Adv. Appl. Prob. **19**, 560–598 (1987)
Printed in N. Ireland

TRANSIENT BEHAVIOR OF REGULATED BROWNIAN MOTION, I: STARTING AT THE ORIGIN

JOSEPH ABATE* AND
WARD WHITT**, *AT & T Bell Laboratories*

Abstract

A natural model for stochastic flow systems is regulated or reflecting Brownian motion (RBM), which is Brownian motion on the positive real line with constant negative drift and constant diffusion coefficient, modified by an impenetrable reflecting barrier at the origin. As a basis for understanding how stochastic flow systems approach steady state, this paper provides relatively simple descriptions of the moments of RBM as functions of time. In Part I attention is restricted to the case in which RBM starts at the origin; then the moment functions are increasing. After normalization by the steady-state limits, these moment c.d.f.'s (cumulative distribution functions) coincide with gamma mixtures of inverse Gaussian c.d.f.'s. The first moment c.d.f. thus coincides with the first-passage time to the origin starting in steady state with the exponential stationary distribution. From this probabilistic characterization, it follows that the kth-moment c.d.f is the k-fold convolution of the first-moment c.d.f. As a consequence, it is easy to see that the $(k+1)$th moment approaches its steady-state limit more slowly than the kth moment. It is also easy to derive the asymptotic behavior as $t \to \infty$. The first two moment c.d.f.'s have completely monotone densities, supporting approximation by hyperexponential (H_2) c.d.f.'s (mixtures of two exponentials). The H_2 approximations provide easily comprehensible descriptions of the first two moment c.d.f.'s suitable for practical purposes. The two exponential components of the H_2 approximation yield simple exponential approximations in different regimes. On the other hand, numerical comparisons show that the limit related to the relaxation time does not predict the approach to steady state especially well in regions of primary interest. In Part II (Abate and Whitt (1987a)), moments of RBM with non-zero initial conditions are treated by representing them as the difference of two increasing functions, one of which is the moment function starting at the origin studied here.

APPROACH TO STEADY STATE; RELAXATION TIMES; DIFFUSION PROCESSES; QUEUES; STOCHASTIC FLOW SYSTEMS; FIRST-PASSAGE TIMES; INVERSE GAUSSIAN DISTRIBUTION; COMPLETE MONOTONICITY

1. Introduction and summary

We focus on *regulated* (or reflecting) *Brownian motion* (RBM), which is Brownian motion on the positive half line with constant negative drift μ and

Received 27 February 1986; revision received 23 June 1986.
* Postal address: AT & T Bell Laboratories, LC2W-E06, 184 Liberty Corner Road, Warren, NJ 07060, USA.
** Postal address: AT & T Bell Laboratories, MH2C-178, Murray Hill, NJ 07974, USA.

constant diffusion coefficient σ^2, modified by an impenetrable reflecting barrier at the origin. RBM is a natural model for stochastic flow systems; see Harrison (1985). It is the standard diffusion process used to approximate queues; see Gaver (1968), Kleinrock (1976) and Newell (1982). For queues it arises as the limit in the heavy-traffic limit theorems; see Borovkov (1965), (1984), Iglehart and Whitt (1970a,b), Coffman and Reiman (1984) and Flores (1985).

We investigate the transient behavior of RBM. We describe the time-dependent behavior of a system that is stationary except for the initial condition. We thus hope to better understand how stochastic flow systems approach steady state. We develop easily comprehensible closed-form approximations and we prove theorems providing insight about the qualitative behavior.

1.1. *Regulated Brownian motion starting at the origin.* A key idea is to focus on a special initial condition. In particular, we consider RBM *starting at the origin.* This initial condition is easier to analyze because the moments as functions of time are then increasing. We are thus able to normalize by the steady-state limits and analyze these moment functions *probabilistically.* This initial condition is also of interest because it routinely arises in simulations and systems that occasionally restart empty. This special case also provides insight about what happens more generally. Moreover, in Part II (Abate and Whitt (1987a)), we apply approximations for this special case to generate approximations for other initial conditions. In Part II we show that moment functions with non-zero initial conditions can be represented as the difference of two increasing functions, one of which corresponds to the moment function starting at the origin. Methods in this paper are then applied again in Part II to analyze the second increasing function.

It should be noted that the general transient marginal distribution of RBM is well known. In particular, if $R(t)$ represents the state of RBM at time t, then

$$\begin{aligned} P(R(t) \leqq y \mid R(0) = x) = 1 - \Phi\left(\frac{-y + x + \mu t}{\sigma t^{\frac{1}{2}}}\right) \\ - \exp(2\mu y / \sigma^2)\Phi\left(\frac{-y - x - \mu t}{\sigma t^{\frac{1}{2}}}\right) \end{aligned} \tag{1.1}$$

where $\Phi(t)$ is the standard normal c.d.f. (cumulative distribution function) having mean 0 and variance 1; p. 49 of Harrison (1985). If $\mu < 0$, then $P(R(t) \leqq y \mid R(0) = x) \to 1 - \exp(2\mu y / \sigma^2)$ as $t \to \infty$, so that the steady-state distribution is exponential. We believe that (1.1) has great value for understanding the essential nature of the transient behavior of stochastic flow systems. We also believe that there is more to be learned by looking further into the transient behavior.

In this paper we primarily study the moments $E(R(t)^k \mid R(0)=0)$, intending to exploit the fact that the process starts at 0. Our point of departure is Gaver (1968) and Kleinrock (1976). Gaver introduced regulated Brownian motion $\{R(t): t \geqq 0\}$ as an approximation for the virtual wait at time t in an $M/G/1$ queue. Gaver then calculated the Laplace transform of the mean $E(R(t) \mid R(0)=x)$ and inverted it numerically, using the method in Gaver (1966). Gaver's work is summarized in Section 2.9 of Kleinrock (1976). In formula (2.157) there, Kleinrock also inverts the Laplace transform for the special case $E(R(t) \mid R(0)=0)$ to obtain an explicit expression in the time domain involving an incomplete gamma function.

It turns out, however, that even for a general initial condition it is possible to obtain relatively tractable expressions in the time domain for the conditional moments $E(R(t)^k \mid R(0)=x)$. Recently, Mitchell (1985) obtained an explicit expression in the time domain for the first moment function by calculating and differentiating the Laplace–Stieltjes transform of (1.1). Mitchell applies the Laplace transform of $\Phi((t-a)/b)$ to obtain the following result. We have applied his approach to describe the second-moment function as well. (The proof is sketched in Section 6.) Let $\phi(t)$ be the density of $\Phi(t)$.

Theorem 1.1. *If* $\mu=-1$ *and* $\sigma^2=1$, *then*
(a) (Mitchell)

$$E(R(t) \mid R(0)=x) = 2^{-1} + t^{\frac{1}{2}}\phi\left(\frac{t-x}{\sqrt{t}}\right) - (t-x+2^{-1})\left[1-\Phi\left(\frac{t-x}{\sqrt{t}}\right)\right] - 2^{-1}e^{2x}\left[1-\Phi\left(\frac{t+x}{\sqrt{t}}\right)\right]$$

(b)

$$\begin{aligned}E(R(t)^2 \mid R(0)=x) = 2^{-1} &+ ((x-1)t^{\frac{1}{2}} - t^{\frac{3}{2}})\phi\left(\frac{t-x}{\sqrt{t}}\right)\\ &+ ((t-x)^2+t-2^{-1})\left[1-\Phi\left(\frac{t-x}{\sqrt{t}}\right)\right]\\ &+ e^{2x}(t+x-2^{-1})\left[1-\Phi\left(\frac{t+x}{\sqrt{t}}\right)\right].\end{aligned}$$

Corollary 1.1.1. *If* $\mu=-1$ *and* $\sigma^2=1$, *then*
(a) $ER(t) \mid R(0)=0) = 2^{-1} - (t+1)[1-\Phi(t^{\frac{1}{2}})] + t^{\frac{1}{2}}\phi(t^{\frac{1}{2}})$;
(b) $E(R(t)^2 \mid R(0)=0) = 2^{-1} - (1-2t-t^2)[1-\Phi(t^{\frac{1}{2}})] - t^{\frac{1}{2}}(1+t)\phi(t^{\frac{1}{2}})$.

Theorem 1.1 and Corollary 1.1.1 may seem very special because they are restricted to the case $\mu=-1$ and $\sigma^2=1$. However, as noted by Gaver (1968), this restriction is *without loss of generality*. As apparently has been known for a long time, see pp. 57–59 of Chandrasekhar (1943), all other cases with $\mu < 0$ can be obtained from this one by an appropriate choice of measuring units for space and time. For convenience in applications, it is significant that all calculations need be done for only one pair of parameter values. We thus refer to the case $\mu = -1$ and $\sigma^2 = 1$ as *canonical regulated Brownian motion*. We show how canonical RBM can be constructed and how to obtain results for general μ and σ^2 in Section 2. For the rest of the introduction, let $\mu = -1$ and $\sigma^2 = 1$.

As an immediate consequence of Theorem 1.1, we can describe the asymptotic behavior as $t \to \infty$ by applying elementary properties of the normal distribution, in particular, expansions for the tail $1 - \Phi(t)$ for large t; pp. 175, 193 of Feller (1968). An alternate proof of part (a) based on results for the $M/M/1$ queue on p. 180 of Cohen (1982) is also possible. (Additional details about Corollary 1.1.2 are in unpublished appendices available from the authors.) Let $f(t) \sim g(t)$ mean that $f(t)/g(t) \to 1$ as $t \to \infty$.

Corollary 1.1.2. As $t \to \infty$,

(*a*) $ER(\infty) - E(R(t) \mid R(0) = x)$

$$= (2\pi)^{-\frac{1}{2}} e^x e^{-t/2}(2(1-x)t^{-\frac{3}{2}} - 2(x^2 - 6x + 6)t^{-\frac{5}{2}}) + o(e^{-t/2})t^{-\frac{5}{2}})$$

$$\sim \begin{cases} (2/\pi)^{\frac{1}{2}} e^x (1-x) e^{-t/2} t^{-\frac{3}{2}} & \text{if } x \neq 1 \\ -(2/\pi)^{\frac{1}{2}} e^{-t/2} t^{-\frac{5}{2}} & \text{if } x = 1; \end{cases}$$

(*b*) $E(R(\infty)^2) - E(R(t)^2 \mid R(0) = x)$

$$= (2\pi)^{-\frac{1}{2}} e^x e^{-t/2}(8(1-x)t^{-\frac{3}{2}} - 8(x^2 + 9x - 9)t^{-\frac{5}{2}}) + o(e^{-t/2} t^{-\frac{5}{2}})$$

$$\sim \begin{cases} 8(2\pi)^{-\frac{1}{2}} e^x (1-x) e^{-t/2} t^{-\frac{3}{2}} & \text{if } x \neq 1 \\ -8(2\pi)^{-\frac{1}{2}} e^{-t/2} t^{-\frac{5}{2}} & \text{if } x = 1. \end{cases}$$

The constant 2 in the denominator of the exponential argument in Corollary 1.1.2 is the *relaxation time*; see Blanc (1985), Cohen (1982), Keilson (1979) and references there. Corollary 1.1.2 suggests that a simple exponential might be a suitable approximation for sufficiently large t, i.e.,

$$ER(\infty) - E(R(t) \mid R(0) = x) \approx A(x) \exp(-rt) \tag{1.2}$$

for $r = \frac{1}{2}$. A primary purpose of this paper is to investigate this question. Obviously the expression in Theorem 1.1 and the previously established Laplace transform are suitable for generating numerical values. (Theorem 1.1 is especially convenient for obtaining numerical results, using rational approximations for the error function; p. 299 of Abramowitz and Stegun (1972).) We are interested in the possibility of simple approximations such as (1.2) in order to obtain a better understanding. We want to identify structure in the transient behavior of RBM to aid in developing approximations for more complex stochastic flow systems such as the $GI/G/1$ queue.

In fact, we find that such a simple exponential approximation *is* justified for the values of t of primary interest when $R(0)=0$, but not in the form suggested by Corollary 1.1.2 above. We conclude that the rate r should be significantly greater than $\frac{1}{2}$. In particular, we suggest the following simple exponential approximation when $R(0)=0$:

$$\begin{aligned} ER(\infty) - E(R(t) \mid R(0)=0) &\approx [(5-\sqrt{5})/20] \exp(-(3-\sqrt{5})t) \\ &\approx 0{\cdot}138 \exp(-0{\cdot}764t) \quad \text{for} \quad t \geqq 1. \end{aligned} \tag{1.3}$$

(By Corollary 1.1.1, the mean $E(R(t) \mid R(0)=0)$ reaches about 85% of its steady-state limit at $t=1$.) Much of this paper is devoted to justifying approximation (1.3). We summarize our results in the rest of Section 1 and provide additional details in the following sections.

It is worth remarking that the difference between the means in (1.2) and (1.3) also indicates how far the entire distribution $P(R(t) \leqq y \mid R(0)=0)$ is from the exponential limit, because, as noted below after Theorem 1.2, the process $(R(t) \mid R(0)=0)$ is stochastically increasing, so that $P(R(t) > y \mid R(0)=0) \leqq \exp(-2y)$ for all y. Consequently, the L_1 norm of the difference between the c.d.f.'s is the difference between the means, i.e.,

$$\begin{aligned} &\|P(R(\infty) \leqq \cdot) - P(R(t) \leqq \cdot \mid R(0)=0)\|_1 \\ &\quad \equiv \int_0^\infty |P(R(\infty) \leqq y) - P(R(t) \leqq y \mid R(0)=0)| \, dy \\ &\quad = \int_0^\infty P(R(\infty) > y) \, dy - \int_0^\infty P(R(t) > y \mid R(0)=0) \, dy \\ &\quad = ER(\infty) - E(R(t) \mid R(0)=0). \end{aligned}$$

1.2. *Moment c.d.f.'s.* We normalize the moment functions by dividing by the steady-state limits, defining

$$H_k(t) = E(R(t)^k \mid R(0)=0)/E(R(\infty)^k), \qquad t \geqq 0. \tag{1.4}$$

The normalization in (1.4) helps interpretation, because it separates the steady-state limit $E(R(\infty)^k)$ from the proportion of this limit reached at time t. The normalization and the condition $R(0)=0$ also allow us to study the moment functions *probabilistically*.

Theorem 1.2. For each k, $E(R(t)^k \mid R(0)=0)$ is increasing in t, so that $H_k(t)$ in (1.4) is a legitimate c.d.f.

There are several ways to prove Theorem 1.2. One is by noting that the Markov transition kernel of RBM is stochastically monotone, Chapter 4 of

Stoyan (1983), so that the c.d.f. in (1.1) is stochastically increasing in t when $R(0)=0$; Section 1.2 of Stoyan (1983). (Other proofs appear in Section 3.) This stochastic order in t also holds for random initial conditions if $R(0)$ is less than or equal to the steady-state limit in the monotone likelihood-ratio ordering, i.e., if $R(0)$ has (in addition to a possible probability mass at the origin) a density $g(y)$ such that $g(y)/2\exp(-2y)$ is decreasing in y for all y; see Section 3 of Keilson and Kester (1977) for the birth-and-death analog, from which the diffusion result follows. (See van Doorn (1980) for related results.) The stochastic monotonicity argument also shows that $H_k(t)$ in (1.4) is *strictly* increasing. It is easy to see that the conditional first moment is initially *decreasing* if $P(R(0)=x>0)=1$. (This is proved in Part II.)

This probabilistic view leads us to new interpretations of the *moment c.d.f.'s* $H_k(t)$. In fact, Theorem 1.2 above is a trivial corollary of Theorem 1.3 below. Let T_{ab} be the first-passage time from a to b for ordinary (unregulated) Brownian motion with parameters $\mu=-1$ and $\sigma^2=1$, and let $f(t;a,b)$ be its density. Obviously T_{ab} is the same with and without the barrier at 0 when $a>b$, but not otherwise. Recall that T_{x0} has the inverse Gaussian density

$$f(t;x,0)=\frac{x}{\sqrt{2\pi t^3}}\exp\left[-\frac{(x-t)^2}{2t}\right],\qquad t>0, \tag{1.5}$$

with associated c.d.f.

$$F(t;x,0)=\Phi\left(\frac{t-x}{\sqrt{t}}\right)+e^{2x}\Phi\left(\frac{-t-x}{\sqrt{t}}\right),\qquad t\geqq 0, \tag{1.6}$$

and Laplace transform

$$\begin{aligned}\hat{f}(s;x,0)&\equiv E(\exp(-sT_{x,0})=\int_0^t\exp(-st)f(t;x,0)\,dt\\&=\exp\{-x[(1+2s)^{\frac{1}{2}}-1]\};\end{aligned} \tag{1.7}$$

p. 363 of Karlin and Taylor (1975), p. 137 of Johnson and Kotz (1970) and p. 221 of Cox and Miller (1965). The inverse Gaussian distribution in (1.5)-(1.7) plays a central role, both here and in Part II; all the quantities of interest can be expressed in terms of it.

Here is a principal result of this paper, proved in Section 3.

Theorem 1.3. For each k,

$$H_k(t)=\int_0^\infty g_k(x)F(t;x,0)\,dx,$$

where $g_k(x)$ is a gamma density, the density of the sum of k i.i.d. exponential random variables each with mean 1/2, and $F(t;x,0)$ is the first-passage-time c.d.f. in (1.6).

In the case $k=1$, the gamma density $g_k(x)$ in Theorem 1.3 reduces to the exponential stationary distribution. Hence, we have the following interesting corollary. (An alternate proof using a coupling construction appears in Part II.)

Corollary 1.3.1. The first-moment c.d.f. $H_1(t)$ in (1.4) coincides with the c.d.f. of the equilibrium time to emptiness, i.e., the first-passage time to 0 starting in steady state with the exponential stationary distribution.

We can also relate the higher moment c.d.f.'s to the first-moment c.d.f. $H_1(t)$ in a simple way (proved in Section 3).

Corollary 1.3.2. For each $k \geqq 1$, $H_k(t)$ is the k-fold convolution of $H_1(t)$.

Corollary 1.3.2 is convenient for comparing the rate of approach to steady state of the moments $E(R(t)^k \mid R(0)=0)$ for different k. Intuitively, we would expect that higher moments approach steady state more slowly. Corollary 1.3.2 makes this property easy to express and establish.

Corollary 1.3.3. For all $k \geqq 1$ and $t \geqq 0$, $H_k(t) \geqq H_{k+1}(t)$; i.e., the moment c.d.f.'s $H_k(t)$ are stochastically increasing in k.

Theorem 1.3 and Corollary 1.3.2 make it easy to compute the moments of the moment c.d.f.'s. Since $E(T_{x0}^j)$ is known and relatively simple for small j, we obtain simple formulas for the j of primary interest. In particular, since $ET_{x0}=x$, $\operatorname{Var}(T_{x0})=x$ and $E(T_{x0}^3)-3E(T_{x0})E(T_{x0}^2)+2E(T_{x0})^3=3x$, p. 139 of Johnson and Kotz (1970), $E(T_{x0}^2)=x+x^2$ and $E(T_{x0}^3)=3x+3x^2+x^3$. With higher moments, it is natural to work with cumulants (semi-invariants); p. 20 of Johnson and Kotz (1970). Since $H_k(t)$ is the k-fold convolution of $H_1(t)$, the jth cumulant of $H_k(t)$ is just k times the jth cumulant of $H_1(t)$. Similarly, since the inverse Gaussian distribution is infinitely divisible, the jth cumulant of T_{x0} is just x times the jth cumulant of T_{10}.

Corollary 1.3.4. The moment-c.d.f. moments are

$$m_{kj} \equiv \int_0^\infty t^j dH_k(t) = \int_0^\infty g_k(x)E(T_{x0}^j)\,dx.$$

For each $k \geqq 1$, the first three moments of $H_k(t)$ are

$$m_{k1}=\frac{k}{2}, \qquad m_{k2}=\frac{k}{2}+\frac{k(k+1)}{4}=\frac{k(k+3)}{4},$$
$$m_{k3}=\frac{3k}{2}+\frac{3k(k+1)}{4}+\frac{k(k+1)(k+2)}{8}=\frac{k(k+4)(k+5)}{8}.$$

Theorem 1.3 provides a basis for extending Corollary 1.1.2 to describe the asymptotic behavior of the complementary c.d.f. or survival function $1 - H_k(t)$ and the density $h_k(t)$ as $t \to \infty$ for all k. For the first-passage-time density in (1.5), it is easy to see that $f(t; x, 0) \sim xe^x(2\pi)^{-\frac{1}{2}}t^{-\frac{3}{2}} \exp(-t/2)$ as $t \to \infty$. Hence, we obtain the following from Theorem 1.3 (proved in Section 3).

Corollary 1.3.5. For each $k \geqq 1$, as $t \to \infty$

$$h_k(t) \sim k2^k(2\pi)^{-\frac{1}{2}}t^{-\frac{3}{2}} \exp(-t/2)$$

and

$$1 - H_k(t) \sim 2h_k(t) \sim k2^{k+1}(2\pi)^{-\frac{1}{2}}t^{-\frac{3}{2}} \exp(-t/2).$$

Our proof of Theorem 1.3 uses the following relation among various first-passage-time distributions for canonical (unregulated) Brownian motion with negative drift, due to *reversibility*, which is also proved in Section 3. (See Keilson (1979) and Kelly (1979) for background on reversibility.)

Theorem 1.4. For all $x > 0$, $P(T_{0x} \leqq t) = \exp(-2x)P(T_{x0} \leqq t)$, $t \geqq 0$.

Combining (1.5) and Theorem 1.4, we have the following corollary.

Corollary 1.4.1.

$$f(t; 0, x) = \exp(-2x)f(t; x, 0) = \frac{x}{\sqrt{2\pi t^3}} \exp\left[\frac{-(x+t)^2}{2t}\right], \qquad t \geqq 0.$$

1.3. *Laplace transforms*. The results in Section 1.2 are established by probabilistic methods in Section 3. Additional insight can be obtained by considering Laplace transforms. Let $f(y, t)$ be the density of the c.d.f. (1.1) at time t under the condition $R(0) = 0$. As noted by Gaver (1968) in (2.11) there, the time-transform of $f(y, t)$ has an especially simple exponential form, namely,

$$\hat{f}(y, s) \equiv \int_0^\infty \exp(-st)f(y, t)\, dt = s^{-1}r_1(s) \exp(-r_1(s)y) \tag{1.8}$$

where $r_1(s) = 1 + (1 + 2s)^{\frac{1}{2}}$. Note that the Laplace transform of the inverse Gaussian density $f(t; x, 0)$ in (1.7) is $\exp(-xr_2(s))$ where $r_2(s) = (1 + 2s)^{\frac{1}{2}} - 1$. The functions $r_1(s)$ and $-r_2(s)$ are the roots of the equations $r^2 - 2r - 2s = 0$; see (2.6) and (2.7) of Gaver (1968), so that $r_1r_2 = 2s$ and $r_1 - r_2 = 2$. Consequently, the Laplace transform of the first-passage-time density $f(t; 0, x)$ in Corollary 1.4.1 is $\exp(-x(r_2 + 2)) = \exp(-xr_1(s))$.

We wish to draw attention to the *separable form* of (1.8): the space variable y appears separately as a simple factor in the exponent. It is thus easy to obtain the time-transform of the moment function:

$$
\begin{aligned}
\hat{m}_k(s) &\equiv \int_0^\infty \exp(-st) E(R(t)^k \mid R(0)=0)\, dt \\
&= \int_0^\infty \exp(-st) \int_0^\infty y^k f(y,t)\, dy\, dt \qquad (1.9) \\
&= \int_0^\infty y^k s^{-1} r_1(s) \exp(-r_1(s) y)\, dy = k!\, s^{-1} r_1(s)^{-k}.
\end{aligned}
$$

The time-transform of the moment c.d.f. $H_k(t)$ in (1.4) is just a scalar times (1.9). More interestingly, we can represent the time-transform of the density $h_k(t)$ of $H_k(t)$ as a simple product. In particular,

$$
\hat{h}_k(s) \equiv \int_0^\infty \exp(-st) h_k(t)\, dt = [2/r_1(s)]^k. \qquad (1.10)
$$

Hence, we have another proof that $h_k(t)$ is the k-fold convolution of $h_1(t)$, as stated in Corollary 1.3.2. Moreover, we can apply (1.10) to calculate the moments given in Corollary 1.3.4. Since

$$
\hat{h}_k(s) = \left(\frac{2}{r_1(s)}\right)^k \equiv \left(\frac{\sqrt{1+2s}-1}{s}\right)^k,
$$

and $\sqrt{1+2s} = 1 + s - (1/2)s^2 + (1/2)s^3 - (5/8)s^4 + O(s^5)$, we obtain $m_{k1} = k/2$, $m_{k2} = k(k+3)/4$ and $m_{k3} = k(k+4)(k+5)/8$ as in Corollary 1.3.4.

In Section 4 we apply (1.10) to derive the following recursion.

Theorem 1.5. For each $k \geqq 1$,

$$
h_{k+1}(t) = 2[1 - H_k(t)] - 2[1 - H_{k-1}(t)], \qquad t \geqq 0,
$$

or, equivalently,

$$
2[1 - H_k(t)] = h_2(t) + \cdots + h_{k+1}(t), \qquad t \geqq 0,
$$

where $H_0(t) = 1$.

Corollary 1.5.1. The second-moment c.d.f. $H_2(t)$ is the stationary-excess (or equilibrium residual life) c.d.f. associated with $H_1(t)$, i.e.,

$$
h_2(t) = 2[1 - H_1(t)] = 4(1+t)[1 - \Phi(t^{\frac{1}{2}})] - 4t^{\frac{1}{2}}\phi(t^{\frac{1}{2}}).
$$

Note that Corollaries 1.3.2 and 1.5.1 say that $H_2(t)$ is simultaneously the convolution of $H_1(t)$ with itself and the associated stationary-excess c.d.f.. Except for the measuring units (time scale), this turns out to *characterize* $H_1(t)$, as we show in Section 4.3.

Corollary 1.5.2. A c.d.f. $H(t)$ on the positive real line with mean m has its convolution equal to its stationary-excess c.d.f. if and only if $H(t)=H_1(t/2m)$, $t\geqq 0$.

We also can invert the transforms to obtain explicit expressions for $h_k(t)$ in the time domain in terms of parabolic cylinder functions; see Chapter 19 of Abramowitz and Stegun (1972). In particular, from (4.22) and (4.25) on p. 240 of Oberhettinger and Badii (1973) together with $D_{-n-1}(z)$ in Section 8.1.2 on p. 326 of Magnus et al. (1966), we can identify the inverse $h_k(t)$ of the transform (1.10). The following provides another proof of Corollary 1.3.5.

Theorem 1.6. For each k,

$$h_k(t)=k2^k(2\pi)^{-\frac{1}{2}}t^{-\frac{3}{2}}\exp(-t/2)w(k,t)$$

where

$$w(k,t)=t^{(k+1)/2}\exp(t/4)u(k+\tfrac{1}{2},t^{\frac{1}{2}})\to 1 \quad\text{as}\quad t\to\infty$$

and

$$u(k+\tfrac{1}{2},t)=\exp(t^2/4)(k!)^{-1}\int_t^\infty(u-t)^k\exp(-u^2/2)\,du$$

using 19.8.1 and 19.14.3 of Abramowitz and Stegun (1972).

1.4. *Approximations for the moment c.d.f.'s.* Looking at the conditional moments $E(R(t)^k \mid R(0)=0)$ via the associated moment c.d.f.'s $H_k(t)$ in (1.4) also suggests an approximation scheme. Since the moments of $H_k(t)$ are readily available via Corollary 1.3.4, it is natural to approximate $E(R(t)^k \mid R(0)=0)$ by fitting a convenient c.d.f. to the moments of $H_k(t)$. Using moments as the basis for fitting c.d.f.'s is also likely to be appropriate if we are primarily interested in a good fit for relatively large t, which is our goal here. In particular, for $k=1$ and 2, we suggest using H_2 (hyperexponential) distributions for this purpose. They are mixtures of two exponentials: i.e., an H_2 density has the form

$$h(t)=p_1\lambda_1\exp(-\lambda_1 t)+p_2\lambda_2\exp(-\lambda_2 t),\ t\geqq 0, \tag{1.11}$$

where $p_1+p_2=1$. We fit an H_2 distribution by matching the first three moments to the three parameters in (1.11). The H_2 approximation in (1.11) may itself seem rather complicated. It is significant in part because it is the basis for developing the simple exponential approximation (1.3). As we will show, one exponential component of (1.11) alone tends to be an excellent approximation for $H_1(t)$ in the region of primary interest, and this leads to (1.3).

This H_2 approximation procedure is supported for $k=1$ and 2, which are the cases of primary interest, by the following result. Recall that a function $h(t)$ is *completely monotone* if it has nth derivative $h^{(n)}(t)$ for all n and $(-1)^n h^{(n)}(t) \geqq 0$ for all n and t; see Keilson (1979). By Bernstein's theorem, a probability density function is completely monotone if and only if it is a mixture of exponential densities. The following is proved in Section 4.

Theorem 1.7. For $k=1$ and 2, the moment c.d.f. $H_k(t)$ has a completely monotone density $h_k(t)$, i.e., is a mixture of exponential c.d.f.'s.

Corollary 1.7.1. There is one and only one H_2 c.d.f. matching the first three moments of $H_k(t)$ for $k=1$ and 2.

It is not possible to fit an H_2 c.d.f. to the first three moments m_{k1}, m_{k2} and m_{k3} of $H_k(t)$ for $k \geqq 3$ because $c_3^2 \equiv (m_{32} - m_{32}^2)/m_{32}^2 = 1$ and $m_{33}/m_{32}^2 m_{31} = 41/81 < 1.5$ and $c_k^2 < 1$ for $k \geqq 4$; Section 3 of Whitt (1982). In fact, we will show that the density $h_3(t)$ is actually increasing near the origin, so that $h_3(t)$ is not even monotone; see Section 4.3.

Complete monotonicity also implies nice structural properties for the distribution; see Keilson (1979).

Corollary 1.7.2. For $k=1$ and 2, the moment density $h_k(t)$ and the complementary c.d.f. $1-H_k(t)$ are log-convex. (The latter is equivalent to $H_k(t)$ being DFR; e.g., p. 74 of Keilson (1979).)

We now briefly indicate how the stationary-excess relationship in Corollary 1.5.1 can be applied. We can combine Corollary 1.7.2 with the continuous-time analog of Theorem 3.1 (i) of Whitt (1985) to obtain a stochastic comparison stronger than Corollary 1.3.3.

Corollary 1.7.3. $H_2(t)$ is stochastically greater than $H_1(t)$ in the likelihood-ratio ordering; i.e., the ratio $h_2(t)/h_1(t)$ of the densities is non-decreasing in t.

1.5. *The approach to steady state.* We sought the results above to understand better how RBM approaches steady state. Intuitively, the physics of the process is not too hard to understand. Except for the barrier at the origin, regulated Brownian motion is a homogeneous process moving with constant negative drift. The barrier counteracts this downward motion. Because of the barrier at the origin, $E(R(t) \mid R(0)=0)$ increases. Since $E(R(t) \mid R(0)=0)$ increases, the barrier moves farther away as t increases, so that the *rate* of increase of $E(R(t) \mid R(0)=0)$ *decreases* as t increases.

The rate can actually be measured in several ways. The *absolute rate* is described by the density $h_1(t)$, which is indeed decreasing as a consequence of Theorem 1.7. The rate relative to the amount to go is described by the *failure rate* or *hazard rate* $h_1(t)/[1-H_1(t)]$, which is the derivative of $-\log[1-H_1(t)]$. By Corollary 1.7.2, $\log[1-H_1(t)]$ is convex, so that the failure rate is also decreasing.

Practically, this analysis means that the approach to steady state of $H_1(t)$ should probably not be especially well described by Corollary 1.1.2. The inverse of the relaxation time is a candidate approximation for the rate; i.e., Corollary 1.1.2 suggests that $h_1(t)/[1-H_1(t)] \approx \frac{1}{2}$. However, Corollary 1.7.2 shows that the failure rate is decreasing, so that the limit $\frac{1}{2}$ is a lower bound. While we do not directly consider queueing processes here, our results support empirical findings for queues in Roth (1981), Odoni and Roth (1983), Lee (1985) and Lee and Roth (1986) showing that the inverse of the relaxation time is only a crude lower bound on the rate of approach to steady state. The relaxation time seems to describe the approach to steady state only for very large t, beyond the region of primary practical interest. In other words, the relaxation time seems to describe the system response to small perturbations from equilibrium.

These observations are supported by Table 1 which gives numerical values for the complementary first-moment c.d.f. $1-H_1(t)$, the density $h_1(t)$ and the failure rate $h_1(t)/[1-H_1(t)]$. (Table 1 is based on Corollary 1.1.1 and values of $\phi(t)$ and $\Phi(t)$ from Table 26.1 in Abramowitz and Stegun (1972).) We are primarily interested in the times required for $E(R(t) \mid R(0)=0)$ to reach 85%–99% of the steady-state value $E(R(\infty))$. Table 1 shows that this occurs for $1 \leqq t \leqq 4$. The appropriate rate (failure rate) in this region is approximately 1·5-2·0 times the inverse of the relaxation time.

A candidate approximation for the approach to equilibrium is a simple exponential fit to the first moment of $H_k(t)$, i.e.,

$$(1.12)\quad 1-H_k(t)=\frac{E(R(\infty)^k)-E(R(t)^k \mid R(0)=0)}{E(R(\infty)^k)} \approx \exp\left(-t/m_{k1}\right), \qquad t \geqq 0,$$

where m_{k1} is the mean of $H_k(t)$. (By Corollary 1.3.4, $m_{k1}=k/2$.) For example, for $k=1$, $m_{11}=\frac{1}{2}$ whereas the relaxation time, say τ_1, from Corollary 1.1.2, is $\tau_1=2$. The estimate $1/m_{11}=2$ for the rate of approach to steady state in (1.12) is thus four times greater than predicted by the relaxation time alone, i.e., $1/\tau_1=0{\cdot}5$. We remark that approximation (1.12) for $k=1$ was suggested for queues by Davis (1960); m_{11} coincides with his notion of *build-up time.*

We have seen that τ_1^{-1}, the inverse of the relaxation time, is properly viewed as a *lower bound* on the rate of approach of $H_1(t)$ to steady state because of the log-convexity of $H_1(t)$ established in Corollary 1.7.2. In a certain sense, m_{11}^{-1} is an *upper bound* on this rate. By Theorem 1.7, the actual distribution $H_1(t)$ is a mixture of exponential distributions, so that $H_1(t)$ is more variable or spread out than the single exponential distribution with the same mean, as given in (1.12). This can be made precise using the convex stochastic ordering; p. 8 of Stoyan (1983). In a certain sense, the convex ordering shows that the

TABLE 1
Numerical values associated with the first-moment c.d.f. $H_1(t)$

Time t	$t^{\frac{1}{2}}$	Complementary c.d.f. $1-H_1(t)$	Density $h_1(t)$	Failure rate $h_1(t)/[1-H_1(t)]$
0·0004	0·02	0·968	38·9	40·2
0·01	0·10	0·850	7·02	8·26
0·04	0·20	0·719	3·07	4·27
0·09	0·30	0·604	1·778	2·94
0·16	0·40	0·505	1·152	2·28
0·25	0·50	0·419	0·791	1·89
0·36	0·60	0·346	0·562	1·62
0·49	0·70	0·284	0·408	1·44
0·64	0·80	0·231	0·301	1·30
0·81	0·90	0·187	0·223	1·19
1·00	1·00	0·151	0·1666	1·11
1·21	1·10	0·120	0·1248	1·04
1·44	1·20	0·095	0·0935	0·98
1·69	1·30	0·075	0·0700	0·93
1·96	1·40	0·059	0·0524	0·89
2·25	1·50	0·046	0·0391	0·86
2·56	1·60	0·035	0·0291	0·82
2·89	1·70	0·027	0·0215	0·80
3·24	1·80	0·021	0·0159	0·77
3·61	1·90	0·015	0·0116	0·75
4·00	2·00	0·012	0·0085	0·74
6·25	2·50	0·0024	0·0016	0·67
9·00	3·00	0·00041	0·00025	0·63
12·25	3·50	0·000056	0·000033	0·60
16·00	4·00	0·0000062	0·0000039	0·58

approach to steady state is slower than an exponential with the same mean. The following is another consequence of Theorem 1.7.

Corollary 1.7.3. For all $t \geqq 0$, $\int_t^\infty \exp(-2u)\, du \leqq \int_t^\infty [1-H_1(u)]\, du$.

The practical significance of this analysis is that we have precise relations supporting the idea that (1.12) is optimistic, whereas the relaxation time in Corollary 1.1.2 is pessimistic.

As a more refined approximation for $k=1$ and 2, we propose the H_2 approximation (1.11) based on the first three moments of $H_k(t)$. For example, when $k=1$, the first three moments of the c.d.f. $H_1(t)$ are $m_{11}=\frac{1}{2}$, $m_{12}=1$ and $m_{13}=\frac{15}{4}$. Hence, after carrying out the H_2 c.d.f. fit as indicated in Section 5, our approximation is

$$
\begin{aligned}
(1.13)\qquad 1 - H_1(t) &\approx \left(\frac{5+\sqrt{5}}{10}\right)\exp\left(-(3+\sqrt{5})t\right) + \left(\frac{5-\sqrt{5}}{10}\right)\exp\left(-(3-\sqrt{5})t\right) \\
&\approx 0{\cdot}7236\exp(-5{\cdot}2356t) + 0{\cdot}2764\exp(-0{\cdot}7639t), \qquad t \geqq 0.
\end{aligned}
$$

For this approximate H_2 c.d.f., say $\tilde{H}_1(t)$, numerical values of the complementary c.d.f. $1-\tilde{H}_1(t)$, density $\tilde{h}_1(t)$ and failure rate $\tilde{h}_1(t)/[1-\tilde{H}_1(t)]$ are displayed in Table 2. The approximations in (1.12) and (1.13) are compared with the exact values for $1-H_1(t)$ in Table 3.

In fact, both approximations (1.12) and (1.13) can be viewed as H_2 approximations. The simple exponential can be thought of as the natural H_2 fit to only the first moment (the two component exponentials in the H_2 both have

TABLE 2
Numerical values associated with the H_2 approximation (1.13) for the first-moment c.d.f. $H_1(t)$

Time t	Complementary c.d.f. $1-\tilde{H}_1(t)$	Density $\tilde{h}_1(t)$	Failure rate $\tilde{h}_1(t)/[1-\tilde{H}_1(t)]$
0	1·00	3·99	3·99
0·0004	0·998	3·99	3·99
0·01	0·961	3·81	3·96
0·04	0·855	3·28	3·83
0·09	0·710	2·56	3·61
0·16	0·558	1·83	3·28
0·25	0·424	1·19	2·81
0·36	0·320	0·74	2·31
0·49	0·246	0·44	1·79
0·64	0·195	0·26	1·33
0·81	0·159	0·17	1·07
1·00	0·133	0·118	0·89
1·21	0·111	0·091	0·82
1·44	0·092	0·072	0·78
1·69	0·071	0·059	0·78
1·96	0·062	0·047	0·76
2·25	0·050	0·038	0·76
2·56	0·039	0·030	0·76
2·89	0·030	0·023	0·76
3·24	0·023	0·0178	0·76
3·61	0·0175	0·0134	0·76
4·00	0·0130	0·0099	0·76
6·25	0·0023	0·00178	0·76
9·00	0·00029	0·00022	0·76
12·25	0·000024	0·000018	0·76
16·00	0·0000014	0·0000010	0·76

TABLE 3
A comparison of the hyperexponential, exponential and asymptotic approximations with exact values of $1 - H_1(t)$, the complementary first-moment c.d.f.

Time t	Exact in Corollary 1.1.1	3-moment H_2 Fit in (1.13)	2-moment H_2 Fit bal. means	1-moment fit exponential e^{-2t} in (1.12)	Asymptotic as $t \to \infty$ $(8/\pi)^{\frac{1}{2}} t^{-\frac{3}{2}} e^{-t/2}$
0·01	0·85	0·96	0·97	0·98	
0·04	0·72	0·86	0·94	0·92	
0·09	0·60	0·71	0·77	0·84	
0·16	0·51	0·56	0·60	0·73	
0·25	0·42	0·42	0·49	0·61	
0·36	0·35	0·32	0·41	0·49	
0·49	0·28	0·25	0·27	0·38	
0·64	0·23	0·20	0·21	0·28	
0·81	0·19	0·16	0·15	0·20	
1·00	0·15	0·13	0·11	0·14	0·97
1·21	0·12	0·11	0·09	0·09	0·66
1·44	0·095	0·092	0·069	0·056	0·45
1·69	0·075	0·071	0·058	0·034	0·31
1·96	0·059	0·062	0·048	0·020	0·22
2·25	0·046	0·050	0·039	0·011	0·15
2·56	0·035	0·039	0·033	0·0060	0·108
2·89	0·027	0·030	0·027	0·0031	0·077
3·24	0·021	0·022	0·022	0·0015	0·054
3·61	0·015	0·018	0·018	0·0007	0·038
4·00	0·012	0·013	0·014	0·0003	0·017
6·25	0·0024	0·0023	0·0038	0·000004	0·0045
9·00	0·00041	0·00029	0·00075	—	0·00066
12·25	0·000056	0·000024	0·00011	—	0·000081

the same mean), while (1.13) is the fit to the first three moments. There is also a commonly used two-moment H_2 fit, where the third parameter is eliminated by assuming balanced means, i.e., $p_1\lambda_1^{-1} = p_2\lambda_2^{-1}$; see (3.7) of Whitt (1982). The two-moment H_2 approximation is also displayed in Table 3. Obviously using more moments helps: the three-moment fit (1.13) is significantly better than the two-moment fit, which in turn is significantly better than the one-moment fit (1.12). Even the three-moment H_2 fit (1.13) is not good for small t, but (1.13) seems adequate for engineering purposes for $t > 0{\cdot}25$ or, equivalently, for $H_1(t) \geqq 0{\cdot}60$. The asymptotic value $(8/\pi)^{\frac{1}{2}} t^{-\frac{3}{2}} \exp(-t/2)$ from Corollary 1.1.2 is also displayed in Table 3. It only seems to be a reasonable approximation for $t \geqq 4.0$ with $H_1(4) = 0{\cdot}988$. It is about as accurate as the H_2 approximation only for $t \geqq 9$ with $H_1(9) = 0{\cdot}9996$. It is not highly accurate even then. The asymptotic value gets to about 10% error when t is about 49 with $1 - H_1(49) \approx 10^{-15}$.

In Table 3, we see that the asymptotic value is indeed an upper bound on $1-H_1(t)$, as indicated by Corollary 1.7.2. The fact that the exponential tail crosses $1-H_1(t)$ exactly once illustrates the convex stochastic ordering in Corollary 1.7.3. In contrast, note that the three-moment H_2 approximation crosses $1-H_1(t)$ three times. Since the H_2 c.d.f. (1.13) is more variable than the exponential (1.12) in the convex stochastic ordering, these tails cross only once too. In general, approximations (1.12) and (1.13) tend to overestimate $1-H_1(t)$ for small t and underestimate for large t.

In a certain sense, the H_2 density (1.13) also provides a bound on the true distribution $H_1(t)$. In particular, both (1.12) and (1.13) are bounds in a stochastic ordering based on Laplace transforms; see Theorem 2 in Whitt (1984). As a consequence, for $k=1$ we have all moments of the exponential approximation (1.12) less than the corresponding moments of $H_1(t)$, which are in turn less than the corresponding moments of the H_2 approximation (1.13). (We derive all moments of $H_1(t)$ in Section 4.)

1.6. *Three regimes.* In our view, there are roughly *three regimes,* with successively decreasing rates of approach to equilibrium. The first regime applies to small t, the second to medium t, and the third to large t. The first two regimes and their rates are reasonably well described by the two exponential components of the H_2 c.d.f. approximation for $H_1(t)$, whereas the third regime is described by the relaxation time. To describe $E(R(t) \mid R(0)=0)$, the second regime seems most important for practical purposes. Of course, by Corollary 1.7.2, the actual rate of approach to equilibrium is steadily decreasing, going from ∞ at $t=0$ to about 0·5 as $t \to \infty$, but thinking of three regimes seems to help for practical understanding.

We regard 5·236 ($1/5{\cdot}236=0{\cdot}191$), 0·764 ($1/0{\cdot}764=1{\cdot}309$) and 0·500 ($1/0{\cdot}50=2{\cdot}00$) as appropriate approximate rates (relaxation times) for $H_1(t)$ in the three regimes. Table 1 shows that 0·76, the rate in the second regime, more realistically describes the rate of approach to steady state than the inverse of the relaxation time, 0·50, when the process is for practical purposes near steady-state, e.g. $H(t) \approx 0{\cdot}95$. Note that 1·309 falls between our 'bounds' 0·50 and 2·0.

The three-regime interpretation is significant because it leads to *simple exponential approximations in each regime.* In particular, it is natural to approximate the complementary c.d.f. $1-H_1(t)$ by $\exp(-5{\cdot}236t)$ in the first regime, by $0{\cdot}276\exp(-0{\cdot}764t)$ in the second regime and by $(8/\pi)^{\frac{1}{2}}t^{-\frac{3}{2}}\times \exp(-t/2)$ or $A\exp(-0{\cdot}500t)$ where $A \approx (8/\pi)^{\frac{1}{2}}t^{-\frac{3}{2}}$ in the third regime. In the second regime, which covers the region of greatest interest where $0{\cdot}85 \leqq H_1(t) \leqq 0{\cdot}99$, we are suggesting just using the second exponential component of (1.13). This is the proposed simple exponential approximation (1.3).

TABLE 4
The performance of the three simple exponential approximations for the complementary c.d.f. $1-H_1(t)$ in the three regimes

Time t	Exact $1-H_1(t)$	First regime $e^{-5\cdot 236t}$	Second regime $0\cdot 276e^{-0\cdot 764t}$	Third regime $(8/\pi)^{\frac{1}{2}}t^{-\frac{3}{2}}e^{-t/2}$
0	1·000	1·00		
0·05	0·690	0·77		
0·10	0·587	0·59		
0·15	0·517	0·46		
1·0	0·151		0·132	
2·0	0·057		0·060	
3·0	0·025		0·028	
4·0	0·0115		0·0130	
5·0	0·0056		0·0061	
6·0	0·0028		0·0028	
7·0	0·0015		0·0013	
9·0	$4\cdot 1\times 10^{-4}$		$2\cdot 9\times 10^{-4}$	$6\cdot 6\times 10^{-4}$
16·0	$6\cdot 2\times 10^{-6}$			$8\cdot 4\times 10^{-6}$
25·0	$4\cdot 1\times 10^{-8}$			$4\cdot 8\times 10^{-8}$
49·0	$1\cdot 0\times 10^{-15}$			$1\cdot 1\times 10^{-15}$

The performance of these simple approximations for $1-H_1(t)$ in the separate regimes is shown in Table 4. The excellent performance should be expected from Table 3 because the individual exponential components dominate in the different regimes. In the first regime, $\exp(-0\cdot 764t)\approx 1$, while in the second regime $\exp(-5\cdot 236t)\approx 0$. Of course, there are times between the first two regimes where the full H_2 approximation (1.13) would be preferred, but the main point is that a simple exponential is appropriate for the region of primary interest, the second regime. For numerical values, it is natural to use Theorem 1.1 or the Laplace transforms. For understanding, it is natural to use (1.3) and (1.13).

The simple exponential approximation (1.3) based on the second component of the three-moment H_2 approximation (1.13) can be justified for $1\leqq t\leqq 7$ another way. We can proceed empirically and do a log-linear regression from the exact values of $1-H_1(t)$ for $t=1, 2, \cdots, 7$ as given in Table 4. This regression reproduces the exact values to the given accuracy and yields the approximation $1-H_1(t)\approx 0\cdot 272\exp(-0\cdot 764t)$, providing additional justification for (1.3). This empirical procedure for developing simple exponential approximations for queues is used by Roth (1981), Odoni and Roth (1983), Lee (1985) and Lee and Roth (1986). Our analysis here and in Abate and Whitt (1986a,b), (1988) supports and complements their work.

Although we have considered only the special case in which $R(0)=0$, we believe that our results provide useful insight more generally. For other

TABLE 5
A comparison of the hyperexponential, exponential and asymptotic (Corollary 1.3.5) approximations with exact values of $1-H_2(t)$, the complementary second-moment c.d.f.

Time t	Exact by numerical inversion	3-moment fit hyperexponential in (5.11)	1-moment fit exponential e^{-t} in (1.12)	Asymptotic as $t\to\infty$ $8(2/\pi)^{\frac{1}{2}}t^{-\frac{3}{2}}e^{-t/2}$
0·01	0·982	0·980	0·990	
0·1	0·858	0·819	0·905	
0·5	0·539	0·542	0·607	14·05
1·0	0·333	0·324	0·368	3·88
1·5	0·216	0·209	0·223	1·64
2·0	0·144	0·141	0·135	0·83
3·0	0·068	0·069	0·050	0·27
4·0	0·033	0·035	0·018	0·068
7·0	0·0045	0·0047	0·0009	0·0104
9·0	0·0012	0·0012	0·0001	0·0027

reasonable initial conditions, e.g., $R(0)\leqq 3/4$, we would expect that the rate of approach to steady state of $ER(t)$ would be about $1/1{\cdot}309=0.76$ (second regime) instead of 0·50 (third regime) when $ER(t)/ER(\infty)\approx 0{\cdot}95$. In other words, the rate of approach for t reasonably large should be about 1.5 times the inverse of the relaxation time.

1.7. *Complementary-c.d.f. c.d.f.'s.* The focus of this paper is on the moments of RBM and the associated moment c.d.f.'s $H_k(t)$ in (1.4), but other characteristics of RBM can be treated in a similar way. In this subsection we briefly discuss the complementary c.d.f. $P(R(t)>y\mid R(0)=0)$. From (1.1), we obtain the following expression for the complementary-c.d.f. c.d.f., which we denote by $H_y(t)$:

$$H_y(t)\equiv\frac{P(R(t)>y\mid R(0)=0)}{P(R(\infty)>y)}=\exp(2y)\Phi\left(\frac{-y-t}{t^{\frac{1}{2}}}\right)+\Phi\left(\frac{-y+t}{t^{\frac{1}{2}}}\right),\qquad t\geqq 0. \tag{1.14}$$

Since $(R(t)\mid R(0)=0)$ is stochastically increasing in t, $H_y(t)$ is a legitimate c.d.f.. As a byproduct of the proof of Theorem 1.3, we obtain the following representation.

Theorem 1.8. The complementary-c.d.f. c.d.f. $H_y(t)$ in (1.14) coincides with the first-passage-time c.d.f. $P(T_{y0}\leqq t)$ in (1.6).

Corollary 1.8.1. For all positive x and y, $H_{x+y}(t)$ is the convolution of $H_x(t)$ and $H_y(t)$.

Corollary 1.8.2. The mean and variance of $H_y(t)$ are both y.

Since the squared coefficient of variation is $c_y^2 = y^{-1}$, clearly an H_2 approximation is not appropriate for $y > 1$.

Since $P(R(t) > y \mid R(0) = 0) \leqq \exp(-2y)$ the complementary c.d.f. is necessarily small for large y, so as y increases we are considering rare events. The complementary-c.d.f. c.d.f. $H_y(t)$ in (1.14) is interesting because we normalize by dividing by the limit. The behavior of $H_y(t)$ is interesting and easy to describe. Since the first-passage-time distribution (1.5) is infinitely divisible, $ET_{y0} = y$ and $\operatorname{Var} T_{y0} = y$, we can invoke a central limit theorem to obtain the following result.

Corollary 1.8.3. $\lim_{y\to\infty} H_y(y^{\frac{1}{2}}(t + y)) = \Phi(t)$.

The practical meaning of Corollary 3 is that for large y the c.d.f. $H_y(t)$ has almost all its increase in the interval $(y - 2y^{\frac{1}{2}},\ y + 2y^{\frac{1}{2}})$. As y gets large, first $P(R(\infty) > y) = \exp(-2y)$ gets small, but even relative to $P(R(\infty) > y)$ the complementary c.d.f. $P(R(t) > y \mid R(0) = 0)$ is small for $t < y - 2y^{\frac{1}{2}}$.

As in Section 1.3, we can also use Laplace transforms. In particular, we can simply integrate the time-transformed density $\hat{f}(s, x) = s^{-1}r_1(s)\exp(-r_1(s)x)$ in (1.8) to obtain the time-transformed complementary c.d.f. $\hat{H}_y(s) = \int_0^\infty \exp(-st)P(R(t) > y \mid R(0) = 0)\, dt = s^{-1}\exp(-r_1(s)y)$, using the separability as before.

1.8. *Density c.d.f.'s.* As in Section 3, let $f(y, t)$ be the density of RBM at time t, i.e., the density of the c.d.f. in (1.1) when $R(0) = 0$. Paralleling (1.4) and (1.14), let $h_y(t)$ be the associated *normalized density function,* defined by

$$h_y(t) = \frac{f(y, t)}{f(y, \infty)} = 2^{-1}e^{2y}f(y, t),\ t \geqq 0. \tag{1.15}$$

Let $\hat{f}(y, s)$ and $\hat{h}_y(s)$ be the Laplace transforms with respect to time. From (1.8), we obtain

$$\begin{aligned} \hat{f}(y, s) &= \frac{r_1}{s}\exp(-r_1 y) = (2\exp(-2y))\frac{r_1}{2s}\exp(-r_2 y) \\ &= (2\exp(-2y))\frac{r_1}{2s}\hat{f}(s; y, 0) \end{aligned} \tag{1.16}$$

so that

$$\hat{h}_y(s) = \frac{r_1}{2s}\hat{f}(s; y, 0). \tag{1.17}$$

Since $r_1/2s$ is the Laplace transform of $\Phi(t^{\frac{1}{2}}) + (2\pi t^3)^{-\frac{1}{2}}\exp(-t/2)$,

$$h_y(t) = t^{-\frac{1}{2}}\phi\left(\frac{t - y}{\sqrt{t}}\right) + \Phi\left(\frac{t - y}{\sqrt{t}}\right). \tag{1.18}$$

(Alternatively, we can obtain (1.18) directly from (1.1) by differentiating with respect to y in the case $x = 0$.)

From Theorem 1.8 we can easily deduce the following.

Theorem 1.9. The normalized density function in (1.15) satisfies

$$h_y(t) = H_y(t) - 2^{-1}\frac{dH_y(t)}{dy} \tag{1.19}$$

and

$$\frac{dh_y(t)}{dt} = f(t; y, 0)\left(\frac{y-1}{2y} + \frac{y}{2t}\right), \qquad t \geqq 0, \tag{1.20}$$

so that $h_y(t)$ is increasing in t (a legitimate c.d.f.) if and only if $y \geqq 1$.

Typical values of $h_y(t)$ as a function of y and t are given in Table 6.

TABLE 6

The ratio in (1.18) of the density $f(y, t)$ starting at the origin to the steady-state limit $f(y, \infty) = 2e^{-2y}$ as a function of the state y and the time t

Time	State y							
t	0·20	0·50	0.90	1·00	1·50	2·00	3·00	12·00
0·05	1·68	0·26	0·0014	0·0002	0·0000	0·0000		
0·10	1·58	0·67	0·057	0·024	0·000	0·0000		
0·20	1·39	0·96	0·32	0·22	0·015	0·0003	—	
0·30	1·29	1·039	0·54	0·42	0·080	0·007	0·0000	
0·50	1·18	1·064	0·77	0·68	0·29	0·076	0·0013	
0·75	1·11	1·055	0·89	0·83	0·51	0·24	0·020	
1·00	1·078	1·044	0·937	0·90	0·66	0·40	0·077	
1·50	1·041	1·026	0·977	0·96	0·83	0·64	0·26	
2·00	1·024	1·016	0·990	0·980	0·90	0·78	0·46	
2·50	1·015	1·010	0·995	0·989	0·94	0·86	0·62	
3·00	1·0093	1·0068	0·998	0·994	0·965	0·91	0·73	
3·50	1·0061	1·0045	0·9989	0·9965	0·978	0·94	0·81	
4·00	1·0041	1·0031	0·9994	0·9980	0·986	0·962	0·87	0·0000
5·00	1·0019	1·0014	0·9999	0·9992	0·9936	0·983	0·93	0·0022
6·00	1·0009	1·0007	1·0000−	0·9997	0·9970	0·9917	0·967	0·015
7·00	1·0005	1·0004	1·0000+	0·9999	0·9986	0·9959	0·982	0·055
8·00	1·0002	1·0002	1·0000+	0·9999	0·9993	0·9979	0·991	0·13
9·00	1·0001		1·0000+	1·0000−	0·9996	0·9989	0·995	0·24
10·00			1·0000+	—	0·9998	0·9994	0·9974	0·37
12·00			1·0000+	—	0·9999	0·9998	0·9993	0·62
14·00			1·0000+	—	1·0000−	1·0000−	0·9998	0·80
16·00			1·0000+	—	—	—	0·9999	0·90
20·00			1·0000+	—			1·0000−	0·981

1.9. *The rest of this paper.* The rest of this paper is organized as follows. In Section 2 we construct canonical RBM and indicate how to obtain corresponding results for general μ and σ^2. In Section 3 we discuss the moment c.d.f.'s in (1.4) and apply probabilistic argument to establish Theorems 1.2-1.4. In Section 4 we analyze the moment c.d.f.'s using Laplace transforms, expanding on Section 1.3 above. We also prove Corollary 1(a) to Theorem 1.1 and Theorems 1.5-1.7 there. In Section 5 we indicate how to construct the H_2-c.d.f. approximations for the moment c.d.f.'s and expand on Section 1.4-1.7 above. We also construct and evaluate the H_2 approximation for the second-moment c.d.f. $H_2(t)$ there. The H_2 approximation for $H_2(t)$ performs even better than the H_2 approximation for $H_1(t)$. Finally, we prove Theorem 1.1 in Section 6. Other results for non-zero initial conditions appear in Part II.

1.10. *Related work.* Related results for queues are described in Abate and Whitt (1987a,b,c), (1988)). In (1987b), (1988), we show that much of the nice structure for RBM also holds for the queue-length process in the $M/M/1$ queue. Contrary to the impression given by much of the literature, it *is* possible to give a relatively nice description of the transient behavior of the $M/M/1$ queue, again under the special initial condition considered here. As an analog of Theorem 1.3, we show that the normalized kth factorial moment function coincides with a negative binomial mixture of first-passage-time distributions (convolutions of the busy-period distribution). As an analogue to (1.8), we show that the time-transform for the probability mass function as a function of time has a simple geometric form. As an analog to Part II, we also obtain results for the $M/M/1$ queue with non-zero initial condition.

In (1987c), we apply heavy-traffic limit theorems establishing convergence to RBM to approximate the transient behavior of the queue-length process in the $GI/G/1$ queue. However, we do *not* apply RBM directly. Instead, we apply the heavy-traffic limit theorems *twice,* and use the exact results for the $M/M/1$ queue. This permits us to consider more explicitly the role of the traffic intensity in the transient behavior of the $GI/G/1$ queue. (This provides a significant improvement when the traffic intensity is not near 1, but not otherwise.) The interarrival-time and service-time distributions enter in the approximations only through their first two moments. The second moments alter the $M/M/1$ description only via a change in the time scale, just as the parameters μ and σ^2 do here, see Corollary 2.3.2. Our theoretically-based approximations for the transient behavior of the $GI/G/1$ queue are consistent with the empirical results of Roth (1981), Odoni and Roth (1983), Lee (1985) and Lee and Roth (1986). The H_2 approximation for the moment c.d.f. $H_1(t)$ again seems suitable for practical engineering purposes. As in (1.3), a simple exponential approximation is obtained by using the second component of this

H_2 c.d.f. The present paper was motivated by our desire to analyze the queues; the results here help us meet this objective.

2. Canonical regulated Brownian motion

In this section we mathematically define and characterize regulated Brownian motion (RBM) in terms of standard Brownian motion (BM). We construct canonical versions of both BM and RBM, having unit drift and unit variance, from which all other versions can be constructed by an appropriate choice of measuring units for time and space. (Canonical RBM coincides with the 'dimensionless diffusion' suggested by Newell; see p. 614 of Gaver (1968), but our construction is different.) Having a canonical version is useful because all difficult calculations only need to be done for this one special case.

Let $\{B(t):t \geqq 0\}$ be standard BM without drift, where $B(0)=0$, $B(t) \stackrel{d}{=} N(0, t)$, $\stackrel{d}{=}$ means equal in distribution and $N(\mu, \sigma^2)$ represents a random variable with a normal distribution having mean μ and variance σ^2; see Chapter 7 of Karlin and Taylor (1975). Let $B(t; \mu, \sigma^2, X)$ represent BM with drift μ and variance (diffusion) coefficient σ^2, starting at the random initial position X, and define it by

$$B(t; \mu, \sigma^2, X) = \sigma B(t) + \mu t + X, \qquad t \geqq 0. \tag{2.1}$$

In (2.1) we assume that X is independent of $\{B(t): t \geqq 0\}$.

Since

$$\{B(at): t \geqq 0\} \stackrel{d}{=} \{a^{\frac{1}{2}}B(t): t \geqq 0\} \tag{2.2}$$

for any $a > 0$, where equality in distribution applies to the entire stochastic process, see p. 351 of Karlin and Taylor (1975), we can change the scale of time and space to transform to and from canonical versions with σ^2 replaced by 1 and μ replaced by -1, 0 or $+1$ if $\mu < 0$, $\mu = 0$ or $\mu > 1$, respectively. For simplicity, we only consider the principal case of interest $\mu < 0$, but similar results hold in the other cases.

Proposition 2.1. If $\mu < 0$, then $\{aB(bt; \mu, \sigma^2, X): t \geqq 0\} \stackrel{d}{=} \{B(t; -1, 1, aX): t \geqq 0\}$ and $\{B(t; \mu, \sigma^2, X): t \geqq 0\} \stackrel{d}{=} \{a^{-1}B(b^{-1}t; -1, 1, aX): t \geqq 0\}$ for

$$a = |\mu|/\sigma^2, \qquad b = \sigma^2/\mu^2, \qquad \mu = -1/ab \quad \text{and} \quad \sigma^2 = 1/a^2 b. \tag{2.3}$$

Proof. By (2.1) and (2.2),

$$aB(bt; \mu, \sigma^2, a^{-1}X) = a\sigma B(bt) + a\mu bt + aa^{-1}X$$

so that

$$\{aB(bt;\mu,\sigma^2,a^{-1}X):t\geqq 0\}\stackrel{d}{=}\{a\sigma b^{\frac{1}{2}}B(t)+a\mu bt+X:t\geqq 0\}$$
$$\stackrel{d}{=}\{B(t;a\mu b,a^2\sigma^2 b,X):t\geqq 0\}.$$

We obtain $a\mu b=-1$ and $a^2\sigma^2 b=1$ with $a=a^2b/ab=|\mu|/\sigma^2$ and $b=\sigma^2/\mu^2$.

RBM is the modification of BM corresponding to the imposition of an impenetrable 'reflecting' barrier at the origin. Let r be the reflecting barrier function mapping the space of continuous real-valued functions on $[0,\infty)$ into itself, defined by

$$r(y)\equiv r(y)(t)=\max\left\{y(t),\,y(t)-\inf_{0\leqq s\leqq t}y(s)\right\},\qquad t\geqq 0. \tag{2.4}$$

The following elementary proposition identifies important properties of r.

Proposition 2.2. For any $a>0$ and continuous function $y\equiv\{y(t):t\geqq 0\}$, $r(y)\equiv\{r(y)(t):\ t\geqq 0\}$ is a continuous function satisfying $r(ay)=ar(y)$ and $r(\{y(at):\ t\geqq 0\})=\{r(y)(at):t\geqq 0\}$.

Let RBM $R(t;\mu,\sigma^2,X)$ be defined in terms of $B(t;\mu,\sigma^2,X)$ by

$$\begin{aligned}\{R(t;\mu,\sigma^2,X):t\geqq 0\}&=r(\{B(t;\mu,\sigma^2,X):t\geqq 0\})\\&=\left\{\max\left\{B(t;\mu,\sigma^2,X),\,B(t;\mu,\sigma^2,X)-\inf_{0\leqq s\leqq t}B(s;\mu,\sigma^2,X)\right\}:t\geqq 0\right\}.\end{aligned}\tag{2.5}$$

It is easy to see that $R(t;\mu,\sigma^2,X)$ defined in (2.5) is a diffusion process (continuous-time stochastic process with continuous sample paths and the strong Markov property); see Harrison (1985).

As an immediate consequence of Propositions 2.1 and 2.2, we have the following result.

Proposition 2.3. If $\mu<0$, then

$$\{aR(bt;\mu,\sigma^2,X):t\geqq 0\}\stackrel{d}{=}\{R(t;-1,1,aX):t\geqq 0\}$$

and

$$\{R(t;\mu,\sigma^2,X):\ t\geqq 0\}\stackrel{d}{=}\{a^{-1}R(b^{-1}t;-1,1,aX):\ t\geqq 0\}$$

for a and b in (2.3).

To obtain moments for the general (μ,σ^2) case from canonical RBM, we can apply the following.

Corollary 2.3.1. If $\mu<0$, then

$$E[R(t;\mu,\sigma^2,X)^k]=a^{-k}E[R(b^{-1}t;-1,1,aX)^k]$$

for a and b in (2.3).

Let $H_k(t; \mu, \sigma^2)$ represent the kth-moment c.d.f. in (1.4) for (μ, σ^2)-RBM; i.e.; let

$$H_k(t; \mu, \sigma^2) = \frac{E[R(t; \mu, \sigma^2, 0)^k]}{E[R(\infty; \mu, \sigma^2, 0)^k]}, \qquad t \geqq 0. \tag{2.6}$$

Corollary 2.3.2. If $\mu < 0$, then $H_k(t; \mu, \sigma^2) = H_k(b^{-1}t; -1, 1) \equiv H_k(b^{-1}t)$ for $b^{-1} = \mu^2/\sigma^2$.

From Corollary 2.3.2, we see that the parameters μ and σ^2 alter the moment c.d.f.'s by a simple transformation of the time scale, depending on the single parameter b. It is significant that the transient behavior of (μ, σ^2)-RBM can be analyzed by considering three separate phenomena: (1) the approach to steady state of canonical RBM, (2) the impact of μ and σ^2 on the steady-state distribution, which is determined by $a^{-1} = \sigma^2/|\mu|$ and (3) the impact of μ and σ^2 on the time scale, which is determined by $b^{-1} = \mu^2/\sigma^2$.

Propositions 2.1 and 2.3 tell us that we can always choose the measuring units for space and time appropriately so that, when $\mu < 0$, we only need consider $\{B(t; \mu, \sigma^2, X): t \geqq 0\}$ and $\{R(t; \mu, \sigma^2, X): t \geqq 0\}$ for the special cases of $\mu = -1$ and $\sigma^2 = 1$. This special case of RBM coincides with the 'dimensionless diffusion' defined on p. 615 of Gaver (1968) in the classical way via the forward differential (Fokker–Planck) equation. In the general case with $P(X = x) = 1$,

$$\frac{\partial F}{\partial t} = -\mu \frac{\partial F}{\partial y} + \frac{\sigma^2}{2} \frac{\partial^2 F}{\partial y^2} \tag{2.7}$$

where $F \equiv F(y; t, x; \mu, \sigma^2)$ is the c.d.f. (cumulative distribution function)

$$F(y; t, X; \mu, \sigma^2) = P(R(t; -\mu, \sigma^2, X) \leqq y), \qquad y \geqq 0. \tag{2.8}$$

The initial condition for (2.7) is specified by

$$F(y; 0, x; \mu, \sigma^2) = \begin{cases} 1, & y < x \\ 0, & y \geqq x \end{cases} \tag{2.9}$$

and the boundary condition is

$$F(y; t, x; \mu, \sigma^2) = 0, x > 0, t > 0 \quad \text{and} \quad y \leqq 0. \tag{2.10}$$

As Gaver indicates, we can also obtain the canonical dimensionless form, say $F(y; t, x) \equiv F(y; t, x; -1, 1)$, directly from (2.7) by performing a change of variables for y and t corresponding to (2.3). The marginal distributions of regulated Brownian motion are described by the c.d.f. F in (2.8). The solution to (2.7)-(2.10) is given in (1.1). Other derivations are given in Harrison (1985).

3. Supporting probabilistic arguments

Let $R(t) \equiv R(t; -1, 1, 0)$ represent canonical RBM starting at the origin. We are interested in $m_k(t) = E[R(t)^k]$, especially for $k = 1$ and 2. We look at the closely related moment c.d.f.'s $H_k(t) = m_k(t)/m_k(\infty) = (2^k/k!)m_k(t)$, $t \geqq 0$, defined in (1.4), which we obtain from the fact that the kth moment of the exponential distribution having mean λ^{-1} is $(k!)\lambda^{-k}$.

Proof of Theorem 1.3. Even with general initial conditions, we can easily find an integral expression for each moment of $R(t)$. We can simply apply (1.1) with

$$E(R(t)^k \mid R(0) = x) = \int_0^\infty ky^{k-1}P(R(t) > y \mid R(0) = x)\, dy; \tag{3.1}$$

p. 150 of Feller (1971). However, we can obtain significant simplification by assuming $R(0) = 0$. Let $M(t)$ be the *maximum process* associated with canonical unregulated Brownian motion with negative drift, starting at the origin, i.e.,

$$M(t) = \max\{B(s, -1, 1, 0) : 0 \leqq s \leqq t\}, \qquad t \geqq 0. \tag{3.2}$$

A key property is that $R(t) \equiv (R(t) \mid R(0) = 0)$ is equal in distribution to $M(t)$ for each t (but not as stochastic processes); p. 14 of Harrison (1985). Moreover, we have the familiar inverse property connecting the maximum process to first-passage times, namely,

$$P(M(t) \geqq x) = P(T_{0x} \leqq t) \tag{3.3}$$

for all positive t and x. These properties together with Theorem 1.4 are all we need:

$$\begin{aligned} &E(R(t)^k \mid R(0) = 0) \\ &= \int_0^\infty kx^{k-1}P(R(t) > x \mid R(0) = 0)\, dx = \int_0^\infty kx^{k-1}P(M(t) > x)\, dx \\ &= \int_0^\infty kx^{k-1}P(T_{0x} \leqq t)\, dx = \int_0^\infty kx^{k-1}\exp(-2x)P(T_{x0} \leqq t)\, dx. \end{aligned} \tag{3.4}$$

After normalizing, we get

$$\begin{aligned} \frac{E(R(t)^k \mid R(0) = 0)}{E(R(\infty)^k)} &= \frac{2^k}{(k-1)!}\int_0^\infty x^{k-1}\exp(-2x)P(T_{x0} \leqq t)\, dx \\ &= \int_0^\infty g_k(x)F(t; x, 0)\, dx, \end{aligned}$$

as claimed in Theorem 1.3.

Proof of Corollary 1.3.2. The density $f(t; x+y, 0)$ of $T_{x+y,0}$ is the convolution of $f(t; x, 0)$ and $f(t; y, 0)$ because $T_{x,0} \stackrel{d}{=} T_{x+y,y}$ for all $y > 0$ and $T_{x+y,0} = T_{x+y,x} + T_{x,0}$ where $T_{x+y,x}$ and $T_{x,0}$ are independent. Similarly, $g_k(x)$ is the k-fold convolution of $g_1(x)$. Hence, the convolution of $h_{k-1}(t)$ and $h_1(t)$ is

$$\begin{aligned}\int_0^t h_{k-1}(s)h_1(t-s)\,ds &= \int_0^t ds\int_0^\infty g_{k-1}(x)f(s; x, 0)\,dx\int_0^\infty g_1(y)f(t-s; y, 0)\,dy\\ &= \int_0^\infty g_{k-1}(x)\,dx\int_0^\infty g_1(y)\,dy\int_0^t f(s; x, 0)f(t-s; y, 0)\,ds\\ &= \int_0^\infty g_{k-1}(x)\int_0^\infty g_1(y)f(t; x+y, 0)\,dy\,dx\\ &= \int_0^\infty dz\int_0^z f(t; z, 0)g_{k-1}(x)g_1(z-x)\,dx\\ &= \int_0^\infty g_k(z)f(t; z, 0)\,dz = h_k(t).\end{aligned}$$

Proof of Corollary 1.3.5. Note that

$$\begin{aligned}(2\pi)^{\frac{1}{2}}t^{\frac{3}{2}}\exp(t/2)h_k(t) &= \int_0^\infty \frac{2^k x^{k-1}\exp(-2x)}{(k-1)!}x\exp(-x^2/2t)e^x\,dx\\ &= k2^k\int_0^\infty \frac{x^k e^{-x}}{k!}\exp(-x^2/2t)\,dx \to k2^k \quad \text{as} \quad t\to\infty.\end{aligned}$$

Given that $h_k(t) \sim A_k t^{-\frac{3}{2}}\exp(-t/2)$, it suffices to consider the limit of $t^{\frac{3}{2}}e^{t/2}\int_t^\infty A_k u^{-\frac{3}{2}}\exp(-u/2)\,du$ as $t\to\infty$. (If $h(t)\sim g(t)$, then $\int_t^\infty h(u)\,du = \int_t^\infty g(u)[h(u)/g(u)]\,du$ where $\sup\{|[h(u)/g(u)]-1| : u \geqq t\} \to 0$ as $t\to\infty$.) After making the change of variables to $y = u - t$, we obtain

$$t^{\frac{3}{2}}e^{t/2}\int_t^\infty A_k u^{-\frac{3}{2}}\exp(-u/2)\,du = \int_0^\infty A_k\left(\frac{t}{y+t}\right)^{\frac{3}{2}}\exp(-y/2)\,dy \to 2A_k \quad \text{as} \quad t\to\infty.$$

Proofs of Theorem 1.2. We give four proofs of Theorem 1.2. First, it is an immediate corollary to Theorem 1.3. Second, it is a trivial consequence of the third term of (3.4): the maximum process obviously has non-decreasing sample paths, so that $P(M(t) > y)$ is increasing in t for all y. The monotonicity of the moments then follow directly from (3.4) or the fact that the entire distribution $P(R(t) \leqq y \mid R(0) = 0)$ is stochastically increasing in t; i.e., $P(R(t) > y \mid R(0) = 0)$ is increasing in t for each $y > 0$. Third, this stochastic order also follows from corresponding results for birth-and-death processes; see van Doorn (1980) and references cited there. The stochastic order carries over when sequences of the birth-and-death processes converge to regulated Brownian motion; see Iglehart and Whitt (1970b) or Stone (1963). The appropriate birth-and-death process to consider is of course the number of customers in an $M/M/1$ queue. Finally, the fourth proof is the stochastic monotonicity argument in Section 1.2. See Remark 4.1 for a fifth proof in the case $k=1$.

Proofs of Theorem 1.4. First, Theorem 1.4 can easily be verified by direct calculation using the known distributions for T_{0x} and T_{x0}; see (11) on p. 14 and (2) on p. 46 of Harrison (1985). (Use the symmetry: $1-\Phi(x)=\Phi(-x)$.) Second, a more revealing proof is provided by exploiting reversibility. The key property is easily expressed in terms of discrete-time Markov chains with integer state space, stationary distribution π and transition matrix P. Reversibility holds if and only if the detailed balance conditions hold, i.e., $\pi_i P_{ij}=\pi_j P_{ji}$; p. 5 of Kelly (1979). As an easy corollary (closely related to the Kolmogorov criterion, p. 21 of Kelly (1979)), we have for any path $[i_1,\dots,i_n]$, i.e., sequence of n successive states, the stationary distribution is

$$\begin{aligned} P([i_1,\cdots,i_n]) &= \pi_{i_1}P_{i_1i_2}P_{i_2i_3}\cdots P_{i_{n-1}i_n} \\ &= \pi_{i_n}P_{i_ni_{n-1}}\cdots P_{i_2i_1}=P([i_n,\cdots,i_1]). \end{aligned} \tag{3.5}$$

The specific process we want to consider is a simple random walk on the non-negative integers with a reflecting barrier at 0. At each transition this process moves up 1 with probability p and down 1 with probability $1-p$. The stationary distribution π is geometric. Let T_{ab}^c be the first-passage time to b from a given that c is not visited first after leaving a. The reversibility implies that

$$\pi_0 P(T_{0n}^0=k)=\pi_n P(T_{n0}^n=k) \tag{3.6}$$

for all k and n. (See Sumita (1984) and especially Doney (1984) for related results.) Now consider the same simple random walk on the integers *without the barrier*. We represent the first-passage-time events $\{T_{0n}\leqq k\}$ and $\{T_{n0}\leqq k\}$ for this unrestricted random walk as a union of disjoint events. Let L_{jk} be the time of the *last* visit to j in $\{0,1,\cdots,k\}$ starting at j. We can write

$$\{T_{0n}\leqq k\}=\bigcup_{j=0}^{k}\bigcup_{m=1}^{k-j}(\{L_{0k}=j\}\cap\{T_{0n}^0=m\})$$

$$\{T_{n0}\leqq k\}=\bigcup_{j=0}^{k}\bigcup_{m=1}^{k-j}(\{L_{nk}=j\}\cap\{T_{n0}^n=m\})$$

so that

$$P(T_{0n}\leqq k)=\sum_{j=0}^{k}\sum_{m=1}^{k-j}P(L_{0k}=j)P(T_{0n}^0=m)$$

$$P(T_{n0}\leqq k)=\sum_{j=0}^{k}\sum_{m=1}^{k-j}P(L_{nk}=j)P(T_{n0}^n=m).$$

Obviously, $P(L_{0k}=j)=P(L_{nk}=j)$ for all j, k and n because of the homogeneity. Hence, we can apply (3.6) to obtain

$$\pi_0 P(T_{0n} \leqq k) = \pi_n P(T_{n0} \leqq k) \tag{3.7}$$

for each k and n. Formulas (3.6) and (3.7) extend easily to the associated continuous-time Markov processes by applying (3.6) and (3.7) to the jump chain. Since the reverse path visits each state the same number of times as the forward path, the time required to traverse the paths in each direction is identical. Hence,

$$\pi_0 P(T_{0n} \leqq t) = \pi_n P(T_{n0} \leqq t) \tag{3.8}$$

holds for the homogeneous birth-and-death process on the integers. Theorem 1.4 is obtained as the limit, after appropriate normalization, when a sequence of birth-and-death processes are considered that converge to Brownian motion; Stone (1963). We use the fact that the geometric stationary probability mass function π_n for the $M/M/1$ queue converges to the exponential stationary density of Brownian motion (a local limit theorem, in this case elementary).

Remark 3.1. For related special properties of first-passage times for Brownian motion and related processes, see p. 66 of Cox and Miller (1965), Doney (1984), Phatarfod et al. (1971), Sumita (1984) and Takacs (1967). Reflection and reversibility are obviously powerful tools.

4. Supporting Laplace-transform arguments

In this section we return to the Laplace transforms introduced in Section 1.3. We treat the first two moment c.d.f.'s this way and then relate the other moment c.d.f.'s to it.

4.1. *The first-moment c.d.f.* We describe the first-moment c.d.f. $H_1(t)$ in (1.4) in more detail and prove Corollary 1.1.1(a). By (2.16) and (2.26) of Gaver (1968), the Laplace transform of the time-dependent mean $m_1(t)$ is

$$\int_0^\infty e^{-st} m_1(t)\, dt = (s[1+(1+2s)^{\frac{1}{2}}])^{-1}. \tag{4.1}$$

Let $\gamma(t)$ be the gamma density with mean 1 and shape parameter $\frac{1}{2}$, i.e.,

$$\gamma(t) = (2\pi t)^{-\frac{1}{2}} \exp(-t/2), \qquad t \geqq 0. \tag{4.2}$$

Let $\gamma_e(t)$ be the stationary-excess or equilibrium residual-life density associated with $\gamma(t)$; i.e., $\gamma(t)=\int_t^\infty \gamma(u)\, du$; p. 28 of Cox (1962).

Theorem 4.1. The first-moment c.d.f. of canonical RBM is

$$H_1(t) = 1 - 2(1+t)[1-\Phi(t^{\frac{1}{2}})] + 2t^{\frac{1}{2}}\phi(t^{\frac{1}{2}}), \tag{4.3}$$

which has density

$$h_1(t) = 2t^{-\frac{1}{2}}\phi(t^{\frac{1}{2}}) - 2[1 - \Phi(t^{\frac{1}{2}})] = 2\gamma(t) - \gamma_e(t),\ t \geqq 0, \tag{4.4}$$

Laplace–Stieltjes transform

$$\int_0^\infty e^{-st}\, dH_1(t) = \int_0^\infty e^{-st} h_1(t)\, dt = \hat{h}_1(s) = 2/[1 + (1 + 2s)^{\frac{1}{2}}], \tag{4.5}$$

and nth moment

$$m_n(H_1) = \int_0^\infty t^n\, dH_1(t) = (n+1)^{-1} m_{2n} \tag{4.6}$$

where $m_{2n} = (2n-1)(2n-3)\cdots(3)(1) = 2^n \pi^{-\frac{1}{2}}\Gamma((2n+1)/2)$ is the $(2n)$th moment of $N(0, 1)$. ($m_n(H_1) = 1/2$, 1 and 15/4 for $n = 1$, 2 and 3.)

Proof. We directly invert the transform of $H_1(t)$ using (4.1). We apply standard arguments for the inversion, referring to Abramowitz and Stegun (1972). By (29.2.6) there, the factor s in the denominator allows us to identify (2.19) as the transform of the alleged density $h_1(t)$. By (29.2.12), the transform of $e^{t/2}h_1(t)$ is $2[1 + (2s)^{\frac{1}{2}}]^{-1} = 2(1 - (2s)^{\frac{1}{2}})/(1 - 2s) = 2(2s)^{\frac{1}{2}}/(2s - 1) - 2/(2s - 1)$. By (29.3.8) and (29.3.38),

$$f(t) = (2/\pi t)^{\frac{1}{2}} \exp(-t/2) + \operatorname{erf}((t/2)^{\frac{1}{2}}) - 1$$

where $\operatorname{erf}(z)$ is the error function defined in (7.1.1) there, from which we obtain (4.4).

By integrating (4.4), we see that the c.d.f. $H_1(t)$ itself can be expressed as

$$\begin{aligned} H_1(t) &= 4\Phi(t^{\frac{1}{2}}) - 2 - 2t[1 - \Phi(t^{\frac{1}{2}})] - 2\int_0^{\sqrt{t}} y^2\phi(y)\, dy \\ &= (4 + 2t)\Phi(t^{\frac{1}{2}}) - 2(1 + t) - 2(1/\pi^{\frac{1}{2}})\int_0^{t/2} u^{\frac{1}{2}}e^{-u}\, du \\ &= 2 + (4 + 2t)[\Phi(t^{\frac{1}{2}}) - 1] - 2(1/\pi^{\frac{1}{2}})\gamma(\tfrac{3}{2}, t/2) \end{aligned} \tag{4.7}$$

where $\gamma(a, t)$ is the incomplete gamma function in Section 6.5.2 of Abramowitz and Stegun (1972). (Formula (4.7) coincides with (2.157) in Kleinrock (1976).) Obtain (4.3) from (4.7) via

$$1 - 2\pi^{-\frac{1}{2}}\gamma(\tfrac{3}{2}, t/2) = 2[1 - \Phi(t^{\frac{1}{2}})] + 2t^{\frac{1}{2}}\phi(t^{\frac{1}{2}}), \tag{4.8}$$

invoking (6.5.22) and (26.2.30) of Abramowitz and Stegun (1972).

To obtain (4.6), make the change of variables $x = t^{\frac{1}{2}}$:

$$\begin{aligned}
\int_0^\infty t^n h(t)\,dt &= 4\int_0^\infty x^{2n}(\phi(x) - x[1 - \Phi(x)])\,dx \\
&= 2m_{2n} - 4\int_0^\infty x^{2n+1}(1 - \Phi(x))\,dx \\
&= 2m_{2n} = (4/(2n+2))\int_0^\infty x^{2n+2}\,d\Phi(x) \\
&= 2m_{2n} - (n+1)^{-1}m_{2n+2} = (n+1)^{-1}m_{2n}.
\end{aligned}$$

Remarks. 4.1. By Lemma 2, p. 175, of Feller (1968), the function $h_1(t^2)/2$ is non-negative, so that we have a fifth proof of Theorem 1.2 for $k = 1$, i.e., that $H_1(t)$ is a legitimate c.d.f.

4.2. Corollary 1.1.1(a) is equivalent to (4.3) because $ER(\infty) = \frac{1}{2}$.

4.3. The analysis above can also be done with the complementary c.d.f. $H_1^c(t) = 1 - H_1(t)$. If m_1 is the mean of $H_1(t)$, then $m_1^{-1}(1 - H_1^c(t))$ is the density of the associated stationary-excess distribution, say $H_1^*(t)$. The kth moments m_k of these two distributions are thus related by $m_k(H_1^*) = m_{k+1}(H_1)/m_1(H_1)(k+1)$; p. 64 of Cox (1962). From (4.1) or (1.8) it follows that the Laplace transform of $H_1^c(t)$ is

$$(4.9)\qquad \hat{H}_1^c(s) = \int_0^\infty \exp(-st)H_1^c(t)\,dt = [1 + s + (1+2s)^{\frac{1}{2}}]^{-1} = 2r_1(s)^{-2},$$

which has expansion

$$(4.10)\qquad \hat{H}_1^c(s) = \frac{1}{2} - \frac{s}{2} + \frac{15}{24}s^2 + o(s^2) \quad \text{as} \quad s \to 0$$

so that, in agreement with (4.6), the first three moments of $H_1(t)$ are $m_1 = \frac{1}{2}$, $m_2 = 1$ and $m_3 = \frac{15}{4}$.

4.4. Corollary 1.1.2(a) for the case $R(0) = 0$ can also be obtained directly from (4.9) by applying Heaviside's theorem; p. 254 of Doetsch (1974).

4.5. The mean $m_1(t) = E(R(t) \mid R(0) = 0)$ is linked in another curious way to the gamma density $\gamma(t)$ in (4.2). Let $U(t)$ represent the renewal function, i.e., the expected number of renewals in the interval $(0, t)$, in an ordinary renewal process with renewal-interval density $\gamma(t)$. The Laplace transform of $U(t)$ is thus $\hat{\gamma}(s)/s(1 - \hat{\gamma}(s))$ where $\hat{\gamma}(s) = (1+2s)^{-\frac{1}{2}}$, the Laplace transform of $\gamma(t)$. Then the time-dependent excess life $U(t) - t$ has Laplace transform

$$\begin{aligned}
\hat{U}(s) - s^{-2} &= \frac{1}{s[(1+2s)^{-\frac{1}{2}} - 1]} - \frac{1}{s^2} \\
&= \frac{1}{2}\left[\frac{1}{s} - \frac{1}{1 + s + (1+2s)^{\frac{1}{2}}}\right] \\
&= 1/s[1 + (1+2s)^{\frac{1}{2}}] = \hat{m}(s).
\end{aligned}$$

Hence, $U(t)-t$ based on $\gamma(t)$ coincides with $m_1(t)$. That $\gamma(t)$ should appear at all is evidently due to the fact that $\gamma(t)=(2\pi t)^{-\frac{1}{2}}\phi([x-t])/\sqrt{t})$, the density of unrestricted BM, for $x=0$. Thus, this excess-life relationship is evidently part of the connection between RBM and BM. More generally, let $T_f(s)\equiv\hat{f}(s)[1-\hat{f}(s)]^{-1}-(ms)^{-1}$ be the Laplace transform of the density of $U(t)-t$ when $\hat{f}(s)$ is the transform of the interrenewal density $f(t)$ having mean m. It is well known that $U(t)-t=0$, so that $T_f(s)=0$, when $f(t)$ is exponential. By above, $U(t)-t$ is $m_1(t)$ when $f(t)=\gamma(t)$. It is also easy to see that $h_1(t)$ is the unique fixed point of the operator T, i.e., $T_f(s)=\hat{f}(s)$ if and only if $f(t)=h_1(t/2m)$. Being a fixed point of the operator T turns out to be equivalent to the characterization in Corollary 1.5.2.

Proof of Theorem 1.7. Since $h_2(t)=2[1-H_1(t)]$, by Corollary 1.5.1, $h_2(t)$ is completely monotone if $h_1(t)$ is. We must show that the nth derivative $h_1^{(n)}(t)$ satisfies $(-1)^n h_1^{(n)}(t)\geqq 0$ for all n and t. By direct calculation, the first derivative of $h_1(t)$ is $h_1^{(1)}(t)=-(2\pi)^{-\frac{1}{2}}t^{-\frac{3}{2}}\exp(-t/2)$. By induction, for each n, the nth derivative is of the form

$$h_1^{(n)}(t)=(-1)^n\sum_{k=0}^{\infty}a_{nk}t^{-(2k+1)/2}\exp(-t/2)$$

where $\{a_{nk}:k\geqq 0\}$ is a sequence of non-negative constants for each n.

4.2. *The second-moment c.d.f.* We now investigate the c.d.f. $H_2(t)=m_2(t)/m_2(\infty)=2m_2(t)$ defined in (1.4). As indicated on pp. 611–612 of Gaver (1968), the transform of $H_2(t)$ can be obtained by differentiating the time-transformed density associated with (1.1). We omit details of the supporting calculations here.

Theorem 4.2. The Laplace transform of $m_2(t)$ is $2[s(1+[1+2s]^{\frac{1}{2}})^2]^{-1}$ and the Laplace transform of the complementary second-moment c.d.f. $H_2^c(t)=1-H_2(t)$ is

$$\hat{H}_2^c(s)\equiv\int_0^\infty e^{-st}H_2^c(s)\,ds=\frac{1}{s}\frac{([1+2s]^{\frac{1}{2}}-1+s)}{([1+2s]^{\frac{1}{2}}+1+s)}=s^{-3}[1+s-(1+2s)^{\frac{1}{2}}], \tag{4.11}$$

which has expansion

$$=1-\frac{5s}{4}+\frac{7}{4}s^2+o(s^2)\quad\text{as}\quad s\to 0, \tag{4.12}$$

so that the first three moments of $H_2(t)$ are $m_1=1$, $m_2=\frac{5}{2}$ and $m_3=21/2$.

4.3. *Higher-moment c.d.f.'s.* We now focus on $H_k(t)$ in (1.4) for $k\geqq 3$.

Proof of Theorem 1.5. Let $\psi(s)=(1+2s)^{\frac{1}{2}}$, $r_1(s)=\psi(s)+1$ and $r_2(s)=\psi(s)-1$, so that $r_1(s)r_2(s)=2s$. Then the transform of the density $h_1(t)$ is $\hat{h}_1(s)=2/r_1(s)=r_2(s)/s$. Consequently, by (1.10), the transform of $H_1^c(t)=1-H_1(t)$ is

$$\hat{H}_1^c(s)=\frac{1-\hat{h}_1(s)}{s}=\frac{1}{s}\left(1-\frac{2}{r_1(s)}\right)=\frac{r_2(s)}{sr_1(s)}=\frac{\hat{h}_1^2(s)}{2}=\frac{\hat{h}_2(s)}{2}.$$

By induction, $\hat{H}_{k+1}^c(s)=\hat{h}_1(s)\hat{H}_k^c(s)+\hat{H}_1^c(s)$ because

$$\begin{aligned}\hat{H}_{k+1}^c&=s^{-1}(1-\hat{h}_{k+1})=s^{-1}(1-\hat{h}_1^{k+1})\\&=s^{-1}([1-\hat{h}_1]-[\hat{h}_1-\hat{h}_1^{k+1}])=\hat{H}_1^c+\hat{h}_1\hat{H}_k^c,\end{aligned}$$

from which Theorem 1.5 follows.

Proof of Corollary 1.5.2. If a c.d.f. $H(t)$ with mean m has its convolution equal to its stationary-excess c.d.f. then its Laplace–Stieltjes transform $\hat{h}(s)$ must satisfy $[1-\hat{h}(s)]/ms=\hat{h}(s)^2$, which leads to a quadratic equation with solution

$$\hat{h}(s)=2/[1+\sqrt{1+4ms}]=\hat{h}_1(2ms).$$

Behavior of $H_k(t)$ and $h_k(t)$ as $t\to 0$. The asymptotic behavior of $H_k(t)$ as $t\to 0$ for each k can be determined by invoking Tauberian theorems, e.g., Theorem 1 on p. 443 of Feller (1971). Theorem 1.5 here then provides the means to extend these results to the densities $h_k(t)$. (Alternatively, Theorem 4, p. 446, of Feller can be used after showing that $h_k(t)$ is ultimately monotone.) The transform $\hat{h}_k(s)=(2/[1+\sqrt{1+2s}])^k$ is easily seen to be regularly varying at ∞ with exponent $k/2$, so that $\hat{H}_k(s)$ is regularly varying with exponent $(k-2)/2$ and $h_k(t)\sim A_k t^{(k-2)/2}$ as $t\to 0$. Hence, $h_k(t)$ is *increasing* at $t=0$ for all $k\geqq 3$, so that $h_k(t)$ cannot be monotone for $k\geqq 3$, let alone completely monotone as is the case for $k=1$ and 2.

4.4. *Transform inversion.* There are numerous algorithms for the numerical inversion of Laplace transforms. For example, there are three standard routines currently available from the ACM library of software algorithms: Algorithm 368 (Stehfest (1970)); Algorithm 486 (Veillon (1974)); Algorithm 619 (Piessens and Huysmans (1984)). The first is referred to as the Gaver–Stehfest method. Gaver (1966) first introduced the procedure. Stehfest (1970) discovered a weighting function which vastly improves the accuracy of the method. We employ this method for our results. It yields good (but not exceptional) accuracy on a small computer (e.g., a PC) with little programming effort.

The other two procedures are based on a Fourier series method first introduced by Dubner and Abate (1968). For other procedures, see the excellent review paper by Davies and Martin (1979) which evaluates 14 methods on 16 test problems.

5. Hyperexponential approximations

5.1. *Fitting H_2 distribution to three moments.* Given the first three moments m_1, m_2 and m_3, the H_2 parameters λ_1, λ_2, p_1 and p_2 for the density in (1.11) can be obtained by solving a quadratic equation, as described in Section 3.1 of Whitt (1982). We briefly describe a derivation originally shown to us by Shlomo Halfin. We start with three equations in three unknowns: $\lambda_1^{-k}p_1 + \lambda_2^{-k}(1-p_1) = m_k/k!$ for $k = 1, 2, 3$. We first reduce this system to two equations in two unknowns and then a single quadratic equation. Letting $y = \lambda_2/\lambda_1$, we express the first two equations as

$$(5.1)\qquad \begin{aligned}(y-1)p_1 &= \lambda_2 m_1 - 1\\ (y^2-1)p_1 &= \lambda_2^2 m_2/2 - 1\end{aligned}$$

so that we can divide the second equation in (5.1) by the first to obtain

$$(5.2)\qquad y = \frac{\lambda_2 m_1 - \lambda_2^2 m_2/2}{\lambda_2^2 m_2/2 - 1}.$$

We next let $w_1 = p_1/\lambda_1 m_1$, so that we can express the second and third equations as $\lambda_1^{-k}w_1 + \lambda_2^{-k}(1-w_1) = v_k/k!$ for $k = 1, 2$, $v_1 = m_2/2m_1$ and $v_2/2 = m_3/6m_1$, which is the same form as the first two original equations, with (w_1, v_1, v_2) playing the role of (p_1, m_1, m_2). Repeating the argument of (5.1) and (5.2), we obtain

$$(5.3)\qquad y = \frac{\lambda_2 v_1 - \lambda_2^2 v_2/2}{\lambda_2^2 v_2/2 - 1}.$$

Equations (5.2) and (5.3) are the resulting two equations in the two unknowns y and λ_2. These equations can immediately be combined to obtain a single quadratic equation for λ_2^{-1} in terms of the known moments, namely,

$$(m_1 - v_1)\lambda_2^{-2} - \frac{(m_2 - v_2)}{2}\lambda_2^{-1} + \frac{(v_1 m_2 - m_1 v_2)}{2} = 0$$

or

$$(5.4)\qquad (3m_1 y)\lambda_2^{-2} - (x + 1.5y^2 + 3m_1^2 y)\lambda_2^{-1} + m_1 x = 0$$

where $x = m_1 m_3 - 1.5m_2^2$ and $y = m_2 - 2m_1^2$. The solution to (5.4) is

$$(5.5)\qquad \lambda_i = \{x + 1.5y^2 + 3m_1^2 y \pm \sqrt{(x + 1.5y^2 + 3m_1^2 y)^2 - 12m_1^2 xy}\}/6m_1 y$$

for x and y above, $p_1 = (m_1 - \lambda_2^{-1})/(\lambda_1^{-1} - \lambda_2^{-1})$ and $p_2 = 1 - p_1$. In order to have an H_2 fit it is necessary and sufficient that either $x > 0$ and $y > 0$ or $x = 0$ and $y = 0$, with the latter yielding an exponential distribution; see Section 3.1 of Whitt (1982). For $k = 1$ and 2, $x > 0$ and $y > 0$, so that an H_2 fit is possible. For $k = 3$, $y = 0$ but $x > 0$, so that an H_2 fit is not possible. For $k \geqq 4$, $y < 0$, so that an H_2 fit is not possible. Other distributions could be fit to the first three moments to approximate $H_k(t)$ for $k \geqq 3$, but we do not pursue this because only $k = 1$ and 2 seem to be of significant practical interest.

The nature of the resulting H_2 distribution is perhaps better understood using a different parameter triple, namely, (m_1, c^2, r) where m_1 is the mean, c^2 is the squared coefficient of variation ($c^2 = (m_2 - m_1^2)/m_1^2$), and r is the proportion of the total mean in the component with the smaller mean, i.e.,

$$r = p_1\lambda_1^{-1}/(p_1\lambda_1^{-1} + p_2\lambda_2^{-1}); \tag{5.6}$$

see (16) of Whitt (1984). The squared coefficient of variation gives a first-order description of the variability, while the parameter r indicates the shape of the distribution for given first two moments. An attractive feature of c^2 and r is that they are independent of the measuring units, so that the numerical values have meaning independent of m_1. The standard two-moment H_2 fit is based on the case of 'balanced means' corresponding to $r = 1/2$. The parameter r increases with the third moment (and thus also the skewness) for given first two moments.

An alternate way to derive the parameters from the moments that goes directly to r is first to normalize by setting $d^3 = m_3/m_1^3$ and then solve

$$\begin{aligned} \gamma &= 3(c^2 - 1)^2/(d^3 - 9c^2 + 3) \\ \alpha &= \pm[(c^2 - 1)/(c^2 - 1 + 2\gamma^2)]^{\frac{1}{2}} \\ 2r &= 1 - \alpha(\gamma - 1), \qquad p_i = (1 \pm \alpha)/2 \\ \lambda_1 &= p_1/rm_1, \qquad \lambda_2 = p_2/(1 - r)m_1. \end{aligned} \tag{5.7}$$

We remark that (5.7) turns out to be especially convenient for the $M/M/1$ queue, because then $\gamma = 2$ independent of the traffic intensity; see Abate and Whitt (1987b). The classical two-moment H_2 fit with $r = \frac{1}{2}$ arises when $\gamma = 1$. The parameter γ is decreasing in r for given first two moments. As additional checks related to (5.7), there are the relations

$$\begin{aligned} &\lambda_2^{-1} > m_1(c^2 + 1)/2 > m_1 > \lambda_1^{-1} > 0 \\ &\lambda_1^{-1} + \lambda_2^{-1} = m_1[2 + (c^2 - 1)/\gamma]. \end{aligned} \tag{5.8}$$

For $c^2 > 3$, $p_1 > \frac{1}{2} > p_2$. The parameter γ in (5.7) can take on negative values when $1 < c^2 < 3$, but α and γ must have the same sign. For the H_2 approximation to the first-moment c.d.f. in (1.13), $r = 0.276$. Since this r is quite different from $\frac{1}{2}$, we should expect that the three-moment H_2 fit would perform significantly better than the standard two-moment fit based on $r=\frac{1}{2}$, as is demonstrated in Table 3.

In some applications, especially when it is desirable to obtain a good fit for small t, we may want to fit the H_2 c.d.f. to the first two moments and the value of the density at the origin. (Here, by (4.4), $h_1(t) \sim (2\pi t)^{-\frac{1}{2}}$ so that $h_1(0)=\infty$.) It is significant that the fitting scheme (5.7) is easily adapted for this purpose, it suffices simply to replace the equation for γ in (5.7) by

$$(5.9) \qquad \gamma = 2(c^2-1)(1-[h(0)m_1]^{-1})/(c^2-3+2[h(0)m_1]^{-1})$$

where $h(0)$ is the value of the density at 0.

5.2. *H_2 approximations via Laplace transforms.* It is also possible and natural to obtain the H_2 approximation (1.13) for the c.d.f. $H_1(t)$ directly from the Laplace transform. In particular, we can recognize that the transform $\hat{H}_1^c(s)$ of $1-H_1(t)$ in (4.9) looks something like the transform of a linear combination of two exponentials. As an approximation, (4.9) can be converted to the general H_2 form $(s+a)/(bs^2+cs+d)$ if we can approximate $(1+2s)^{\frac{1}{2}}$ by a rational approximation of the form $(e+fs)/(g+hs)$. In particular, we propose using a Padé approximant; see Baker (1975) and Section 16.4 of Luke (1969). This does not *a priori* guarantee negative real roots, but if we use $(1+2s)^{\frac{1}{2}} \approx (2+3s)/(2+s)$, both of which have Taylor's series of the form $1+s-s^2/2+o(s^2)$ as $s \to 0$, then we obtain

$$(5.10) \qquad \hat{H}_1^c(s) \approx (s+2)/(s^2+6s+4),$$

which yields precisely (1.13) when inverted. This of course must occur because we have specified $\hat{H}_1^c(0)$ and the first two derivatives at $s=0$, which corresponds to the first three moments. Theorem 1.7 guarantees that we will obtain two real negative roots for $H_1(t)$ in this case.

5.3 *The second-moment c.d.f.* For $k=2$, the H_2 approximation obtained by (5.5) or (5.7) is

$$(5.11) \qquad H_2(t) \approx 1-0.5\exp(-2t)-0.5\exp(-2t/3), \; t \geqq 0.$$

The approximation (5.11) is compared with the exact values in Table 5 (p. 575). The exact values of $H_2(t)$ were obtained by inverting the transform in (4.11) as described in Section 4.4. (Instead we could apply Corollary 1.1.1(b).) In Table 5 these exact values are compared with the approximating exponential c.d.f. $1-e^{-t}$ in (1.12) with $m_{21}=1$ and the asymptotic value from Corollary 1.3.5 as well as the H_2 c.d.f. in (5.11). Note that the quality of the H_2 approximation in (5.11) for $H_2(t)$ is even better than the quality of the approximation for $H_1(t)$. This can be explained in part by the fact that the $H_2(t) \sim 2t$ as $t \to 0$, agreeing with (5.11). In other words, (5.11) turns out to match the value of the density at 0, i.e., $h_2(0)=2$, as well as the first three moments. As a consequence, (5.11) is a good approximation for all $t \geqq 0$. In contrast, $H_1(t) \sim (8t/\pi)^{\frac{1}{2}}$ as $t \to 0$.

The parameter r for (5.11) is $r = 0{\cdot}250$. Since $r = 0{\cdot}276$ for (1.11) and $r = 0{\cdot}250$ for (5.11), it might be better to use $r \approx 0{\cdot}25$ instead of $r = 0{\cdot}50$ for other H_2 approximations of this kind if only two moments are available.

6. Proof of Theorem 1.1

We briefly sketch the argument used by Mitchell (1985) to prove Theorem 1.1(a). It also extends easily to prove Theorem 1.1(b). We start with the Laplace transform (with respect to the space variable) of the density of (1.1), i.e.,

$$\hat{f}(s; t, x) = \int_0^\infty \exp(-sy) f(y; t, x)\, dy. \tag{6.1}$$

The Laplace transform of $\Phi((t-a)/b)$, $t \geqq 0$, is

$$\begin{aligned} \hat{\Phi}_{a,b}(s) &= \int_0^\infty \exp(-sy)\Phi\left(\frac{y-a}{b}\right) dy \\ &= s^{-1}\{\Phi(-ab^{-1}) + [1-\Phi(bs-ab^{-1})]\exp(-s(2a-b^2s)/2)\}, \end{aligned} \tag{6.2}$$

so that

$$\hat{f}(s; t, x) = [A + B_1(s)C_1(s)] - E(s)[A - B_2(s)C_2(s)] \tag{6.3}$$

where

$$\begin{aligned} A &= 1 - \Phi\left(\frac{x-t}{\sqrt{t}}\right) = \Phi\left(\frac{t-x}{\sqrt{t}}\right) \\ B_1(s) &= \exp(-(x-t)s + ts^2/2) \\ B_2(s) &= \exp((x-t)(s+2) + t(s+2)^2/2) \\ C_1(s) &= 1 - \Phi(st^{\frac{1}{2}} - (x-t)t^{-\frac{1}{2}}) \\ C_2(s) &= 1 - \Phi(t^{\frac{1}{2}}(s+2) + (x-t)t^{-\frac{1}{2}}) \\ E(s) &= s/(s+2). \end{aligned} \tag{6.4}$$

The first moment is then

$$\begin{aligned} E(R(t) \mid R(0)=x) &= \frac{-\partial \hat{f}(s; x, t)}{\partial s}\bigg|_{s=0} \\ &= -B_1(0)C_1(0) - B_1(0)C_1'(0) + E'(0)[A - B_2(0)C_2(0)] \end{aligned}$$

while the second moment is

$$\begin{aligned} E(R(t)^2 \mid R(0)=x) &= \left.\frac{\partial^2 \hat{f}(s;t,x)}{\partial s^2}\right|_{s=0} \\ &= B_1(0)C_1''(0) + 2B_1'(0)C_1'(0) + B_1''(0)C_1(0) \\ &\quad + 2^{-1}(A - B_2(0)C_2(0)) + B_2'(0)C_2(0) + B_2(0)C_2'(0). \end{aligned}$$

Additional details appear in unpublished appendices.

Acknowledgment

We thank Chris Mitchell for showing us how to establish Theorem 1.1.

References

ABATE, J. AND WHITT, W. (1987a) Transient behavior of regulated Brownian motion, II: non-zero initial conditions. *Adv. Appl. Prob.* **19,** 599–631.

ABATE, J. AND WHITT, W. (1987b) Transient behavior of the $M/M/1$ queue: Starting at the origin. *Queueing Systems* **2.**

ABATE, J. AND WHITT, W. (1987c) Approximate transient behavior of the $GI/G/1$ queue.

ABATE, J. AND WHITT, W. (1988) Transient behavior of the $M/M/1$ queue via Laplace transforms. *Adv. Appl. Prob.* **20** (1).

ABRAMOWITZ, M. AND STEGUN, I. A. (1972) *Handbook of Mathematical Functions.* Dover, New York.

BAKER, G. A., JR. (1975) *Essentials of Padé Approximants.* Academic Press, New York.

BLANC, J. P. C. (1985) The relaxation time of two queueing systems in series. *Commun. Statist.-Stochastic Models* **1,** 1–16.

BOROVKOV, A. A. (1965) Some limit theorems in the theory of mass service, II. *Theory. Prob. Appl.* **10,** 375–400.

BOROVKOV, A. A. (1984) *Asymptotic Methods in Queueing Theory.* Wiley, New York.

CHANDRASEKHAR, S. (1943) Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* **15,** 1–89 (in *Selected Papers on Noise and Stochastic Processes,* ed. N. Wax, Dover, New York, 1954, pp. 3–91).

COFFMAN, E. G. AND REIMAN, M. I. (1984) Diffusion approximations for computer/communications systems. In *Mathematical Computer Performance and Reliability,* ed. G. Iazeolla, P. J. Courtois and A. Hordijk, Elsevier, Amsterdam, 33–53.

COHEN, J. W. (1982) *The Single Server Queue,* 2nd edn. North-Holland, Amsterdam.

COX, D. R. (1962) *Renewal Theory.* Methuen, London.

COX, D. R. AND MILLER, H. D. (1965) *The Theory of Stochastic Processes.* Wiley, New York.

DAVIES, B. AND MARTIN, B. (1979) Numerical inversion of the Laplace transform. *J. Comp. Phys.* **33,** 1–32.

DAVIS, D. (1960) The build-up time of waiting lines. *Nav. Res. Log. Qtrly.* **2,** 185–193.

DOETSCH, G. (1974) *Introduction to the Theory and Application of the Laplace Transformation.* Springer-Verlag, New York.

DONEY, R. A. (1984) Letter to the Editor. *J. Appl. Prob.* **21,** 673–674.

DUBNER, H. AND ABATE, J. (1968) Numerical inversion of Laplace transforms by relating them to the finite Fourier cosine transform. *J. ACM* **15,** 115–123.

FELLER, W. (1968) *An Introduction to Probability Theory and its Applications,* Vol. II, 3rd edn. Wiley, New York.

FELLER, W. (1971) *An Introduction to Probability Theory and Its Applications,* Vol. II, 2nd edn. Wiley, New York.

FLORES, C. (1985) Diffusion approximations for computer communications networks. In *Computer Communications,* Proc. Symp. Appl. Math. **31,** American Math. Society, Providence, 83–124.

GAVER, D. P., JR. (1966) Observing stochastic processes and approximate transform inversion. *Operat. Res.* **14,** 444–459.

GAVER, D. P., JR. (1968) Diffusion approximations and models for certain congestion problems. *J. Appl. Prob.* **5,** 607–623.

HARRISON, J. M. (1985) *Brownian Motion and Stochastic Flow Systems.* Wiley, New York.

IGLEHART, D. L. AND WHITT, W. (1970a) Multiple channel queues in heavy traffic, I. *Adv. Appl. Prob.* **2,** 150–177.

IGLEHART, D. L. AND WHITT, W. (1970b) Multiple channel queues in heavy traffic II: sequences, networks and batches. *Adv. Appl. Prob.* **2,** 355–369.

JOHNSON, N. L. AND KOTZ, S. (1970). *Distributions In Statistics, Continuous Univariate Distributions* I, Wiley, New York.

KARLIN, S. AND TAYLOR, H. M. (1975) *A First Course in Stochastic Processes,* 2nd edn. Academic Press, New York.

KEILSON, J. (1979) *Markov Chain Models—Rarity and Exponentiality.* Springer-Verlag, New York.

KEILSON, J. AND KESTER, A. (1977) Monotone matrices and monotone Markov processes. *Stoch. Proc. Appl.* **5,** 231–241.

KELLY, F. P. (1979) *Reversibility and Stochastic Networks.* Wiley, New York.

KLEINROCK, L. (1976) *Queueing Systems, Vol. 2: Computer Applications.* Wiley, New York.

LEE, I. (1985) *Stationary Markovian Queueing Systems: An Approximation for the Transient Expected Queue Length,* M. S. dissertation, Department of Electrical Engineering and Computer Science, MIT, Cambridge.

LEE, I. AND ROTH, E. (1986) Stationary Markovian queueing systems: an approximation for the transient expected queue length.

LUKE, Y. L. (1969) *The Special Functions and Their Approximants.* Academic Press, New York.

MAGNUS, W., OBERHETTINGER, F. AND SONI, R. P. (1966) *Formulas and Theorems for the Special Functions of Mathematical Physics.* Springer-Verlag, New York.

MITCHELL, J. C. (1985) Lost-sales inventory systems with a service objective, I: stationary demand, linear procurement costs and fixed lead times.

NEWELL, G. F. (1982) *Applications of Queueing Theory,* 2nd edn. Chapman and Hall, London.

OBERHETTINGER, F. AND BADII, L. (1973) *Tables of Laplace Transforms.* Springer-Verlag, New York.

ODONI, A. R. AND ROTH, E. (1983) An empirical investigation of the transient behavior of stationary queueing systems. *Operat. Res.* **31,** 432–455.

PHATARFOD, R. M., SPEED, T. P. AND WALKER, A. M. (1971) A note on random walks. *J. Appl. Prob.* **8,** 198–201.

PIESSENS, R. AND HUYSMANS, R. (1984) Algorithm 619. Automatic numerical inversion of the Laplace transform. *ACM Trans. Math. Softw.* **10,** 348–353.

ROTH, E. (1981) *An Investigation of the Transient Behavior of Stationary Queueing Systems,* Ph.D. Dissertation, M.I.T., Cambridge.

STEHFEST, H. (1970) Algorithm 368. Numerical inversion of Laplace transforms. *Comm. ACM* **13,** 47–49 (erratum **13,** 624).

STONE, C. (1963) Limit theorems for random walks, birth and death processes, and diffusion processes. *Illinois J. Math.* **7,** 638–660.

STOYAN, D. (1983) *In Comparison Methods for Queues and Other Stochastic Models,* ed. D. J. Daley. Wiley, Chichester.

SUMITA, U. (1984) On conditional passage time structure of birth-death processes. *J. Appl. Prob.* **21,** 10–21.

TAKÁCS, L. (1967) *Combinatorial Methods in the Theory of Stochastic Processes.* Wiley, New York.

VAN DOORN, E. (1980) *Stochastic Monotonicity and Queueing Applications of Birth–Death Processes.* Lecture Notes in Statistics **4,** Springer-Verlag, New York.

VEILLON, F. (1974) Algorithm 486. Numerical inversion of Laplace transform. *Comm. ACM* **17,** 587–589.

WHITT, W. (1982) Approximating a point process by a renewal process, I: two basic methods. *Operat. Res.* **30,** 125–147.

WHITT, W. (1984) On approximations for queues, III: mixtures of exponential distributions. *AT & T Bell Lab. Tech. J.* **63,** 163–175.

WHITT, W. (1985) The renewal-process stationary-excess operator. *J. Appl. Prob.* **22,** 156–167.