SIAM J. APPL. MATH.
Vol. 51, No. 2, pp. 525-541, April 1991

012

ASYMPTOTIC ANALYSIS FOR A CLOSED PROCESSOR-SHARING SYSTEM WITH SWITCHING TIMES: NORMAL USAGE*

ALBERTO BERSANI† AND CECILIA SCIARRETTA†

Abstract. In a closed processor-sharing system with CPU-switching times, the probability distribution of the waiting time W, $P_\varepsilon(t)=\mathbb{P}\{W>t \mid \varepsilon \geqq 0\}$ where ε is the switching time is studied. The waiting time is the time needed for a particular job, called a *tagged* job, to be served by the CPU. The switching time ε is the time spent by the CPU to transfer from one job to the next. In this paper, ε is used as a perturbative parameter. The system consists of a bank of $N+1$ terminals, with N large, in series with a CPU, which feeds back to the terminals.

Let p/q be the ratio of the mean required service time to the mean think time. The state of the system is characterized by the traffic intensity $\rho = Np/q = O(1)$.

Let $\pi_i(\varepsilon)$ be the stationary probabilities, i.e., the probabilities of having i jobs ($i=0,1,\cdots,N$) requiring service by the CPU, when the system is in the steady state. We write $\pi_i(\varepsilon)$ as a polynomial in ε. In order to compute the equilibrium distribution $P_\varepsilon(t)$, the first two moments $E(W)$ and $E(W^2)$ and the variance $\sigma^2(W)$ for the waiting time, two ordinary differential equations of the third order are written down and solved for normal usage ($\rho<1$) by means of asymptotic expansions in $1/N$ and ε. The first perturbative terms in $1/N$ and ε are computed for $P_\varepsilon(t)$, $E(W)$, $E(W^2)$, and $\sigma^2(W)$. It is noted that the introduction of a nonzero switching time (that is, $\varepsilon>0$) implies an increase in the expectation value $E(W)$ and in the variance $\sigma^2(W)$ of the waiting time W. Finally, numerical results are used to illustrate experimentally the range of validity of the approximate expressions that have been obtained.

Key words. queueing networks, processor-sharing, state-dependent queues, asymptotic expansions

AMS(MOS) subject classifications. 60K25, 68M20, 90B22, 34E20

1. Introduction. Several authors have recently studied a model of a processor-sharing system consisting of a bank of $(N+1)$ terminals (see Fig. 1).

The system is assumed closed, i.e., the number of terminals is finite and constant. The users alternate between thinking periods, during which they generate jobs, and waiting periods, during which they wait for the CPU to provide service to their jobs.

The think times and the required service times are assumed to be random variables independently, identically, and exponentially distributed, with mean values $1/p$ and $1/q$, respectively. We define the parameter ρ as follows:

$$\rho = N\cdot\frac{p}{q}, \tag{1.1}$$

that is, ρ is the traffic intensity in the CPU.

The discipline of service is assumed to be processor-sharing: if i jobs are present in the CPU, they receive service "simultaneously" at $1/i$ times the rate a job receives service when it is the only one present.

Several authors have been interested in the waiting (or response) time, that is, the time W needed for a tagged job J_* to be served by the CPU. By tagged job we mean a particular job J_* entering the CPU at $t=0$, so that the probability that J_* finds i jobs at its arrival is π_i, the probability of having i jobs in the CPU. While for the CPU there is no difference between J_* and the other jobs requiring service, we will distinguish

* Received by the editors August 17, 1988; accepted for publication (in revised form) March 29, 1990. This research was supported by Italsiel S.P.A., Rome, and the graduate fellowship program of the Istituto Nazionale di Alta Matematica "F. Severi."

† Istituto Nazionale di Alta Matematica, "F. Severi," P. le Aldo Moro, 5, 00185 Rome, Italy.

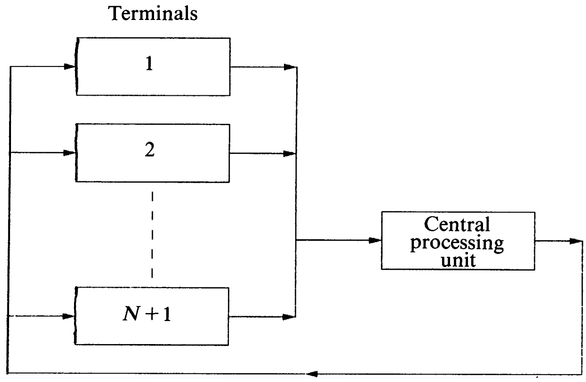

FIG. 1. *The physical system.*

the state of the CPU according to the presence of J_*. We will indicate with "i" the state with $i+J_*$ jobs in the CPU, and with "$\tilde{i}$" every state of the CPU in which i jobs, different from J_*, are served.

Mitra [6] studied the behavior of the waiting time equilibrium probability distribution and its moments; he found the solution of this problem in matrix form. However, when the system is large (i.e., N is large) the matrix inversions needed to evaluate Mitra's formulae are somewhat cumbersome. In order to avoid this, Mitra and Morrison [8] turned the problem into one solving a second-order differential equation through the introduction of a generating function. For the normal usage case ($\rho<1$) the approximate solutions for $E(W)$ and $E(W^2)$—the first two moments of the waiting time—were given by means of asymptotic series in powers of $1/N$. It is also possible to investigate the open system (see Fig. 2), that is, a queue of jobs entering a CPU, which serves them and does not feed them back to the terminals.

For an open system we assume that the job arrivals are a Poisson process with

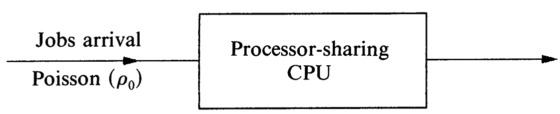

FIG. 2. *The open system.*

rate $\rho_0<1$. It is possible to obtain this system from the closed one, with the limiting procedure:

$$N\to\infty,\quad \frac{p}{q}\to 0,\quad \rho\to\rho_0.$$

Morrison [10] found the explicit expression of the waiting-time distribution $\mathbb{P}\{W>t\}=P(t)$ for an open system with CPU-processor-sharing discipline. Using essentially the same techniques as in [8], Morrison [11] arrived at the approximate solution for the waiting time distribution for the closed system in the normal usage case, as a perturbative series in powers of $1/N$, whose leading order term coincides with the explicit solution for the open system. Moreover, in [11] Morrison investigated the heavy ($\rho\simeq 1$) and very heavy ($\rho>1$) cases. The model has been generalized to systems with multiple job classes in [9]-[12].

However, in actual processor-sharing systems the jobs are not served simultaneously; in fact the CPU, in presence of i jobs requiring service, transfers the control from one job to the next one i times in each unit of service time. Hence it is only approximately true that the time spent on each job per unit time is $1/i$, due to the time spent to transfer the control between jobs. The time spent by the CPU to transfer from one job to the next will be called the "switching time." In [3], [6], [8]-[12], the existence of the switching times is neglected. However, in general, they are not negligible, particularly when we are dealing with a large number of jobs requiring service.

The dependence of the behavior of the CPU on the number of jobs requiring service may be described by a state-dependent queueing system. That is, a system where some of the quantities of interest depend on variables representing the load (or state) of the server, e.g., the CPU. State-dependent systems are used to model, for example, systems with discouraged arrivals, finite capacity or finite buffer content, finite population size, or finally multiple-server system. Several authors have treated state-dependent systems [1], [2], [4], [5], [7]. The load-dependence consists, in general, of assuming that the arrival rate and/or the service rate depend on the number of jobs in the CPU. Benedetti [1], for a server with the processor-sharing discipline, has assumed a monotonically decreasing dependence of the service rate on the load of the CPU.

If we define the CPU efficiency as

$$\mu(i)=\frac{\text{mean of total service time required by } i \text{ jobs}}{\text{mean of total time spent by CPU to process } i \text{ jobs}},$$

we observe that, while for zero-switching times we have $\mu(i)=1$ ($i=1,\cdots,N+1$), when we introduce the switching time ε, the efficiency is a decreasing function of i, that we will choose to be

$$\mu(i)=\frac{1}{1+i\varepsilon},\qquad i=1,\cdots,N+1. \tag{1.2}$$

This assumption is due to Benedetti, and characterizes the state dependence of a single-server processor-sharing system. Increasing the number of jobs in the CPU corresponds to a considerable decrease of $\mu(i)$. Since the time spent by the CPU in transferring its service from one job to the next must be only a small fraction of the mean required service time, ε is a small parameter and will be used as a perturbative parameter. We will derive expressions written as power series of ε, such that the leading order term ($\varepsilon=0$) will give the corresponding expressions in the zero-switching time models, described in [3], [6], [8]-[12].

In §§ 2 and 3 we describe the system; the stationary probabilities $\pi_i(\varepsilon)$, defined as the probabilities to have i jobs requiring service from the CPU in the steady state, are computed as polynomials of ε.

In order to study the waiting time problem, we consider the waiting time equilibrium distribution, defined as follows:

$$P_\varepsilon(t)=\mathbb{P}(W>t\,|\,\varepsilon\geqq 0),\qquad 0\leqq t<\infty \tag{1.3}$$

and its first two moments $E(W)$ and $E(W^2)$.

In order to express $E(W)$ and $E(W^2)$ in power series of $1/N$ and ε, we introduce the generating functions $C(N, z, \varepsilon)$ and $G(N, z, \varepsilon)$, the former satisfying a third-order linear ordinary differential equation.

Furthermore, the case of normal usage ($\rho<1$) is studied, and the following expressions to first order in $1/N$ and ε for $E(W)$, $E(W^2)$, $P_\varepsilon(t)$ and the variance

$\sigma^2(W)$ are obtained:

(1.4) $$E(W) \sim \frac{1}{1-\rho} - \frac{2\rho^2}{N(1-\rho)^3} + \frac{(1+\rho)\varepsilon}{(1-\rho)^3},$$

(1.5) $$\frac{1}{2}E(W^2) \sim \frac{2}{(2-\rho)(1-\rho)^2} + \frac{2\rho^2(8\rho-13)}{N(1-\rho)^4(2-\rho)^2} - \frac{2\varepsilon(\rho^3+5\rho^2-7\rho-4)}{(2-\rho)^2(1-\rho)^4},$$

(1.6) $$P_\varepsilon(t) \sim \left[1 + \frac{\rho^2}{N(1-\rho)^2} - \frac{\varepsilon\rho}{(1-\rho)^2}\right] \cdot P(t) + \frac{\rho}{(1-\rho)^2}\left(\varepsilon - \frac{\rho}{N}\right) T^{3/2} K_3(2\sqrt{T}),$$

(1.7) $$\sigma^2(W) \sim \frac{2+\rho}{(1-\rho)^2(2-\rho)} + \frac{4\rho^2(\rho^2+4\rho-9)}{N(1-\rho)^4(2-\rho)^2} - \frac{2\varepsilon(3\rho^3+7\rho^2-14\rho-4)}{(1-\rho)^4(2-\rho)^2},$$

where $P(t)$ is the corresponding distribution for the zero-switching time open system, $T=(1-\rho)t$, and $K_i(z)$ are the modified Bessel functions of the third kind.

We remark that the effect of the first-order correction in ε in our model consists of an increase of $P_\varepsilon(t)$, $E(W)$ and the variance $\sigma^2(W)$. Moreover, for $\varepsilon=0$, (1.4)-(1.6) reduce to the expressions for the corresponding quantities obtained by Mitra and Morrison in their zero-switching time model [8, eqs. (62), (63)], [11, eq. (3.34)]. The correction terms in ε in (1.6) are $O(\varepsilon/\alpha^2)$, where $\alpha=1-\rho$, so that, for $0<\alpha\ll 1$, we assume that $\varepsilon/\alpha^2 \ll 1$, rather than $\varepsilon \ll 1$. On the other hand, we already know, from [11], that (1.6) holds for $N \gg 1$, $N\alpha^2 \gg 1$. Moreover, in (2.8) we will show that it is sufficient to assume $\varepsilon \ll \bar{\varepsilon}$, with $\bar{\varepsilon}$ given by (2.6)—when $\rho \leqq N/(N+1)$—or $\varepsilon \ll \tilde{\varepsilon}$, with $\tilde{\varepsilon}$ given by (2.7)—when $1>\rho \geqq N/(N+1)$—together with the previous conditions on ε, α, N, to guarantee that (1.4)-(1.7) are satisfactory approximations of the exact quantities (see also Tables 1-4).

In § 4 some numerical results concerning the distribution of the waiting time and its moments are discussed. Some comparisons between the explicit expressions of $\mathscr{G}(N, \varepsilon) = G(N, 1, \varepsilon)$ and $E(W)$—given by (2.3) and (2.12), respectively—and their approximations, are given. The range of validity of our results is investigated, varying the two parameters N and ε.

2. The equations describing the model. It is known [3] that queueing models for computer networks may be described by means of Markovian birth-and-death processes, where i, the number of jobs requiring service from the CPU, is a stochastic variable. The think times and the required service times are random variables exponentially, independently, and identically distributed, with means $1/p$ and $1/q$, respectively. Without loss of generality we can assume that $q=1$, so that $\rho = N \cdot p$. In order to evaluate, for a tagged job J_* arriving at the CPU, the waiting time distribution and its first two moments, let us introduce the stationary probabilities

$$\pi_i(\varepsilon) = \lim_{t\to\infty} \pi_i(\varepsilon, t)$$
$$= \lim_{t\to\infty} \mathbb{P}\{i \text{ jobs in the CPU at time } t \mid \varepsilon \geqq 0\}, \quad i=1, \cdots, N.$$

For the closed system, described in § 1, where we make use of (1.2), Benedetti [1] has shown that

$$\pi_i(\varepsilon)=\frac{1}{\mathcal{G}(N,\varepsilon)}\binom{N}{i} i!\,p^i \prod_{k=0}^{i}(1+k\varepsilon), \tag{2.1}$$

where the partition (or normalization) function $\mathcal{G}(N,\varepsilon)$ is determined by the condition

$$\sum_{i=0}^{N}\pi_i(\varepsilon)=1; \tag{2.2}$$

thus

$$\mathcal{G}(N,\varepsilon)=\sum_{i=0}^{N}\binom{N}{i} i!\,p^i \prod_{k=0}^{i}(1+k\varepsilon). \tag{2.3}$$

Let us introduce the generating function

$$G(N,z,\varepsilon)=\left[\sum_{i=0}^{N}\pi_i(\varepsilon)z^i\right]\mathcal{G}(N,\varepsilon), \qquad z\in\mathbb{C}, \quad |z|\leqq 1 \tag{2.4}$$

so that

$$G(N,1,\varepsilon)=\mathcal{G}(N,\varepsilon)$$

and

$$G(N,z,\varepsilon)=\sum_{i=0}^{N}\binom{N}{i}(pz)^i i! \prod_{k=0}^{i}(1+k\varepsilon). \tag{2.5}$$

If $N\gg 1$, $\rho\leqq N/(N+1)$, $\varepsilon\ll\bar{\varepsilon}$, where

$$\begin{aligned}\bar{\varepsilon}&=\frac{2N-\rho(N+1)+\sqrt{4N\cdot[N-\rho(N+1)]}}{\rho(N+1)^2}\\ &=\frac{(1+\sqrt{\alpha})}{(1-\sqrt{\alpha})N}+O\left(\frac{1}{N^2}\right)\end{aligned} \tag{2.6}$$

or $N\gg 1$, $1>\rho>N/(N+1)$, $\varepsilon\ll\tilde{\varepsilon}$, where

$$\tilde{\varepsilon}=\frac{\alpha}{\rho} \tag{2.7}$$

we approximate $\prod_{k=0}^{i}(1+k\varepsilon)$ with $(1+(i(i+1)/2)\varepsilon)$ so that (2.5) becomes

$$\begin{aligned}G(N,z,\varepsilon)&\sim\sum_{i=0}^{N}\left(1+\frac{i(i+1)}{2}\varepsilon\right)\binom{N}{i} i!\,(pz)^i\\ &= G(N,z,0)+\frac{\varepsilon\cdot z}{2}[2\cdot G_z(N,z,0)+z\cdot G_{zz}(N,z,0)],\end{aligned} \tag{2.8}$$

—where G_z denotes the derivative with respect to z—and

$$\frac{1}{\mathcal{G}(N,\varepsilon)}\sim\frac{1}{\mathcal{G}(N,0)}\left\{1-\frac{\varepsilon}{2\mathcal{G}(N,0)}[2G_z(N,z,0)+z\cdot G_{zz}(N,z,0)]_{z=1}\right\}. \tag{2.9}$$

In fact, in this case

$$\pi_i(\varepsilon)\gg\pi_{i+1}(\varepsilon), \qquad i=0,1,\cdots,N-1$$

and the approximation of $\prod_{k=0}^{i}(1+k\varepsilon)$ with $(1+(i(i+1)/2)\varepsilon)$ in (2.5) is valid.

Let us now consider a tagged job J_*, which arrives at the CPU at time $t=0$. Let Q be the number of jobs—different from J_*—present in the CPU at $t=0$; we introduce

$$p_i(\varepsilon, t) = \mathbb{P}\{W > t \text{ and } Q = i \mid \varepsilon \geqq 0\}, \qquad i = 0, \cdots, N.$$

We have

$$p_i(\varepsilon, 0) = 1, \qquad i = 0, 1, \cdots, N \tag{2.10}$$

and

$$P_\varepsilon(t) = \sum_{i=0}^{N} \pi_i(\varepsilon) p_i(\varepsilon, t).$$

Moreover, the probabilities $p_i(\varepsilon, t)$ satisfy the following $(N+1)$ differential equations [1]:

$$\begin{aligned} \frac{dp_i(\varepsilon, t)}{dt} &= (N-i)\frac{\rho}{N} p_{i+1}(\varepsilon, t) \\ &\quad + \frac{i}{(i+1)[1+(i+1)\varepsilon]} \cdot p_{i-1}(\varepsilon, t) \\ &\quad - \left[(N-i)\cdot\frac{\rho}{N} + \frac{1}{1+(i+1)\varepsilon}\right] p_i(\varepsilon, t), \qquad i = 0, 1, \cdots, N \end{aligned} \tag{2.11}$$

with initial conditions given by (2.10).

It is known [1] that the first two moments for the waiting time distribution, by virtue of the solution of (2.11), are given by

$$E(W) = \pi^T(\mathbb{B} + \varepsilon\mathbb{B}^2)\mathbb{U}, \tag{2.12}$$

$$\tfrac{1}{2}E(W^2) = \pi^T(\mathbb{B} + \varepsilon\mathbb{B}^2)(\mathbb{I} - \mathbb{A})^{-1}(\mathbb{B} + \varepsilon\mathbb{B}^2)\mathbb{U}, \tag{2.13}$$

where $\pi^T = (\pi_0(\varepsilon), \cdots, \pi_N(\varepsilon))$; $\mathbb{U} = (1, \cdots, 1)^T$; $\mathbb{I}$ is the identity matrix; $\mathbb{B}$ is the diagonal matrix with entries $B_{i,i} = i+1$; $\mathbb{A}$ is a tridiagonal matrix whose entries are

$$A_{i,i} = -[(i+1)[1+(i+1)\varepsilon]\cdot(N-i)p + i], \qquad i = 0, \cdots, N,$$

$$A_{i,i-1} = i, \qquad i = 0, \cdots, N,$$

$$A_{i,i+1} = (i+1)[1+(i+1)\varepsilon]\cdot(N-i)p, \qquad i = 0, \cdots, N.$$

If we introduce the vector $C = (c_0, \cdots, c_N)^T$, defined as

$$C^T = \pi^T(\mathbb{B} + \varepsilon\mathbb{B}^2)(\mathbb{I} - \mathbb{A})^{-1}, \tag{2.14}$$

we may write

$$E(W) = \sum_{i=0}^{N} c_i, \tag{2.15}$$

$$\frac{1}{2}E(W^2) = \sum_{i=0}^{N} c_i[(i+1) + \varepsilon(i+1)^2], \tag{2.16}$$

where we have used the property $\mathbb{A}\mathbb{U} = 0$.

To obtain an approximate expression for $E(W)$ and $E(W^2)$, we consider the generating function $C(N, z, \varepsilon)$ given by

$$C(N, z, \varepsilon) = \left[\sum_{i=0}^{N} c_i z^i\right] \mathcal{G}(N, \varepsilon) \tag{2.17}$$

and we rewrite equation (2.14) as a differential equation for $C(N, z, \varepsilon)$, multiplying both sides of (2.14) by z^i and summing over i. For the sake of simplicity, let us indicate the derivative with respect to z with a prime. We have

$$\begin{aligned}&[1+\rho(1-z)+\rho(1-z)\varepsilon]C+(1-z)\left[\rho z-1-\frac{2\rho z}{N}+\varepsilon\left(3\rho z-\frac{4\rho z}{N}\right)\right]\\ &\cdot C'+\rho(1-z)\cdot\left[\frac{-z^2}{N}+\varepsilon z^2\left(1-\frac{5}{N}\right)\right]\cdot C''-\frac{\rho}{N}(1-z)\varepsilon z^3\cdot C'''\\ &=H'+\varepsilon(H'+zH''),\end{aligned}\tag{2.18}$$

where $H(N, z, \varepsilon)=z\cdot G(N, z, \varepsilon)$ so that $H(N, 1, \varepsilon)=G(N, 1, \varepsilon)=\mathcal{G}(N, \varepsilon)$.

We note that the presence of ε introduces a third-order derivative in the differential equation for $C(N, z, \varepsilon)$. We expand $C(N, z, \varepsilon)$ to first order in $1/N$ and then we expand to first order in ε, to obtain the following approximate expression:

$$C\sim C_{00}+\frac{1}{N}C_{10}+\varepsilon C_{01}.\tag{2.19}$$

Substituting (2.19) in (2.18) and equating the coefficients of the powers of the same order, we arrive at the following three differential equations:

$$(1-z)(\rho z-1)C'_{00}+[1+(1-z)\rho]C_{00}=H'_{00},\tag{2.20}$$

$$\begin{aligned}&(1-z)(\rho z-1)C'_{01}+[1+(1-z)\rho]C_{01}\\ &\quad=H'_{01}+H'_{00}+zH''_{00}-(1-z)\rho[z(zC_{00})']',\end{aligned}\tag{2.21}$$

$$(1-z)(\rho z-1)C'_{10}+[1+(1-z)\rho]C_{10}=H'_{10}+\rho(1-z)[z^2C'_{00}]',\tag{2.22}$$

where we have

$$H(N, z, \varepsilon)\sim H_{00}+\frac{1}{N}H_{10}+\varepsilon H_{01}.$$

These equations have the general form

$$(1-z)(\rho z-1)f'+[1+(1-z)\rho]f=g(\rho, z),\tag{2.23}$$

where $g(\rho, z)$ represents the nonhomogeneous term. Since from (2.15)-(2.17) we have

$$E(W)=\frac{C(N, 1, \varepsilon)}{\mathcal{G}(N, \varepsilon)},$$

$$\begin{aligned}\frac{1}{2}E(W^2)=\frac{1}{\mathcal{G}(N, \varepsilon)}&\cdot\{C(N, 1, \varepsilon)+C'(N, 1, \varepsilon)+\varepsilon\\ &\cdot[C''(N, 1, \varepsilon)+3C'(N, 1, \varepsilon)+C(N, 1, \varepsilon)]\}\end{aligned}$$

and from (2.19) we have

$$E(W)\sim\frac{1}{\mathcal{G}(N, \varepsilon)}\cdot\left[C_{00}(N, 1, \varepsilon)+\frac{1}{N}C_{10}(N, 1, \varepsilon)+\varepsilon C_{01}(N, 1, \varepsilon)\right],\tag{2.24}$$

$$\begin{aligned}\frac{1}{2}E(W^2)\sim\frac{1}{\mathcal{G}(N, \varepsilon)}\cdot\Big\{&C_{00}(N, 1, \varepsilon)+C'_{00}(N, 1, \varepsilon)+\frac{1}{N}\\ &\cdot[C_{10}(N, 1, \varepsilon)+C'_{10}(N, 1, \varepsilon)]\\ &+\varepsilon\cdot[C_{01}(N, 1, \varepsilon)+C'_{01}(N, 1, \varepsilon)\\ &+C''_{00}(N, 1, \varepsilon)+3C'_{00}(N, 1, \varepsilon)+C_{00}(N, 1, \varepsilon)]\Big\}\end{aligned}\tag{2.25}$$

the problem of evaluating $E(W)$ and $E(W^2)$ is reduced to the problem of computing C_{00}, C_{01}, C_{10} and their derivatives at $z=1$. Equation (2.23) has a singular point at $z=1$ and at $z=1/\rho$, since at these points the coefficient of the leading derivative is zero.

Since we are interested in the behavior of the solution at $z=1$, and we will consider the normal-usage case ($\rho<1$), the point $z=1/\rho>1$ is not playing an important role in our analysis, since we only consider z in the interval [0, 1]. We need to look for the unique solution in the interval [0, 1], which is analytic at $z=1$. The explicit form of the solution of (2.23), which is analytic at $z=1$, is

$$f(z,\rho)=\left[\frac{(1-\rho z)^{\rho}}{(1-z)}\right]^{1/(1-\rho)}\int_z^1\left[\frac{(1-y)^{\rho}}{(1-\rho y)}\right]^{1/(1-\rho)}\cdot g(\rho,y)\,dy,\ |z|\leqq 1. \tag{2.26}$$

The values of $C_{00}(1)$, $C_{01}(1)$, and $C_{10}(1)$ can easily be obtained from (2.19), (2.20), (2.21), and (2.26). In order to compute the derivatives of C_{00}, C_{01}, C_{10} we twice differentiate (2.18) and study the resulting equations at $z=1$. Trivial computations give us the following relations:

$$C'_{00}=\frac{H''_{00}+\rho C_{00}}{(2-\rho)}, \tag{2.27}$$

$$C''_{00}=\frac{H'''_{00}+4\rho C'_{00}}{(3-2\rho)}=\frac{(2-\rho)H'''_{00}+4\rho(H''_{00}+\rho C_{00})}{(3-2\rho)(2-\rho)}, \tag{2.28}$$

$$C'_{01}=\frac{1}{(2-\rho)}[\rho C_{01}+\rho C''_{00}+3\rho C'_{00}+\rho C_{00}+H''_{01}+2H''_{00}+H'''_{00}], \tag{2.29}$$

$$C'_{10}=\frac{1}{(2-\rho)}[\rho C_{10}-\rho C''_{00}-2\rho C'_{00}+H''_{10}]. \tag{2.30}$$

When $\rho<1$, for $N\to\infty$ we may expand $G(N,z,0)$ in an asymptotic series of $1/N$ [8]:

$$G(N,z,0)\sim\frac{1}{1-\rho z}-\frac{\rho^2z^2}{N(1-\rho z)^3}+O\left(\frac{1}{N^2}\right). \tag{2.31}$$

So that, since

$$G_z(N,z,0)\sim\frac{\rho}{(1-\rho z)^2}-\frac{\rho^2 z}{N(1-\rho z)^4}(2+\rho z),$$

$$G_{zz}(N,z,0)\sim\frac{2\rho^2}{(1-\rho z)^3}-\frac{2\rho^2}{N(1-\rho z)^5}(\rho^2z^2+4\rho z+1)$$

for $N(1-\rho)^2\gg 1$ and $\varepsilon\ll(1-\rho)^2$, from (2.8) we have

$$G(N,z,\varepsilon)\sim\frac{1}{1-\rho z}-\frac{\rho^2z^2}{N(1-\rho z)^3}+\frac{\varepsilon\rho z}{(1-\rho z)^3}, \tag{2.32}$$

$$H(N,z,\varepsilon)\sim\frac{z}{1-\rho z}-\frac{\rho^2z^3}{N(1-\rho z)^3}+\frac{\varepsilon\rho z^2}{(1-\rho z)^3}, \tag{2.33}$$

$$\mathcal{G}(N,\varepsilon)\sim\frac{1}{1-\rho}-\frac{\rho^2}{N(1-\rho)^3}+\frac{\varepsilon\rho}{(1-\rho)^3}, \tag{2.34}$$

$$\frac{1}{\mathcal{G}(N,\varepsilon)}\sim(1-\rho)\left[1+\frac{\rho^2}{N(1-\rho)^2}-\varepsilon\frac{\rho}{(1-\rho)^2}\right]. \tag{2.35}$$

Moreover, using (2.33) to compute the first three derivatives with respect to z of $H(N, z, \varepsilon)$, from (2.27)-(2.30), we have

$$C_{00}(1)=\frac{1}{(1-\rho)^2}, \qquad C_{01}(1)=\frac{1+2\rho}{(1-\rho)^4},$$

$$C_{10}(1)=\frac{-3\rho^2}{(1-\rho)^4}, \qquad C'_{00}(1)=\frac{3\rho-\rho^2}{(2-\rho)(1-\rho)^3},$$

$$C''_{00}(1)=\frac{2\rho^2(4-\rho)}{(2-\rho)(1-\rho)^4},$$

$$C'_{01}(1)=\frac{\rho(2\rho^3-15\rho^2+9\rho+16)}{(1-\rho)^5(2-\rho)^2},$$

$$C'_{10}(1)=\frac{-3\rho^2(\rho^3-5\rho^2+2\rho+6)}{(2-\rho)^2(1-\rho)^5},$$

so that, from (2.24) and (2.25), we have—for $E(W)$ and $E(W^2)$, respectively—formulae (1.4) and (1.5).

Finally, the variance $\sigma^2(W)$, obtained from (1.4) and (1.5), is given by (1.7).

Formulae (1.4) and (1.5) are consistent with the corresponding ones for the zero-switching time model [8, eqs. (62)-(63)] if we set $\varepsilon=0$. The effect of the nonzero switching time ε is a growth of the expectation value $E(W)$ and of the variance $\sigma^2(W)$ of the waiting time, due to the positiveness of the third term in the right-hand side of (1.4) and (1.7), for $\rho\in[0, 1]$.

3. The waiting time distribution. For our purpose it is more convenient to introduce $\tilde{Q}_t$, that is, the number of jobs, different from J_*, in the CPU at time t ($\tilde{Q}_0=Q$); then we have

$$g_i(\varepsilon, t)=\mathbb{P}\{W>t \text{ and } \tilde{Q}_t=i \mid \varepsilon\geqq 0\}, \qquad i=0,\cdots,N$$

and

$$P_\varepsilon(t)=\sum_{i=0}^{N} g_i(\varepsilon, t), \tag{3.1}$$

$$g_i(\varepsilon, 0)=\pi_i(\varepsilon), \qquad i=0,\cdots,N. \tag{3.2}$$

Equation (3.2) tells us that the probability that J_*, at its arrival, finds i jobs in the CPU is just the probability that the system is in the state $\tilde{i}=(i$ jobs in the CPU$\}$.

The following differential equation for $g_i(\varepsilon, t)$ ($i=0,\cdots,N$), analogous to (2.11) for $p_i(\varepsilon, t)$ ($i=0,\cdots,N$), can be derived:

$$\begin{aligned}\frac{dg_i}{dt}(\varepsilon, t)=(N-i+1)\frac{\rho}{N}g_{i-1}(\varepsilon, t)-\left[(N-i)\frac{\rho}{N}+\frac{1}{[1+(i+1)\varepsilon]}\right]\\ \cdot g_i(\varepsilon, t)+\left(\frac{i+1}{i+2}\right)\cdot\frac{1}{[1+(i+2)\varepsilon]}g_{i+1}(\varepsilon, t),\end{aligned} \tag{3.3}$$

$$i=0,\cdots,N$$

with (3.2) as initial conditions.

Defining a new vector function $f(\varepsilon, t)$ whose components are given by

$$f_i(\varepsilon, t)=\frac{g_i(\varepsilon, t)}{(i+1)[1+(i+1)\varepsilon]}, \qquad i=0,\cdots,N \tag{3.4}$$

we have that (3.3) becomes

$$(3.5)\qquad \begin{aligned}(i+1)[1+(i+1)\varepsilon]\cdot\frac{df_i}{dt}=\frac{\rho}{N}(N-i+1)i(1+i\varepsilon)\cdot f_{i-1}+(i+1)\cdot f_{i+1}\\ -\left\{\frac{\rho}{N}(N-i)[1+(i+1)\varepsilon]+1\right\}(i+1)\cdot f_i\end{aligned}$$

with initial conditions

$$(3.6)\qquad f_i(\varepsilon,0)=\frac{\pi_i(\varepsilon)}{(i+1)[1+(i+1)\varepsilon]},\qquad i=0,\cdots,N.$$

Let us introduce the generating function for $f_i(\varepsilon, t)$:

$$(3.7)\qquad F(z,t,N,\varepsilon)=\left[\sum_{i=0}^{N} f_i(\varepsilon,t)z^i\right]\cdot\mathscr{G}(N,\varepsilon).$$

Multiplying each side of (3.5) by z^i, and summing over i, we obtain the following partial differential equation for F:

$$(3.8)\qquad \begin{aligned}&\frac{\partial}{\partial t}[(1+\varepsilon)F+(1+3\varepsilon)z\cdot F_z+\varepsilon z^2F_{zz}]\\ &=[(z-1)\rho(1+\varepsilon)-1]F+(1-z)\left\{\frac{\rho}{N}z[2-N+(4-3N)\varepsilon]+1\right\}F_z\\ &\quad-\frac{\rho}{N}z^2(1-z)[(N-5)\varepsilon-1]\cdot F_{zz}-\frac{\rho}{N}\varepsilon z^3(z-1)\cdot F_{zzz}\end{aligned}$$

and from (3.6) we have the initial condition

$$(3.9)\qquad [\varepsilon z^2F_{zz}+(1+3\varepsilon)z\cdot F_z+(1+\varepsilon)F]_{t=0}=G(N,Z,\varepsilon).$$

Because of (3.1), we may express the waiting time distribution in terms of F and G, that is,

$$(3.10)\qquad P_\varepsilon(t)=\frac{1}{\mathscr{G}(N,\varepsilon)}[(1+\varepsilon)F+(1+3\varepsilon)F_z+\varepsilon F_{zz}]_{z=1}.$$

To eliminate the time-dependence in (3.8) let us take the Laplace transform of F with respect to t:

$$\bar{F}(z,s,N,\varepsilon)=\int_0^\infty e^{-st}F(z,t,N,\varepsilon)\,dt;$$

using (3.9) and simple properties of the Laplace transform, (3.8) becomes

$$(3.11)\qquad \begin{aligned}&[(z-1)\rho(1+\varepsilon)-1]\cdot\bar{F}+(1-z)\left\{\frac{\rho}{N}z[2-N+(4-3N)\varepsilon]+1\right\}\cdot\bar{F}_z\\ &\quad-\frac{\rho}{N}z^2(1-z)[\varepsilon(N-5)-1]\bar{F}_{zz}-\frac{\rho}{N}\varepsilon z^3(z-1)\bar{F}_{zzz}\\ &=[(1+\varepsilon)\bar{F}+(1+3\varepsilon)z\cdot\bar{F}_z+\varepsilon z^2\bar{F}_{zz}]s-G(N,z,\varepsilon).\end{aligned}$$

Moreover, the Laplace transform of the waiting time distribution may be written as follows:

$$(3.12)\qquad \overline{P_\varepsilon}(s)=\frac{1}{\mathscr{G}(N,\varepsilon)}[(1+\varepsilon)\bar{F}+(1+3\varepsilon)\bar{F}_z+\varepsilon\cdot\bar{F}_{zz}]_{z=1};$$

setting $z=1$ in (3.11), we have

$$[(1+\varepsilon)\bar{F}+(1+3\varepsilon)\bar{F}_z+\varepsilon\bar{F}_{zz}]_{z=1}=\frac{1}{s}[\mathscr{G}(N,\varepsilon)-\bar{F}(1,s,N,\varepsilon)]$$

so that

$$\overline{P_\varepsilon}(s)=\frac{1}{s}\left[1-\frac{\bar{F}(1,s,N,\varepsilon)}{\mathscr{G}(N,\varepsilon)}\right]. \tag{3.13}$$

We may now use (2.9) to write

$$\overline{P_\varepsilon}(s)\sim\frac{1}{s}\left\{1-\frac{\bar{F}(1,s,N,\varepsilon)}{\mathscr{G}(N,0)}\cdot\left[1-\frac{\varepsilon}{2\mathscr{G}(N,0)}(2G_z(N,z,0)+G_{zz}(N,z,0))_{z=1}\right]\right\}. \tag{3.14}$$

Let us expand $\bar{F}$ and G in a power series of $1/N$ and ε, as follows:

$$\begin{aligned}
&\bar{F}(z,s,N,\varepsilon)=\varphi(z,s,\varepsilon)+\frac{1}{N}\psi(z,s,\varepsilon)+O\left(\frac{1}{N^2}\right),\\
&G(N,z,\varepsilon)=G_0(z,\varepsilon)+\frac{1}{N}G_1(z,\varepsilon)+O\left(\frac{1}{N^2}\right),\\
&\varphi(z,s,\varepsilon)=\bar{F}_0(z,s)+\varepsilon\bar{F}_1(z,s)+O(\varepsilon^2),\\
&\psi(z,s,\varepsilon)=\bar{K}_0(z,s)+\varepsilon\bar{K}_1(z,s)+O(\varepsilon^2),\\
&G_0(z,\varepsilon)=G_{00}(z)+\varepsilon G_{01}(z)+O(\varepsilon^2).
\end{aligned} \tag{3.15}$$

Substituting (3.15) in (3.11) and equating the coefficients of the same powers of $1/N$ and ε, we have

$$[\rho z^2-z(s+\rho+1)+1]\bar{F}_0'+[z\rho-(s+\rho+1)]\bar{F}_0=-G_{00}, \tag{3.16}$$

$$[\rho z^2-z(s+\rho+1)+1]\bar{F}_1'+[z\rho-(s+\rho+1)]\bar{F}_1=-(z\rho-\rho-s)[z(z\bar{F}_0)']'-G_{01}, \tag{3.17}$$

$$[\rho z^2-z(\rho+s+1)+1]\bar{K}_0'+[\rho z-(s+\rho+1)]\bar{K}_0=(z-1)\rho[z^2\bar{F}_0']'-G_{10}. \tag{3.18}$$

For $\rho<1$, and $N\gg 1$, using expansion (2.32), (3.16)-(3.18) may be rewritten as follows:

$$[\rho z^2-z(s+\rho+1)+1]\bar{F}_0'+[z\rho-(s+\rho+1)]\bar{F}_0=\frac{-1}{1-\rho z}, \tag{3.19}$$

$$\begin{aligned}&[\rho z^2-z(s+\rho+1)+1]\bar{F}_1'+[z\rho-(s+\rho+1)]\bar{F}_1\\&\quad=-(z\rho-\rho-s)[z(z\bar{F}_0)']'-\frac{z\rho}{(1-\rho z)^3},\end{aligned} \tag{3.20}$$

$$\begin{aligned}&[\rho z^2-z(s+\rho+1)+1]\bar{K}_0'+[z\rho-(s+\rho+1)]\bar{K}_0\\&\quad=(z-1)\rho(z^2\bar{F}_0')'+\frac{\rho^2z^2}{(1-\rho z)^3}.\end{aligned} \tag{3.21}$$

The analytic solution of (3.19) is already known [11], and it is possible to express $\bar{F}_1$ as an integral, depending on $\bar{F}_0$ and its derivatives. Reasoning in a similar way as for (2.20)-(2.22), we may consider the general form of (3.19)-(3.21) given by

$$[\rho z^2-z(\rho+s+1)+1]f'+[z\rho-(s+\rho+1)]f=\Omega(z,\rho), \tag{3.22}$$

where $\Omega(z, \rho)$ is the nonhomogeneous term; equation (3.22) is solved by

$$f = \frac{-r(1-\rho r z)^{\nu-1}}{(z-r)^{\nu}} \int_r^z \frac{\Omega(y, \rho)(y-r)^{\nu-1}}{(1-\rho r y)^{\nu}} \, dy, \tag{3.23}$$

where

$$r = \frac{1}{2\rho}\{(1+\rho+s) - [(1+\rho+s)^2 - 4\rho]^{1/2}\}, \qquad 0 < r < 1$$

is the value of $z \in [0, 1]$ for which the coefficient of the leading derivative in (3.22) is zero and $\nu = (1-\rho r^2)^{-1} > 0$. We note that (3.23) is the unique solution of (3.22) analytic at $z = r$. From (3.23) we obtain the following solutions for (3.19), (3.20), and (3.21):

$$\bar{F}_0 = \frac{r(1-\rho r z)^{\nu-1}}{(z-r)^{\nu}} \int_r^z \frac{(y-r)^{\nu-1}}{(1-\rho y)(1-\rho r y)^{\nu}} \, dy, \tag{3.24}$$

$$\begin{aligned} \bar{F}_1 = {} & \frac{-r(1-\rho r z)^{\nu-1}}{(z-r)^{\nu}} \int_r^z \frac{(y-r)^{\nu-1}}{(1-\rho r y)^{\nu}} \\ & \cdot \left\{ [\rho(1-y)+s](y(y\bar{F}_0)_y)_y - \frac{y\rho}{(1-\rho y)^3} \right\} dy, \end{aligned} \tag{3.25}$$

$$\begin{aligned} \bar{K}_0 = {} & \frac{\rho r(1-\rho r z)^{\nu-1}}{(z-r)^{\nu}} \int_r^z \frac{(y-r)^{\nu-1}}{(1-\rho r y)^{\nu}} \\ & \cdot \left\{ (1-y)[y^2 \bar{F}_{0y}]_y - \frac{\rho y^2}{(1-\rho y)^3} \right\} dy. \end{aligned} \tag{3.26}$$

Let us remark that in (3.11) the leading derivative is multiplied by ε/N so that for $N \to \infty$, $\varepsilon \to 0$, we expect a singular perturbation phenomenon (see [13]). In this case, in general, some boundary conditions must be dropped when $N \to \infty$ and/or $\varepsilon \to 0$; we avoid this difficulty requiring that $\bar{F}_0$, $\bar{F}_1$, $\bar{K}_0$ are analytic for $z \in [0, 1]$.

Though we have an explicit expression for $\bar{F}_0$, $\bar{F}_1$, and $\bar{K}_0$, it is rather difficult to obtain the inverse Laplace transform of (3.24)–(3.26); we may then look for an approximate expression for the solutions, written as an asymptotic series in the parameter $\alpha = 1 - \rho$, assuming $\alpha \ll 1$. Thus, stretching z in such a way that $z = 1 - \alpha w$, and rescaling t so that $t = T/\alpha$, and consequently setting $s = \alpha\kappa$, let us rewrite $\bar{F}_0(z)$, $\bar{F}_1(z)$, and $\bar{K}_0(z)$ in the following way:

$$\bar{F}_0 = \frac{1}{\alpha}\Phi_0(w, \alpha), \qquad \bar{F}_1 = \frac{\rho}{\alpha^3}\Phi_1(w, \alpha),$$

$$\bar{K}_0 = \frac{\rho^2}{\alpha^3}\Omega_0(w, \alpha).$$

Expanding Φ_0, Φ_1, and Ω_0 in power series of α

$$\Phi_0 \sim \Phi_{00} + \alpha\Phi_{01}, \quad \Phi_1 \sim \Phi_{10} + \alpha\Phi_{11}, \quad \Omega_0 \sim \Omega_{00} + \alpha\Omega_{01},$$

we rewrite (3.19), (3.20), and (3.21) in terms of κ, w, and α. The corresponding solutions for Ω_0 and Φ_0, bounded for $w \to \infty$, are [11]

$$\Phi_{00} = \int_0^{\infty} \frac{\exp[-(1+w)\eta]}{(1+\kappa\eta)} \, d\eta, \tag{3.27}$$

$$\Omega_{00} = -\frac{1}{2}\int_0^{\infty} \frac{\eta^2 \exp[-(1+w)\eta]}{(1+\kappa\eta)} \, d\eta. \tag{3.28}$$

Here we focus on the equation for $\Phi_1(w, \alpha)$:

$$\begin{aligned}&\{(1-\alpha w)[1+\alpha(w+\kappa)-\alpha^2 w]-1\}(1-\alpha)\Phi_{1w}\\&\qquad+\alpha(1-\alpha)[\alpha^2 w-\alpha(w+\kappa)-1]\Phi_1\\&\qquad=(w+\kappa-\alpha w)[\alpha^4\Phi_0-3\alpha^3(1-\alpha w)\Phi_{0w}+\alpha^2(1-\alpha w)^2\Phi_{0ww}]\\&\qquad-\frac{\alpha(1-\alpha)(1-\alpha w)}{(w+1-\alpha w)^3},\end{aligned} \tag{3.29}$$

where $\Phi_{iw}=-\alpha\Phi_{iz}$, $\Phi_{iww}=\alpha^2\Phi_{izz}$, $i=0, 1$.

Substituting the expansions of Φ_0 and Φ_1 in (3.29), we obtain at the zeroth order in α:

$$\Phi_{10}-\kappa\Phi_{10w}=\frac{1}{(1+w)^3}=\frac{1}{2}\int_0^\infty \eta^2 \exp[-(1+w)\eta]\,d\eta. \tag{3.30}$$

If we require the solution of (3.30) to be bounded for $w\to\infty$, we obtain

$$\Phi_{10}=\frac{1}{2}\int_0^\infty \frac{\eta^2\exp[-(1+w)\eta]}{(1+w\kappa)}\,d\eta=-\Omega_{00}. \tag{3.31}$$

Taking into account (3.14), (3.15), and (2.35), and noting that

$$\overline{P_\varepsilon}(\alpha\kappa)=\frac{1}{\alpha}\int_0^\infty e^{-\kappa T}\mathbb{P}\left\{w>\frac{T}{\alpha}\,\middle|\,\varepsilon\geqq 0\right\}dT,$$

we arrive at the following relation:

$$\begin{aligned}\overline{P_\varepsilon}(\alpha\kappa)&\sim\frac{1}{\kappa}\left\{1-\left[\Phi_0(w=0)+\frac{\varepsilon\rho}{\alpha^2}\Phi_1(w=0)+\frac{\rho^2}{N\alpha^2}\Omega_0(w=0)\right]\right.\\&\qquad\left.\cdot\left(1+\frac{\rho^2}{N\alpha^2}\right)\left[1-\frac{\varepsilon}{G(N,1,0)}[2G_z(N,z,0)+G_{zz}(N,z,0)]_{z=1}\right]\right\}\\&\sim\frac{1}{\kappa}\left\{\left(1+\frac{\rho^2}{N\alpha^2}-\frac{\varepsilon\rho}{\alpha^2}\right)(1-\Phi_0)+\frac{\varepsilon\rho}{\alpha^2}(1-\Phi_1)-\frac{\rho^2}{N\alpha^2}(1+\Omega_0)\right\}_{w=0}.\end{aligned} \tag{3.32}$$

Setting $(1-\Phi_1(w=0))\sim(1-\Phi_{10}(w=0))$, we note that

$$(1-\Phi_1)\sim(1+\Omega_0).$$

Thus

$$\begin{aligned}\overline{P_\varepsilon}(\alpha\kappa)&\sim\frac{1}{\kappa}\left(1+\frac{\rho^2}{N\alpha^2}-\frac{\varepsilon\rho}{\alpha^2}\right)(1-\Phi_0)_{w=0}\\&\qquad+\frac{\rho}{\kappa\alpha^2}\left(\varepsilon-\frac{\rho}{N}\right)(1+\Omega_0)_{w=0}.\end{aligned} \tag{3.33}$$

The inverse Laplace transform of $\overline{P_\varepsilon}(\alpha\kappa)$ is (see [11])

$$\begin{aligned}P_\varepsilon\left(\frac{T}{\alpha}\right)&\sim\left(1+\frac{\rho^2}{N\alpha^2}-\frac{\varepsilon\rho}{\alpha^2}\right)P\left(\frac{T}{\alpha}\right)\\&\qquad+\frac{\rho}{\alpha^2}\left(\varepsilon-\frac{\rho}{N}\right)T^{3/2}K_3(2\sqrt{T}),\end{aligned} \tag{3.34}$$

where $P(t)$ is the waiting time distribution for the open system $(N\to\infty)$ in the zero-switching time model [10, eq. (2.20)], and K_3 is the modified Bessel function of third kind or the McDonald function [14, p. 78].

We remark that setting $\alpha = 1 - \rho \ll 1$ and $T/\alpha = t$ in (3.34), we obtain (1.6) for $P_\varepsilon(t)$, which is valid for $N\alpha^2 \gg 1$, $\varepsilon \ll \alpha^2$.

Since $\lim_{t\to 0} P(t) = 1$, and $\lim_{T\to 0} T^{3/2}K_3(2\sqrt{T}) = 1$, from (3.34) we have that

$$\lim_{T\to 0} P_\varepsilon\left(\frac{T}{\alpha}\right) = 1.$$

Furthermore,

$$\lim_{T\to\infty} P\left(\frac{T}{\alpha}\right) = \lim_{T\to\infty} T^{3/2}K_3(2\sqrt{T}) = \lim_{T\to\infty} P_\varepsilon\left(\frac{T}{\alpha}\right) = 0.$$

In the zero-switching time case we have

$$P_0\left(\frac{T}{\alpha}\right) \sim P\left(\frac{T}{\alpha}\right) + \frac{\rho^2}{N\alpha^2}\left[P\left(\frac{T}{\alpha}\right) - T^{3/2}K_3(2\sqrt{T})\right], \tag{3.35}$$

that is, (3.34) of [11]. Since $P(T/\alpha) > P_0(T/\alpha)$ for every $T > 0$, we have that the second term in the right-hand side of (3.35) is negative for $0 < t < \infty$. Then we may conclude that

$$P\left(\frac{T}{\alpha}\right) < T^{3/2}K_3(2\sqrt{T}).$$

Since (3.34) may be rewritten as

$$P_\varepsilon\left(\frac{T}{\alpha}\right) \sim P\left(\frac{T}{\alpha}\right) + \left(\frac{\rho^2}{N\alpha^2} - \frac{\varepsilon\rho}{\alpha^2}\right)\left[P\left(\frac{T}{\alpha}\right) - T^{3/2}K_3(2\sqrt{T})\right], \tag{3.36}$$

the effect of the presence of ε is an increase of the waiting time distribution, as will be shown in Fig. 3; that is, if $\varepsilon_1 < \varepsilon_2$, then $P_{\varepsilon_1} < P_{\varepsilon_2}$. It is remarkable to observe that, as $T \to 0$, $P_\varepsilon(T/\alpha) \to 1$.

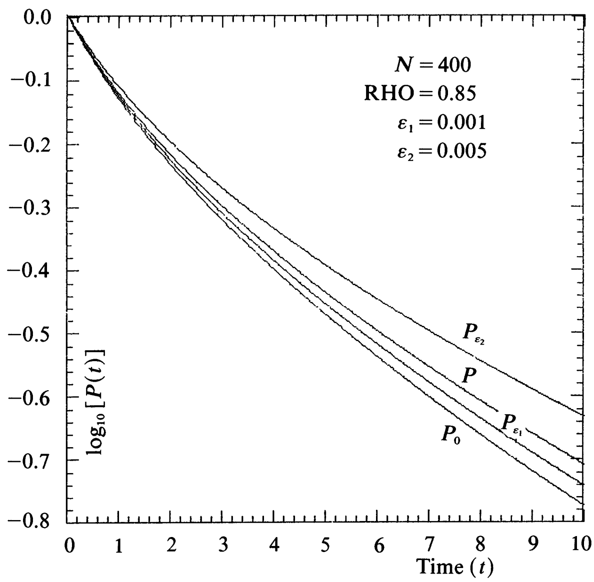

FIG. 3. *Some approximate expressions of $P_\varepsilon(t)$. P is the exact expression of $\mathbb{P}\{w > t\}$ in the open system [10, eq. (2.20)] that we consider here as the zeroth approximation of $P_0(t)$ in the closed system with $\varepsilon = 0$. P_0, given by (3.35) is the corrected value of P to first order in $1/N$ when $\varepsilon = 0$. P_{ε_1} and P_{ε_2}, given by (3.34), are the corrected values of P_0 to first order in ε for $\varepsilon = \varepsilon_1$ and $\varepsilon = \varepsilon_2$, respectively.*

TABLE 1

The dependence of $\mathcal{G}(N, E)$ on N. The exact values of $\mathcal{G}(N, \varepsilon)$, its approximate expression $\mathcal{G}(N, \varepsilon)_{as}$, and their ratio are reported for different values of N. We observe that $\varepsilon \ll \bar{\varepsilon}(N)$ and that if $\varepsilon \ll \alpha^2$, the approximate value $\mathcal{G}(N, \varepsilon)_{as}$ becomes more and more satisfactory for increasing values of N.

$\varepsilon = 0.001$; $\alpha = 0.25$

N	$\mathcal{G}(N, \varepsilon)$	$\mathcal{G}(N, \varepsilon)_{as}$	$\mathcal{G}(N, \varepsilon)_{as}/\mathcal{G}(N, \varepsilon)$
20	3.1791	2.2480	0.7071
50	3.5522	3.3280	0.9369
80	3.6951	3.5980	0.9737
100	3.7512	3.6880	0.9832
200	3.8816	3.8680	0.9965
400	3.9594	3.9580	0.9997
500	3.9763	3.9760	0.9999
1000	4.0117	4.0120	1.0006

TABLE 2

The dependence of $\mathcal{G}(N, \varepsilon)$ on ε. The first row ($\varepsilon = 0$) represents the correction to first order in $1/N$ to the zero-switching time model. This correction is satisfactory as long as $N\alpha^2 \gg 1$. When $\varepsilon \neq 0$ the first-order corrections in $1/N$ and ε are satisfactory if $N\alpha^2 \gg 1$, $\varepsilon \ll \alpha^2$, and $\varepsilon \ll \bar{\varepsilon}$.

$N = 100$; $\alpha = 0.15$; $\bar{\varepsilon} \simeq 0.022$

ε	$\mathcal{G}(N, \varepsilon)$	$\mathcal{G}(N, \varepsilon)_{as}$	$\mathcal{G}(N, \varepsilon)_{as}/\mathcal{G}(N, \varepsilon)$
0	5.4806	4.5259	0.8258
0.0001	5.4895	4.5511	0.8291
0.001	5.5726	4.7778	0.8574
0.008	6.4375	6.5407	1.0160
0.01	6.8004	7.0444	1.0359
0.015	8.2602	8.3037	1.0053
0.02	12.6892	9.5630	0.7536
0.05	5.05×10^{10}	17.1200	3.39×10^{-10}

TABLE 3

The dependence of $E(W)$ on N. $E(W)$ is monotonically increasing in N, since in this table $\varepsilon \ll \bar{\varepsilon}$; when N is very large ($N\alpha^2 \gg 1$) and ε is sufficiently small ($\varepsilon \ll \alpha^2$), $E(W)_{as}$ is a satisfactory approximation of $E(W)$.

$\varepsilon = 0.001$; $\alpha = 0.25$

N	$E(W)$	$E(W)_{as}$	$E(W)_{as}/E(W)$
40	2.8521	2.3120	0.8106
50	2.9691	2.6720	0.8999
80	3.1906	3.2120	1.0067
100	3.2825	3.3920	1.0334
200	3.5097	3.7520	1.0690
400	3.6567	3.9320	1.0753
500	3.6901	3.9680	1.0753
1000	3.7620	4.0400	1.0739

In other words, even if ε multiplies the higher-order term in (3.11), no singular perturbation effect ("boundary layer") [13] appears in (3.34).

4. Some numerical results. In Tables 1 and 2 we present the numerical values of $\mathcal{G}(N, \varepsilon)$ given by (2.3), its asymptotic estimate—denoted by $\mathcal{G}(N, \varepsilon)_{as}$—given by (2.34), and their ratio. In particular, in Table 1 we show the dependence of the previously mentioned quantities on N for ρ and ε fixed, and in Table 2 we show their dependence on ε for ρ and N fixed.

In Tables 3 and 4 we present the numerical values of $E(W)$ given by (2.12), its asymptotic estimate denoted by $E(W)_{as}$ given by (1.4), and their ratio. In particular, in Table 3 we show the dependence of the previously mentioned quantities on N for ρ and ε fixed, and in Table 4 we show their dependence on ε for ρ and N fixed.

TABLE 4

The dependence of $E(W)$ on ε. In a closed system in the zero-switching time model ($\varepsilon = 0$) the mean waiting time $E(W)$ is smaller than in the open system, since we have formally taken the limit $N \to \infty$. When we consider a nonzero ε, $E(W)$ is an increasing function of ε. Too large (and unrealistic) values of ε correspond to large values of $E(W)$ (second column). $E(W)_{as}$ is a satisfactory approximation of $E(W)$ only for $\varepsilon \ll \alpha^2$ and $\varepsilon \ll \bar{\varepsilon}$.

$N = 100$; $\alpha = 0.25$; $\bar{\varepsilon} = 0.029$

ε	$E(W)$	$E(W)_{as}$	$E(W)_{as}/E(W)$
0	3.2389	3.2800	1.0127
0.001	3.2825	3.3920	1.0334
0.005	3.4796	3.8400	1.1036
0.01	3.7999	4.4000	1.1579
0.02	5.1066	5.5200	1.0809
0.03	18.5936	6.6400	0.3571
0.04	61.4090	7.7600	0.1264
0.15	91.8775	20.080	0.2186

This quantitative study of $\mathcal{G}(N, \varepsilon)$ and $\mathcal{G}(N, \varepsilon)_{as}$ and of $E(W)$ and $E(W)_{as}$ shows that for suitable small values of ε ($\varepsilon \ll \alpha^2$) and for N relatively large ($N\alpha^2 \gg 1$) the asymptotic expressions of $\mathcal{G}(N, \varepsilon)$ and $E(W)$, given by $\mathcal{G}(N, \varepsilon)_{as}$ (2.32) and $E(W)_{as}$ (1.4), respectively, are good approximations of their exact values, that is, formulae (2.3) and (2.12).

Finally, in Fig. 3 on the x-axis we report the time t and on the y-axis we report the logarithm in base 10 of several approximate expressions of $P_\varepsilon(t)$. As expected, if $\varepsilon_1 < \varepsilon_2$, we have $P_{\varepsilon_1}(t) < P_{\varepsilon_2}(t)$ for every t.

Acknowledgments. The authors are grateful to Professor Francesco Zirilli for his valuable and helpful suggestions, Dr. Claudio Benedetti for some useful discussions, and Dr. Pasquale Pavone for suggesting the problem.

REFERENCES

[1] C. BENEDETTI, *Waiting time distributions for two realistic models of a shared processor system*, Istituto Nazionale di Alta Matematica, Rome, Italy, preprint.

[2] H. HEFFES, *Moment formulae for a class of mixed multi-job-type queueing networks*, Bell System Tech. J., 61 (1982), pp. 709–745.

[3] L. KLEINROCK, *Queueing Systems*, Vols. I, II, John Wiley, New York, 1975.

[4] C. KNESSL, B. J. MATKOWSKY, Z. SCHUSS, AND C. TIER, *Asymptotic analysis of a state-dependent* M/G/1 *queueing system*, SIAM J. Appl. Math., 46 (1986), pp. 483-505.

[5] ———, *On the performance of state-dependent single server queues*, SIAM J. Appl. Math., 46 (1986), pp. 657-697.

[6] D. MITRA, *Waiting time distributions from closed queueing network models of shared-processor systems*, Performance '81, F. J. Kylstra ed., North-Holland, Amsterdam, pp. 112-129.

[7] D. MITRA AND J. MCKENNA, *Asymptotic expansions for closed markovian networks with state-dependent service rates*, J. Assoc. Comput. Mach., 33 (1986), pp. 568-592.

[8] D. MITRA AND J. A. MORRISON, *Asymptotic expansions of moments of the waiting time in a shared-processor of an interactive system*, SIAM J. Comput., 12 (1983), pp. 789-802.

[9] ———, *Asymptotic expansions of moments of the waiting time in closed and open processor-sharing systems with multiple job classes*, Adv. Appl. Probab., 15 (1983), pp. 813-839.

[10] J. A. MORRISON, *Response-time distribution for a processor-sharing system*, SIAM J. Appl. Math., 45 (1985), pp. 152-167.

[11] ———, *Asymptotic analysis of the waiting-time distribution for a large closed processor-sharing system*, SIAM J. Appl. Math., 46 (1986), pp. 140-170.

[12] J. A. MORRISON AND D. MITRA, *Heavy-usage asymptotic expansions for the waiting time in closed processor-sharing systems with multiple classes*, Adv. Appl. Probab., 17 (1985), pp. 63-85.

[13] A. H. NAYFEH, *Perturbation Methods*, John Wiley, New York, 1973.

[14] G. N. WATSON, *Theory of Bessel Functions*, Cambridge University Press, Cambridge, England, 1966.