by $-N$ in (3.48) in all factors except $1/\Gamma(\theta)$, where we used (3.49) and (3.50). Up to an exponentially small error, (3.47) becomes

$$\frac{\beta}{2} - \frac{N}{\beta} - \frac{N^2}{\beta^3} + O(\beta^{-5}). \tag{3.52}$$

Computing the logarithmic derivative of (3.48), with the scaling (3.50), and equating the result to (3.52) leads to

$$\frac{\beta}{2} - \frac{N}{\beta} - \frac{N^2}{\beta^3} + O(\beta^{-5}) \sim \frac{-\Delta'(\beta) + \beta\Delta(\beta)/2 - \sqrt{2\pi} N! \omega_N (-\beta)^{-N}/2}{\Delta(\beta) - \sqrt{2\pi} N! \omega_N (-\beta)^{-N-1}}, \tag{3.53}$$

where

$$\Delta(\beta) = e^{\beta^2/4} D_N(-\beta) = (-\beta)^N \left[1 - \frac{1}{2} N(N-1)\beta^{-2} + O(\beta^{-4})\right], \tag{3.54}$$

By using the recurrence in (2.23) to infer the behavior of $D'_N(-\beta)$, or by directly expanding (2.6), we obtain

$$\Delta'(\beta) = N(-\beta)^{N-1} \left[-1 + \frac{1}{2}(N-1)(N-2)\beta^{-2} + O(\beta^{-4})\right], \tag{3.55}$$

so that $\Delta'(\beta)/\Delta(\beta) = N/\beta + N(N-1)/\beta^3 + O(\beta^{-5})$ as $\beta \to \infty$. Thus the right-hand side of (3.53), after some further expansion, becomes

$$\frac{\beta}{2} - \frac{N}{\beta} - \frac{N(N-1)}{\beta^3} - \frac{\sqrt{2\pi}}{\Delta(\beta)} N!(-\beta)^{-N} \omega_N [1 + o(1)]. \tag{3.56}$$

Comparing this to (3.52) we see that the first two terms agree automatically, and agreement of the $O(\beta^{-3})$ terms forces

$$\omega_N \sim \frac{-1}{\sqrt{2\pi} N!} (-\beta)^{N-3} \Delta(\beta) N \sim \frac{\beta^{2N-3}}{\sqrt{2\pi}(N-1)!}. \tag{3.57}$$

We also see that this analysis would predict that $\omega_0 = 0$, and indeed $\theta = 0$ is a solution of (3.46) (exactly) when $\beta > 0$.

4. Discussion and extensions

To summarize, we have obtained explicit expressions for the Laplace transform of the transient density for the Halfin–Whitt diffusion process, and then established various spectral properties, for increasing values of the drift parameter β.

In particular we gave the approach to equilibrium for $\beta > 0$, and observed a "phase transition" when $\beta = \beta_* \approx 1.85722$, which was also observed recently by Gamarnik and Goldberg [8]. But, in [8] the authors analyzed the discrete $M/M/s$ model, located the dominant singularity using the approach in [16] and then evaluated it in the limit of $s \to \infty$ with $1 - \rho = O(s^{-1/2})$. In contrast, we started with the limiting diffusion and located its dominant singularity, inferring the large time behavior. Thus while [8] considers the limits in the order $t \to \infty$ and then $s \to \infty$ (with $1 - \rho = O(s^{-1/2})$), here we reverse the order. The fact that the results for β_* agree shows that such a reversal is possible for this particular model. Perhaps this is due to the fact that the diffusion limit for the $M/M/s$ model in the Halfin–Whitt regime does not involve a re-scaling of time t. To provide a formal proof of the suggested interchange of limits is a challenging open problem.

It is possible to study the discrete $M/M/s$ queue using a Laplace transform over time, and then expressing the solution in terms of hypergeometric functions (see [23]). Then in the limit of $s \to \infty$ (with $1 - \rho = O(s^{-1/2})$ and number of customers $n = s + O(\sqrt{s})$) the hypergeometric functions can be approximated by (the simpler) parabolic cylinder functions, and we would thus regain our Theorems 1 and 2. Such an approach however would likely not extend to the more general $GI/M/s$ model, where the analysis of the discrete model seems intractable.

Since the solution in Theorems 1 and 2 is still quite complicated, it would be useful to evaluate it in further limiting cases. Here we considered only large time t, for fixed values of the drift β and initial condition x_0. It is likely that having $\beta \to \pm\infty$ and/or $x_0 \to \pm\infty$ would produce different, and quite possibly simpler, formulas than those in Theorems 4 and 5. Then space and time would also need to be scaled appropriately, and it may be possible to approximate the parabolic cylinder functions by elementary functions, and then asymptotically invert the Laplace transform by using the saddle point method. We are presently examining some of these different limiting cases.

The exact formulas for the Laplace transform in Theorems 1 and 2 could, with some work, be generalized to the following diffusion model. Suppose we divide the $x-$ axis as $(-\infty, \infty) = I_0 \cup \{x_1^*\} \cup I_1 \cup \{x_2^*\} \cup \cdots \cup \{x_m^*\} \cup I_m$. Here the x_j^* are single points ordered as $-\infty < x_1^* < x_2^* < \cdots < x_m^* < \infty$ and the I_j are the intervals $I_j = (x_j^*, x_{j+1}^*)$, with $I_0 = (-\infty, x_1^*)$

and $I_m = (x_m^*, \infty)$. Then consider a one-dimensional diffusion process with a constant diffusion coefficient (say $B(x) = 2$ as in (2.2)), and linear drift functions $C_j(x) = a_j x + b_j$ for $x \in I_j,\ j = 0, 1, \ldots, m$. The transient solution to this model can be obtained similarly as we did here. Denoting again its Laplace transform by $\hat{p}(x;\theta)$, we can express it as a linear combination of two parabolic cylinder functions on each of the subintervals I_j. The coefficients in this linear combination will still depend on the transform variable θ, and there will be a total of $2m$ coefficients. Then we would impose the interface conditions $\hat{p}(x_j^*-;\theta) = \hat{p}(x_j^*+;\theta)$ and $\hat{p}_x(x_j^*-;\theta) = \hat{p}_x(x_j^*+;\theta)$ for $j = 1, 2, \ldots, m-1$ and this will lead to $2m-2$ equations for the undetermined coefficients. The initial condition x_0 will lie in some subinterval, say I_k. Then within this interval we would need to allow for $\hat{p}(x;\theta)$ to have different forms for $x_k^* < x < x_0$ and $x_0 < x < x_{k+1}^*$, and impose the continuity condition $\hat{p}(x_0^-;\theta) = \hat{p}(x_0^+;\theta)$ and jump condition $\hat{p}_x(x_0^-;\theta) - \hat{p}(x_0^+;\theta) = 1$. Thus within I_k we would have two additional unknown coefficients, but these are compensated for by the two interface conditions at x_0. Finally in the subintervals I_0 and I_m, which extend to $\pm\infty$, one of the two parabolic cylinder functions must be eliminated due to its behavior at infinity, and this yields two additional conditions. We have thus shown that the solution will involve $2m+2$ undetermined coefficients (with 4 coming from the interval I_k that contains x_0), but these are uniquely determined by $2m-2$ interface conditions at x_j^* $(1 \le j \le m-1)$, the two jump conditions at x_0, and the two conditions at infinity.

The model analyzed here is a special case of the above, with only two subintervals (thus $m = 1$) $(-\infty, 0)$ and $(0, \infty)$, and $x_1^* = 0, a_0 = -1, a_1 = 0, b_0 = b_1 = -\beta$. Another important special case is the abandonment model considered in [11]. This model is sometimes denoted by $M/M/s + M$ (or more generally $GI/M/s + GI$) and in [11] it is shown that a diffusion limit also has the above form, with now $a_1 = -\eta$ rather than $a_1 = 0$, where η measures the effects of customers abandoning the queue. Note that if $\eta = 1$ the drift is $C_j(x) = -x - \beta$ for $j = 0, 1$ (thus for all x) and then the model collapses to a standard Ornstein–Uhlenbeck process. Using the method outlined above the Laplace transform can be expressed in terms of parabolic cylinder functions, for both $x < 0$ and $x > 0$.

However, the spectral properties of the abandonment model will necessarily be much different than those of the Halfin–Whitt diffusion studied here. For any model with drift parameters $a_0 < 0$ and $a_m < 0$ the spectrum will be purely discrete, and the approach to equilibrium will thus be purely exponential, with $p(x,t) - p(x,\infty) \sim F(x,x_0)\mathrm{e}^{-rt}$. Here F will have different forms on the different subintervals and the relaxation rate r will depend only on the drift parameters $a_0, a_1, \ldots, a_m; b_0, b_1, \ldots, b_m$ and the interface points $x_1^*, x_2^*, \ldots, x_m^*$. Thus the type of phase transition we observed at $\beta = \beta_*$ will simply not occur in the model with abandonment where $a_0 = -1$ and $a_1 = -\eta$. The model analyzed here can be obtained by setting $\eta = 0$, but the limit $\eta \to 0^+$ is highly singular and the continuous part of the spectrum is present only if $\eta = 0$. It may be, however, interesting to study the abandonment model (say its relaxation rate $r = r(\eta,\beta)$) in the limit of $\eta \to 0^+$, which would presumably show how the spectrum, though necessarily discrete, begins to resemble a continuous one, say via the coalescence of some of the poles of the Laplace transform. We are currently investigating some of these spectral properties.

We are also examining properties of first passage times for some of these diffusion models with a piecewise linear drift. These may lead to very different asymptotic properties from those of the transient distribution. In particular, we may see, say as $\beta \to \infty$, some exponentially large time scales. In this case the question whether the diffusion process can adequately describe the original discrete model on very large time scales will become an important.

Finally we note that for the general model with $(m+1)$ subintervals the transient density and its first derivative will be continuous at the interface points x_j^*, but the second derivatives will generally suffer jumps there. For the Halfin–Whitt diffusion, with or without abandonment, the drift function is continuous at $x=0$ so that the jump will occur only in the third derivatives. This can be seen also from Theorems 1 and 2, which show that $\hat{p}_{xx}(0^+;\theta) = \hat{p}_{xx}(0^-;\theta)$.

Acknowledgements

We would like to thank Ton Dieker, David Gamarnik and David Goldberg for stimulating discussions, and for sharing their unpublished work. The work of Charles Knessl was supported partially by NSF grant DMS-05-03745 and NSA grant H 98230-08-1-0102. The work of Johan van Leeuwaarden was supported by a VENI grant from The Netherlands Organization for Scientific Research (NWO).

References

[1] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, 1970 (9th printing).
[2] J.P.C. Blanc, E.A. van Doorn, Relaxation times for queueing systems, in: J.W. de Bakker, M. Hazewinkel, J.K. Lenstra (Eds.), Mathematics and Computer Science, North-Holland, Amsterdam, 1984, pp. 139–162.
[3] S. Borst, A. Mandelbaum, M. Reiman, Dimensioning large call centers, Oper. Res. 52 (2004) 17–34.
[4] E. Brockmeyer, H.L. Halstrøm, A. Jensen, The Life and Works of A.K. Erlang, Trans. Danish Acad. Tech. Sci. 2 (1948) Denmark..
[5] J.W. Cohen, The Single Server Queue, North Holland, Amsterdam, 1982.
[6] A. Erdelyi, Higher Transcendental Functions, vol. 2, MacGraw-Hill, New York, 1953.
[7] P. Flajolet, R. Sedgewick, Analytic Combinatorics, Cambridge University Press, Cambridge, 2009.
[8] D. Gamarnik, D.A. Goldberg, On the exponential rate of convergence to stationarity in the Halfin–Whitt regime I: the spectral gap of the $M/M/n$ queue, 2008, Preprint.
[9] D. Gamarnik, P. Momčilovic, Queues with many servers: the virtual waiting-time process in the QED regime, 2007, Preprint.
[10] N. Gans, G. Koole, A. Mandelbaum, Telephone call centers: tutorial, review and research prospects, Manuf. Serv. Oper. Manag. 5 (2003) 79–141.
[11] O. Garnett, A. Mandelbaum, M. Reiman, Designing a call center with impatient customers, Manuf. Serv. Oper. Manag. 4 (2002) 208–227.
[12] I.S. Gradshteyn, I.M. Ryzhik, Table of Integrals, Series and Products, 5th ed., Academic Press, New York, 1994.
[13] S. Halfin, W. Whitt, Heavy-traffic limits for queues with many exponential servers, Oper. Res. 29 (1981) 567–588.
[14] A.J.E.M. Janssen, J.S.H. van Leeuwaarden, B. Zwart, Refining square root staffing by expanding Erlang C, Oper. Res. (2009) (in press).
[15] P. Jelenković, A. Mandelbaum, P. Momčilovic, Heavy traffic limits for queues with many deterministic servers, Queueing Syst. 47 (2004) 53–69.
[16] S Karlin, J.L. McGregor, Many server queueing processes with Poisson input and exponential service times, Pacific J. Math. 8 (1958) 87–118.
[17] J. Keilson, A review of transient behavior in regular diffusion and birth–death processes, J. Appl. Probab. 1 (1964) 247–266.
[18] J.F.C. Kingman, The heavy traffic approximation in the theory of queues, in: W.L. Smith, W. Wilkinson (Eds.), Proc. Symp. on Congestion Theory, University of North Carolina Press, Chapel Hill, 1965, pp. 137–169.
[19] C. Maglaras, A. Zeevi, Diffusion approximations for a multiclass Markovian service system with guaranteed and "best-effort" service levels, Math. Oper. Res. 29 (2004) 786–813.
[20] A. Mandelbaum, P. Momčilovic, Queues with many servers: the virtual waiting-time process in the QED regime, 2007, Preprint.
[21] F. Pollaczek, Über zwei Formeln aus der Theorie des Wartens vor Schaltergruppen, Elektr. Nachr. 8 (1931) 256–268.
[22] J. Reed, The G/GI/N queue in the Halfin–Whitt regime, 2008, Preprint.
[23] H.M. Srivastava, B.R.K. Kashyap, Special Functions in Queuing Theory and Related Stochastic Processes, Academic Press, Inc., New York-London, 1982.
[24] E.A. van Doorn, Conditions for exponential ergodicity and bounds for the decay parameter of a birth–death process, Ann. Appl. Probab. 17 (1985) 514–530.