THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 60, No. 5, May–June 1981
Printed in U.S.A.

A Class of Closed Markovian Queuing Networks: Integral Representations, Asymptotic Expansions, and Generalizations*

By J. McKENNA, D. MITRA, and K. G. RAMAKRISHNAN

(Manuscript received October 28, 1980)

Closed Markovian networks of queues that have the product form in their stationary probability distributions are useful in the performance evaluation and design of computer and telecommunication systems. Therefore, the efficient computation of the partition function—the key element of the solution in product form—has attracted considerable effort. We present a new and broadly applicable method for calculating the partition function. This method can be applied to very large networks, which were previously computationally intractable. Most of the paper details applications of this approach to a network class which arose in modeling an interactive processor. We show that the partition function and derivatives such as mean values (response times, CPU utilizations, etc.) may be represented by integrals and their ratios. The integrands contain a parameter N which is large for large networks. Next, the classical techniques of asymptotic analysis are applied to derive three main power series expansions in descending powers of N to correspond to normal, high, and very high usage. This work emphasizes multiple terms in the expansions for precision and error analyses.

I. INTRODUCTION

The theoretical results on the product form of the stationary distributions of large classes of Markovian queuing networks continue to have a profound influence on computer communications, computer systems analysis, and traffic theory.[1–4] These results make at least feasible the analysis and synthesis of the large systems of ever increasing complexity being considered in these areas. The subclass of *closed*

* Presented at ORSA-TIMS meeting, Jan. 5–7, 1981, at Boca Raton, Fla.

networks of queues is more difficult to analyze than the open networks because there is no stationary independence of the network nodes. However, the incentive for investigating the closed networks does exist since they have been used to model multiple-resource computer systems,[2,5] multiprogrammed computer systems,[6–8] time-sharing,[2] and window flow control in computer communication networks;[9,10] networks with external inputs subject to blocking require the analysis of a large number of closed networks.[11,12] The closed network model that we shall use for illustrative purposes arose in the modeling of a central processor in a node of a computer network. This network is subject to a variety of processing demands. In recognition of the utility of closed networks, considerable research and commercial interest has been directed towards developing efficient procedures for computing the partition function (the normalizing constant), the only element of the product form solution requiring significant computation.[13–18]

However, as these existing recursive techniques are applied to the problems of particular interest in the Bell System, wherein the constituents of the closed chains are many and the number of chains are many, their shortcomings are observed to be severe in the amount of computing time and memory required and the accuracy attained. A more detailed account appears in Section 2.4 and Section IX. Briefly, the existing recursive techniques are largely ineffectual.

We present a new way to view the problem, which surmounts many of the difficulties associated with large networks. The approach is broadly applicable—as indicated in Section X—even though the paper is a detailed account of applications to a specific class of closed networks. The new approach consists, first, of recognizing that the partition function may be written as an integral with a large parameter N present in the integrand to reflect the large size of the network. Next, the classical techniques of asymptotic analysis are applied to derive an asymptotic power series, typically in descending powers of N. The integrand will have fundamentally different properties in different ranges of the system parameters and this will require correspondingly different expansions. Thus, in this paper we develop three separate series expansions—Proposition 3 (Section IV), Proposition 12 (Section VII), Proposition 17 (Section VIII)—each corresponding to a specific range of values of the usage parameter α. It is worth emphasizing that, commensurate with an objective of providing solutions with any desired accuracy, we give procedures for generating multiple terms in the asymptotic expansions, not just the dominant term. In Section VIII, we unify the preceding results by giving a common expansion that holds uniformly in the system parameters. The uniform expansions introduce in a natural way the parabolic cylinder (or Weber) functions, a classical family of special functions with many antecedents and ties with other special functions.[13] Besides duplicating the specialized expansions derived earlier, the uniform expansion makes available for use the many well-documented and tested expansions that are known in connection with parabolic cylinder functions.

Section IX describes a user-oriented software package that has been written in C-language to implement the approach developed here. We supply results obtained by the package on four test problems that arose in analyzing performance of a Bell System project. Also reported are the results of a comparison with a well known, commercially marketed package that obtains solutions recursively. Our package is able to solve the large problems, which are well beyond the range of the other package, and, surprisingly, solve the small problems as well with errors that have small bounds.

Section X provides the basis for extending the approach developed here to quite general multiprocessor, multidiscipline queuing networks. We show that for most networks that have been shown to have the product form in their stationary probability distribution, the partition function has an integral representation. The expansions appropriate for its computation are not considered here.

Not surprisingly, the new representation of the partition function as an integral—the starting point of our computational procedures—may be exploited anew to derive analytical estimates and bounds of the quantities of interest, such as throughput, mean response time, etc. We demonstrate particularly in Section 5.3 that these formulas explicitly exhibit the system parameters and as such are rather useful as design and synthesis aids. (The bounds are also useful as checks on the computational procedure.) Purely computational procedures by themselves do not yield this particular form of insight into system behavior.

The asymptotic sequences used typically are power series in N^{-1}, where recall N is the generic large parameter.[14] Thus, the number of terms required to achieve the desired accuracy decreases with increasing N. In contrast, with recursive solutions the computational complexity grows with the network size. Also, the asymptotic methods handle increased numbers of *classes* of constituents with little incremental difficulty, while the computational complexity in recursive methods grows geometrically. Thus, the contrasting techniques are not replacements for each other but complementary: loosely speaking, the recursions are most effective for smaller networks, while the asymptotic expansions are most effective for large networks.

The contrasting behavior with respect to a large number of classes is of particular importance in computer communications where, as Reiser,[9] Schwartz,[3] and others have pointed out, traffic corresponding to each source-destination pair is treated as a separate class and the network closure follows from the windows employed in flow control. Reiser has developed heuristics to cope with this situation.

References 11, 15, and 16 also contain results pertaining to large networks.

II. NETWORK MODEL AND KNOWN RESULTS

2.1 Model

In the model (see Fig. 1) each constituent, which may be thought of as a terminal or station, of the closed network spends alternating periods of time in the two nodes that constitute the network—the 'think' node (also node 1) and the 'CPU' node (node 2). The think time in each cycle for each constituent is an independent random variable with an exponential distribution. The time spent in the CPU node depends on many factors since it is here that there occurs interaction between constituents being serviced. We stipulate that the CPU discipline is 'processor sharing'* and that the desired service time (i.e., the time required to service the job if the entire CPU was dedicated to the job) is an independent exponentially distributed random variable.[1,4] The 'think' node is thus an ∞-server center and in the terminology of Ref. 4, nodes 1 and 2 are respectively Type 3 and 2 centers.

We stipulate that there is sufficient statistical inhomogeneity amongst the constituents to justify the existence of several, say p, classes of constituents, with class i having K_i constituents, $1 \le i \le p$. Statistical homogeneity applies within a class in that ρ_{i1} and ρ_{i2} will respectively denote the mean think time and the mean desired service time that are common to all in class i. The variations among these mean values may be quite substantial.

Our involvement with this model arose while modeling behavior of traffic through a processing node of a computer network. The number of classes of constituents is at least five, namely, time sharing; inquiry/response and data-base query; batch and remote job entry; messages and broadcast; data entry/collect and screen type jobs. The mean values $\{\rho_{ij}\}$ are obtained from benchmark measurements. In another variation of this problem, a finer classification of constituents was considered. Our interest is in cases where the individual class populations $\{K_i\}$ extend to several hundred, while the number of classes is of the order of ten.

* In the processor-sharing discipline there is no overt queuing because all, say n, jobs present in the node simultaneously receive service at $1/n$ times the rate given to a single job by the processor. Thus, the rate given to a single specific constituent fluctuates with time. This discipline is the limiting case of the round robin discipline as the time quantum given to each job becomes arbitrarily small.

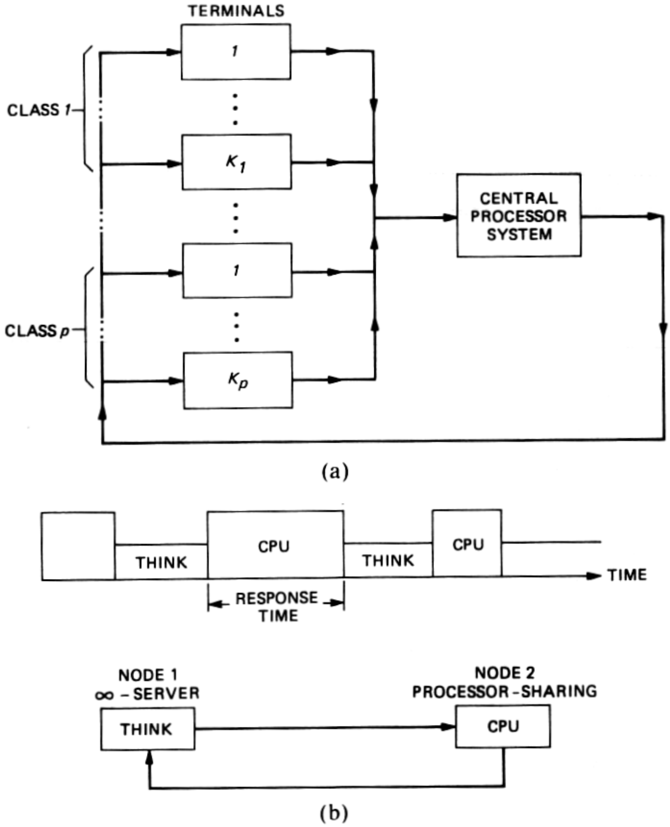

Fig. 1—(a) There are p classes of constituents—shown as terminals—with K_j constituents in class j. (b) Constituents spend alternate periods of time in the think node and processor-sharing CPU node.

2.2 Product form solution

If N_{ij} is the stationary random variable denoting the number of constituents of class i in node j, and if π is the following stationary probability $\pi(n_{11}, \cdots, n_{p1}; n_{12}, \cdots, n_{p2}) = Pr[N_{11} = n_{11}, \cdots, N_{p1} = n_{p1}; N_{12} = n_{12}, \cdots, N_{p2} = n_{p2}]$, then it is known that with the left hand side abbreviated to $\pi(\mathbf{n}_1, \mathbf{n}_2)$,[4]

$$\pi(\mathbf{n}_1, \mathbf{n}_2) = \frac{1}{G(\mathbf{K})} \left(\prod_{i=1}^{p} \frac{\rho_{i1}^{n_{i1}}}{n_{i1}!} \right) \left(\sum_{j=1}^{p} n_{j2} \right)! \left(\prod_{k=1}^{p} \frac{\rho_{k2}^{n_{k2}}}{n_{k2}!} \right), \tag{1}$$

where $G(\mathbf{K}) = G(K_1, K_2, \cdots, K_p)$ is the normalization constant so chosen as to make the sum of all quantities in (1) equal to 1. Explicitly,

$$G(\mathbf{K}) = \sum_{m_p=0}^{K_p} \cdots \sum_{m_1=0}^{K_1} \left(\prod_{i=1}^{p} \frac{\rho_{i1}^{(K_i - m_i)}}{(K_i - m_i)!} \right) \left(\sum_{j=1}^{p} m_j \right)! \left(\prod_{k=1}^{p} \frac{\rho_{k2}^{m_k}}{m_k!} \right). \tag{2}$$

The function $G(\cdot)$ defined on the integer lattice in $\mathbf{R}^p$ is referred to as the partition function.*

2.3 System performance

A number of interrelated system performance measures are obtained from the partition function. We start with $\overline{N_{i1}(\mathbf{K})}$, the mean number of constituents of class i in node 1 ('think'), and obtain directly from (2),

$$\overline{N_{i1}(\mathbf{K})} = \rho_{i1} G(\mathbf{K} - \mathbf{e}_i)/G(\mathbf{K}), \tag{3}$$

where $\mathbf{e}_i$ is our notation for the vector with the i^{th} component unity and all other components zero. Thus, $G(\mathbf{K} - \mathbf{e}_i)$ is the partition function associated with a new population with one less constituent in the i^{th} class. From (3) and Little's theorem applied to class i and node 1, we obtain for the throughput of constituents of class i,

$$\lambda_i(\mathbf{K}) = G(\mathbf{K} - \mathbf{e}_i)/G(\mathbf{K}). \tag{4}$$

The mean response time, i.e., time spent in the CPU in each cycle by class i constituents, is obtained again from an application of Little's theorem:

$$t_i(\mathbf{K}) = K_i G(\mathbf{K})/G(\mathbf{K} - \mathbf{e}_i) - \rho_{i1}. \tag{5}$$

Finally, the utilization of the CPU by constituents of class i,

$$u_i(\mathbf{K}) \triangleq \sum \cdots \sum \frac{n_{i2}}{\sum n_{j2}} \pi(\mathbf{n}_1, \mathbf{n}_2) = \rho_{i2} G(\mathbf{K} - \mathbf{e}_i)/G(\mathbf{K}). \tag{6}$$

The important point to note is that all the mean values given in (3) through (6) are simply obtained from the knowledge of the partition function estimated at two neighboring points of the integer lattice in $\mathbf{R}^p$. As the above quantities are all closely related, we shall henceforth consider only the last, $\{u_i(\mathbf{K})\}$.

Higher moments of N_{i1}, the random number of constituents of class i in node 1, may also be obtained from knowledge of the partition function:

$$\overline{N_{i1}^2(\mathbf{K})} = \{\rho_{i1}^2 G(\mathbf{K} - 2\mathbf{e}_i) + \rho_{i1} G(\mathbf{K} - \mathbf{e}_i)\}/G(\mathbf{K}), \tag{7}$$

and, for $i \neq j$,

$$\overline{N_{i1}N_{j1}(\mathbf{K})} = \rho_{i1}\rho_{j1} G(\mathbf{K} - \mathbf{e}_i - \mathbf{e}_j)/G(\mathbf{K}). \tag{8}$$

Of course, the moments of $\{N_{i2}\}$ are easily derived from moments of $\{N_{i1}\}$.

---

* The distribution in (1) and (2) is also the stationary distribution of other networks. For example, if node 2 is first-come-first-served with class independent service rate $1/\rho_2$, then (1) and (2) with $\rho_{i2} \equiv \rho_2$ is the solution.

2.4 Recursive solutions

The above results explain why the problem of efficiently computing the partition function has excited so many researchers.[15,17–21] For the problem at hand it is easy to arrive at the following recursion by established techniques:

$$G(\mathbf{K}) = \sum_{j=1}^{p} \rho_{j2} G(\mathbf{K} - \mathbf{e}_j) + \prod_{j=1}^{p} \frac{\rho_{j1}^{K_j}}{K_j!}. \tag{9}$$

The boundary conditions are: $G(\mathbf{K}) = 0$ for $\mathbf{K} \not\geq 0$.

Observe that the partition functions themselves can be scaled on account of the linearity and the fact that only ratios of partition function values have physical content. By the same token, implementations of (9), for large $\mathbf{K}$, typically give rise to values which are either very small or very large leading to rather severe problems of overflow and underflow. Proper scaling is only marginally helpful.

The main problems with implementing (9) are with respect to time, memory required during computation, and accuracy. A straightforward application of (9) would require an estimated K^p iterations, where K is the generic class size. Similarly, the storage required would be approximately K^{p-1}. Now these crude estimates can clearly be improved upon by simply pruning or avoiding the computation of intermediate lattice points, but this would be at the cost of increased algorithmic complexity, and the extent of the accrued benefits are not generally known. The underflow/overflow phenomenon that affects the accuracy of the scheme has already been commented on; a no less severe problem is accumulation of round-off errors in a large number of iterations.

The recursive solution in (9) is one of several that can be generated by recently discovered techniques. However, all solutions that we are aware of are recursive and share to varying degrees the three broad categories of limitations just discussed.

III. INTEGRAL REPRESENTATIONS

3.1 Partition function

We start with Euler's integral

$$n! = \int_0^\infty e^{-t} t^n dt. \tag{10}$$

Substituting for the middle term in braces in (2) we obtain*

* If the range of integer subscripts is not stipulated explicitly, then the range is understood to be $[1, p]$.

$$G(\mathbf{K}) = \int_0^\infty e^{-t} \sum_{m_p=0}^{K_p} \cdots \sum_{m_1=0}^{K_1} \left(\prod \frac{\rho_{j1}^{K_j - m_j}}{(K_j - m_j)!}\right)\left(\prod \frac{(t\rho_{j2})^{m_j}}{m_j!}\right) dt$$

$$= \frac{1}{\{\prod K_j!\}} \int_0^\infty e^{-t} \sum_{m_p=0}^{K_p} \cdots \sum_{m_1=0}^{K_1} \prod \binom{K_j}{m_j} \rho_{j1}^{K_j - m_j} (t\rho_{j2})^{m_j} dt$$

$$= \frac{1}{\{\prod K_j!\}} \int_0^\infty e^{-t} \prod (\rho_{j1} + t\rho_{j2})^{K_j} dt. \tag{11}$$

To obtain our final form for the partition function, let N be the generic large parameter to be associated with a large network. Also, for $j = 1, 2, \cdots p$, let

$$K_j = \beta_j N, \tag{12}$$

$$r_j \triangleq \left.\frac{\text{mean think time}}{\text{mean service time}}\right|_{\text{class } j} = \frac{\rho_{j1}}{\rho_{j2}}$$

$$= \gamma_j N. \tag{13}$$

The suggestion in this notation is that $\{\beta_j\}$ and $\{\gamma_j\}$ are 0(1), which is the situation to which our work is primarily directed.

There is considerable latitude in selecting N. The following choice is certainly not essential but as it does ease some of the manipulations, we use it throughout the paper:

$$N = (\prod r_i^{K_i})^{1/\Sigma K_j}. \tag{14}$$

An implication of this choice of N is that

$$\sum \beta_j \log \gamma_j = 0. \tag{15}$$

We return to (11) to observe

$$G(\mathbf{K}) = \left(\prod \frac{\rho_{j2}^{K_j}}{K_j!}\right) \int_0^\infty e^{-t} \prod (r_j + t)^{K_j} dt$$

$$= \left(\prod \frac{\rho_{j2}^{K_j}}{K_j!}\right) N^{\Sigma K_j + 1} \int_0^\infty e^{-Nz} \prod (\gamma_j + z)^{\beta_j N} dz. \tag{16}$$

Finally, we have after using (14),

Proposition 1:

$$G(\mathbf{K}) = \left(N \prod \frac{\rho_{j1}^{K_j}}{K_j!}\right) \int_0^\infty e^{-Nf(z)} dz, \tag{17}$$

$$\text{where } f(z) \triangleq z - \sum \beta_j \log(\gamma_j + z). \quad \square \tag{18}$$

We shall see that typically there is no need to compute the term in braces in (17).

3.2 Representations of mean values and some higher moments

We have two options concerning representations of the physically interesting quantities in (3) through (7). Concerning ourselves with $u_i(\mathbf{K})$ given in (6), we may simply use the respective integral representations for $G(\mathbf{K}-\mathbf{e}_i)$ and $G(\mathbf{K})$ and obtain for $i = 1, 2, \cdots, p$,

$$u_i(\mathbf{K}) = \frac{\beta_i}{\gamma_i} \frac{\hat{N} \int_0^\infty e^{-\hat{N}\hat{f}(z)} dz}{N \int_0^\infty e^{-Nf(z)} dz}, \tag{19}$$

where $\hat{N}$ and $\hat{f}(\cdot)$ are defined analogously to N and $f(\cdot)$ but for a network with one less constituent in the i^{th} class. Obviously N and $\hat{N}$ as well as $f(\cdot)$ and $\hat{f}(\cdot)$ are going to be close to each other, but the above option takes no further notice of this fact.

In the contrasting option, we proceed as follows. Observe that

$$u_i(\mathbf{K}+\mathbf{e}_i)^{-1} = \frac{1}{\rho_{i2}} \frac{G(\mathbf{K}+\mathbf{e}_i)}{G(\mathbf{K})}. \tag{20}$$

Now, from (16),

$$\begin{aligned} G(\mathbf{K}+\mathbf{e}_i) &= \frac{\rho_{i2}}{K_i+1}\left(\prod \frac{\rho_{j2}^{K_j}}{K_j!}\right) \int_0^\infty (r_i+t)e^{-t} \prod_j (r_j+t)^{K_j} dt \\ &= \frac{\rho_{i2}}{K_i+1}\left(N^2 \prod \frac{\rho_{j1}^{K_j}}{K_j!}\right) \int_0^\infty (\gamma_i+z)e^{-Nf(z)} dz, \end{aligned} \tag{21}$$

where to obtain the last equation we have proceeded just as we did from (16) to (17). The point to note is that in the above expression N and $f(\cdot)$ are identical to that used in the expression for $G(\mathbf{K})$ in (17) and (18). Finally, from (20), for $i = 1, 2, \cdots, p$

Proposition 2:

$$u_i(\mathbf{K}+\mathbf{e}_i)^{-1} = \frac{1}{\beta_i + 1/N}\left(\gamma_i + \frac{\int_0^\infty z e^{-Nf(z)} dz}{\int_0^\infty e^{-Nf(z)} dz}\right). \quad \square \tag{22}$$

Notice that the ratio of integrals, the only quantity requiring significant computation, is common to all classes.

The higher moments given in (7) and (8) have similar representations which are easy to derive. We give the form for one term that occurs in (7) which reveals the general pattern:

$$\frac{1}{\rho_{i1}^2}\frac{G(\mathbf{K}+2\mathbf{e}_i)}{G(\mathbf{K})}=\frac{1}{N^2}\frac{1}{\gamma_i^2(\beta_i+1/N)(\beta_i+2/N)} \cdot \left(\gamma_i^2+2\gamma_i\frac{\displaystyle\int_0^\infty ze^{-Nf(z)}dz}{\displaystyle\int_0^\infty e^{-Nf(z)}dz}+\frac{\displaystyle\int_0^\infty z^2e^{-Nf(z)}dz}{\displaystyle\int_0^\infty e^{-Nf(z)}dz}\right). \quad (23)$$

The feature to note is that the i^{th} moment involves integrals of the form $\int_0^\infty z^j e^{-Nf(z)}dz$, $j = 0, 1, \cdots, i$.

3.3 Properties of the function f

We shall later need to recall certain properties of the function $f(z)$ in (18), $z \geq 0$. As a consequence of (15),

$$f(0) = 0. \quad (24)$$

Note

$$\begin{aligned} f^{(i)}(z) &= 1-\sum \beta_j/(\gamma_j+z) \qquad \text{for } i=1 \\ &= (-1)^i(i-1)!\sum \beta_j/(\gamma_j+z)^i \qquad \text{for } i>1, \end{aligned} \quad (25)$$

and the following alternating sign property that holds for $i = 2, 3, \cdots$,

$$(-1)^i f^{(i)}(z) > 0, \qquad z \geq 0. \quad (26)$$

Also, for $i = 2, 3, \cdots$,

$$|f^{(i)}(z)| \leq |f^{(i)}(0)|, \qquad z \geq 0. \quad (27)$$

As the second derivative is positive in $[0, \infty)$, the function is always convex.

Let us view the function $f(z)$ in the interval $\gamma_m < z < \infty$, where $\gamma_m \triangleq -\min_i \gamma_i$ (see Fig. 2). As the figure indicates, the derivative of the function tends to ∞ at both ends of the interval. Coupled with the convexity and the other already established facts, it follows that the function has a unique stationary point, a minimum, in (γ_m, ∞). As in the figure, we let $\hat{z}$ denote this unique stationary point.

This stationary point may be obtained as the unique real solution in (γ_m, ∞) to the equation

$$\sum_{j=1}^{p}\frac{\beta_j}{\gamma_j+z}=1.$$

Thus, the largest real root of a polynomial of order p gives $\hat{z}$.

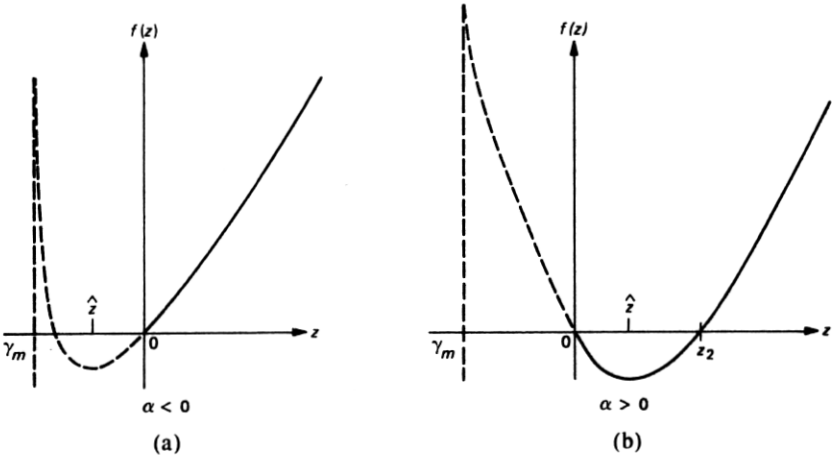

Fig. 2—Sketches of the function $f(z)$.

It will be important for us to distinguish the cases $\hat{z} < 0$ and $\hat{z} \geq 0$. The slope of the function $f(z)$ at the origin effectively indicates which case holds: the first if the slope is positive and the second otherwise. Thus, a key parameter of the system is

$$\alpha \triangleq -f^{(1)}(0) = \sum \beta_i/\gamma_i - 1$$
$$= \sum K_i/r_i - 1. \tag{28}$$

From the previous discussion we thus have that

$$\operatorname*{Min}_{z\geq 0} f(z) = f(0) \quad \text{if} \quad \alpha \leq 0$$
$$= f(\hat{z}) \quad \text{if} \quad \alpha > 0. \tag{29}$$

Equation (29) summarizes the background on the stationary point as needed for most of the paper. For example, $\hat{z}$ is needed only if $\alpha > 0$. However, Section VIII is exceptional in that, while considering small possibly negative α, the corresponding stationary point $\hat{z}$ is required to be known. Note that as $\alpha \to 0$, $\hat{z} \sim \alpha/\sum \beta_i/\gamma_i^2$.

The parameter α, $\alpha > -1$, is an indicator of the traffic intensity, with increasing α corresponding to higher traffic intensities. Our results, theoretical and numerical, show that $\alpha \geq 0$ corresponds to heavy usage corresponding to close to 100 percent utilization of the CPU. 'Normal' usage in large networks will certainly require $\alpha < 0$ and in all likelihood α will not be close to 0. For this reason, the most comprehensive results given here are for the case $\alpha < 0$.

IV. ASYMPTOTIC EXPANSIONS FOR NORMAL TRAFFIC

Throughout Section IV we shall consider $\alpha < 0$.

4.1 Laplace's method

We shall first apply Laplace's method[14,22,23] to obtain asymptotic expansions (see Appendix A for notation) for the integral in the representation of the partition function in (17), and subsequently to the other integrals in (22) and (23). Laplace's method observes that when N is large, the minimum of $Nf(z)$ at $z = 0$ is very sharp and that the dominant contribution to the integral comes from the neighborhood of 0.

In the integral

$$I = \int_0^\infty e^{-Nf(z)} dz, \tag{30}$$

let us change variables,

$$u \triangleq f(z), \tag{31}$$

so that

$$I = \int_0^\infty e^{-Nu} \left(\frac{dz}{du}\right) du. \tag{32}$$

To obtain dz/du, let us begin with the power series convergent in the neighborhood of 0,

$$u = f(z) = \sum_{j\geq 1} f_j z^j, \tag{33}$$

where $f_j = f^{(j)}(0)/j!$. Standard procedures allow the series to be reversed, i.e., give z as a power series in u. Appendix A elaborates on this procedure and explicitly gives the leading terms. Let the series obtained by this procedure be

$$z \triangleq g(u) = \sum_{j\geq 1} g_j u^j, \tag{34}$$

so that

$$\frac{dz}{du} = \sum_{j\geq 0} a_j u^j, \tag{35}$$

where $a_j = (j+1)g_{j+1}$. Substitution of this series in (32) yields

$$I \sim \sum_{j\geq 0} a_j \int_0^\infty e^{-Nu} u^j du$$

and thus

Proposition 3:

$$I \sim \sum_{j\geq 0} \frac{j! a_j}{N^{j+1}} \quad \text{as} \quad N \to \infty. \quad \square \tag{36}$$

The leading three terms of the series are obtained from

$$a_0 = 1/f_1; \qquad a_1 = -2f_2/f_1^3; \qquad a_2 = 3(2f_2^2 - f_1 f_3)/f_1^5. \tag{37}$$

The proof that (36) is an asymptotic expansion is available from various sources.[14, 22, 23] Indeed an explicit proof is available from the bounds that we develop in the following section in the course of an error analysis. Nonetheless, we sketch a proof based on Watson's lemma[24] applied to the integral in (32)—the lemma is anyhow used later. Watson's lemma considers the integral $\int_0^\infty e^{-Nu}h(u)du$ in which

(*i*) $h(u)$ has a convergent power series expansion in the neighborhood of the origin, and

(*ii*) there exist constants c_1, c_2 such that $|h(t)| < c_1 e^{c_2 t}$ for $t \geq 0$, and asserts that an asymptotic expansion for the integral is obtained by replacing $h(u)$ by its power series and integrating term by term. Thus, series reversion to obtain dz/du and a subsequent application of Watson's lemma gives the asymptotic expansion in (36).

The reader will note for future reference that the asymptotic expansion in (36) may also be written as

$$I \sim \sum_{j=1}^{\infty} g^{(j)}(0)/N^j, \tag{38}$$

where $g(\cdot) = f^{-1}(\cdot)$, as follows from (33) and (34).

Asymptotic expansions of integrals of the form $I^{(k)} = \int_0^\infty z^k e^{-Nf(z)}dz$ follow with only slight modifications. Thus, in lieu of (32) we now have

$$I^{(k)} = \int_0^\infty e^{-Nu}\left(z^k \frac{dz}{du}\right)du. \tag{39}$$

In principle, it is straightforward to obtain a power series for $(z^k dz/du)$, which is convergent in the neighborhood of 0 from the power series in (34). Using the following as the defining relation* for the sequence $\{a_j^{(k)}\}$, $j = 0, 1, 2, \cdots$,

$$z^k \frac{dz}{du} = \left(\sum_{j\geq 1} g_j u^j\right)^k \left(\sum_{j\geq 0} (j+1) g_{j+1} u^j\right) = \sum_{j\geq 0} a_j^{(k)} u^{j+k}, \tag{40}$$

the asymptotic expansion for the integral is obtained after term-by-term integration, giving

Proposition 4:

$$I^{(k)} = \int_0^\infty z^k e^{-Nf(z)}dz \sim \sum_{j=0}^{\infty} (j+k)!\,a_j^{(k)}/N^{j+k+1} \tag{41}$$

as $N \to \infty$. □

* In this notation, the sequence $\{a_j\}$ in (35) and (36) is $\{a_j^{(0)}\}$.

Let us consider in greater detail the expansion of the integral $I^{(1)} = \int_0^\infty ze^{-Nf(z)}dz$, which will be needed if (22) is used to compute the mean values. We have derived the following recursive formula which efficiently generates $\{a_j^{(1)}\}$ from the coefficients $\{a_j^{(0)}\}$ that are needed anyhow for the expansion of I: for all $j \geq 0$,

$$a_j^{(1)} = \sum_{k=1}^{j+1} a_{j-k+1}^{(0)} a_{k-1}^{(0)}/k. \tag{42}$$

In particular, the leading three terms of the expansion of $I^{(1)}$ are obtained from,

$$a_0^{(1)} = 1/f_1^2; \quad a_1^{(1)} = -3f_2/f_1^4; \quad a_2^{(1)} = 2(5f_2^2 - 2f_1 f_3)/f_1^6.$$

We have derived, but omit to give, a more general recursive formula—of which (42) is a special case—for generating $\{a_j^{(k+1)}\}$ from $\{a_j^{(k)}\}$.

4.2 Asymptotic expansions for the utilizations

As discussed so far, both of the two options stated in Section 3.2 for representing the mean values [see (19) and (22)] require the development of asymptotic expansions of two separate integrals and the subsequent computation of the ratio. Here we observe that a single asymptotic expansion exists for the mean values. The underlying reason is that the asymptotic sequence $\{N^{-j}\}$, $j = 0, 1, \cdots$, form a multiplicative asymptotic sequence.[14,22]

In particular, if as in (36) and (41),

$$\int_0^\infty e^{-Nf(z)}dz \sim \frac{a_0^{(0)}}{N} + \frac{a_1^{(0)}}{N^2} + 2!\frac{a_2^{(0)}}{N^3} + \cdots$$

and

$$\int_0^\infty ze^{-Nf(z)}dz \sim \frac{a_0^{(1)}}{N^2} + \frac{2!a_2^{(1)}}{N^3} + \frac{3!a_2^{(1)}}{N^4} \cdots,$$

then

$$\frac{\int_0^\infty ze^{-Nf(z)}dz}{\int_0^\infty e^{-Nf(z)}dz} \sim \sum_{j\geq 1} \frac{b_j}{N^j},$$

where the sequence $\{b_j\}$ is obtained by formal substitution.

The following gives the leading terms for the utilizations derived by the above procedure.

Proposition 5:
For $i = 1, 2, \cdots, p$, as $N \to \infty$,

$$\{u_i(\mathbf{K} + \mathbf{e}_i)\}^{-1}(\beta_i + 1/N) \sim \gamma_i + \frac{A_1}{N} + \frac{A_2}{N^2} + \frac{A_3}{N^3}, \tag{43}$$

where

$$A_1 = -1/\alpha; \quad A_2 = 4f_2/\alpha^3; \quad A_3 = -40f_2^2/\alpha^5 - 18f_3/\alpha^4. \quad \square$$

In accordance with an earlier observation, all terms of the asymptotic series other than the first are independent of i.

Proposition 5 contains a justification for treating the parameter α [see (28)] as an indicator of traffic intensity. Using only the dominant term,

$$\{u_i(\mathbf{K} + \mathbf{e}_i)\}^{-1} \sim \gamma_i/\beta_i \tag{44}$$

and

$$\text{utilization of CPU} = \sum u_i(\mathbf{K}) \sim 1 + \alpha. \tag{45}$$

A necessary caveat is that the above, as indeed all results in this section, has been derived for the assumption $\alpha < 0$.

Since the utilization as given by (45) can come close to unity even with $\alpha < 0$, (45) justifies another earlier statement that for large networks normal usage will not extend beyond the range $\alpha < 0$.

V. ERROR ANALYSIS AND PERFORMANCE BOUNDS FOR NORMAL TRAFFIC

Maintaining the restriction $\alpha < 0$ placed in the preceding section, we supplement in two directions the results obtained so far. First, we obtain essential results on the error incurred in truncating the expansions. These results containing information extending beyond what is required as proof of asymptotic expansions are needed for very practical reasons, such as to know how many terms to use and, more importantly, to help define the regime of applicability. In the second part of the section, certain rather special properties of the functions in the representations are used to derive analytical bounds on the network performance measures.

5.1 Completely monotonic functions

The following result on the function $g(\cdot) = f^{-1}(\cdot)$ is key to much of the error analysis:

Proposition 6:

$$(-1)^j g^{(j)}(u) < 0 \quad \text{for} \quad u \geq 0, \qquad j = 1, 2, 3, \cdots \quad \square \tag{46}$$

An inductive proof is given in Appendix B. By virtue of this result, $g^{(1)}(\cdot)$ is a completely monotonic, or alternating, function (see Ref. 23). The importance of this property stems from the role of $g^{(1)}(u) = dz/du$ in the integral representation (32).

5.2 Error bounds

In connection with the expansion of the integral I in Proposition 3, let R_m denote the error that accrues if only the leading m terms are used, i.e.,

$$R_m = I - \sum_{j=0}^{m-1} j!a_j^{(0)}/N^{j+1} = I - \sum_{j=1}^{m} g^{(j)}(0)/N^j. \tag{47}$$

Now by the mean value theorem,[25] for each u there is a $\xi(u)$ in $[0, u]$ such that

$$g^{(1)}(u) = \sum_{j=1}^{m} g^{(j)}(0)u^{j-1}/(j-1)! + g^{(m+1)}(\xi)u^m/m! \tag{48}$$

On substitution in (32),

$$I = \sum_{j=1}^{m} g^j(0)/N^j + R_m, \tag{49}$$

where

$$R_m = \frac{1}{m!}\int_0^\infty e^{-Nu}g^{(m+1)}\{\xi(u)\}u^m du. \tag{50}$$

A simple corollary to (50) and Proposition 6 is

Proposition 7:

$$\begin{aligned} R_m &> 0 \quad \text{if} \quad m \text{ is even,} \\ &< 0 \quad \text{if} \quad m \text{ is odd.} \quad \square \end{aligned}$$

This, of course, means that the terms in the asymptotic expansion alternate in sign and that the partial sums of the asymptotic expansion alternately over- and underestimate the true value of the integral in the following manner.

Proposition 8: For m even,

$$\sum_{i=1}^{m} g^{(i)}(0)/N^i \leq \int_0^\infty e^{-Nf(z)}dz \leq \sum_{i=1}^{m+1} g^{(i)}(0)/N^i. \quad \square$$

The above is quite useful since in most situations the designer would much rather overestimate than underestimate a measure such as CPU utilization. In this context, both the upper and lower bounds are required since ratios of integrals occur in the measures.

An implication of Proposition 6 is that $|g^{(m+1)}(\xi)| < |g^{(m+1)}(0)|$, $\xi > 0$, which together with (50) gives

Proposition 9:

$$\begin{aligned} R_m &< g^{(m+1)}(0)/N^{m+1} \quad \text{if} \quad m \text{ is even} \\ &> g^{(m+1)}(0)/N^{m+1} \quad \text{if} \quad m \text{ is odd.} \quad \square \end{aligned}$$

The above propositions thus state that the error is numerically less than the first neglected term of the series, and has the same sign. In particular, we have an explicit proof that (36) constitutes an asymptotic expansion. More generally, the above results show that, on account of the specialty of the integral, the main results that we require from an error analysis are already present in the expansion.

It is useful to examine in detail $g^{(4)}(0)$ and thence the bound on R_3:

$$|R_3| \leq (b_4|\alpha|^2 + 10|\alpha|b_2b_3 + 15b_2^3)/(|\alpha|^7N^4), \tag{51}$$

where $b_i = (i-1)! \sum \beta_j/\gamma_j^i$. (A look at the proof in Appendix B will convince the reader of the presence of $|\alpha|^{2m+1}N^{m+1}$ in the denominator of the bound for $|R_m|$.) The bound does make the suggestion that in cases where α is so small [i.e., utilization is very large, see (45)] that α^2N is itself small, then $|R_3|$ is large. More generally, in the case of small α^2N, the number of terms in the series requiring computation to meet specifications on the accuracy may be large. Later we return to consider this case further.

5.3 Bounds on mean values

The following bounds which supplement the computational procedures are presented to serve as aids in design and synthesis. For $i = 1, 2, \cdots, p$,

Proposition 10:

$$\left[\{u_i(\mathbf{K} + \mathbf{e}_i)\}^{-1} - \frac{\gamma_i}{\beta_i + 1/N}\right](\beta_i + 1/N)$$

$$\leq \frac{1 + \sqrt{1 + 8f_2/\alpha^2N}}{2|\alpha|N}, \tag{52}$$

$$\geq \frac{|\alpha|}{2f_2}\left(1 - \frac{2}{1 + \sqrt{1 + 16f_2/(\pi\alpha^2N)}}\right). \quad \square \tag{53}$$

Recall that $2f_2 = f^{(2)}(0) = \sum \beta_j/\gamma_j^2$ while α (here $\alpha < 0$) and N are as given in (28) and (14).

To prove Proposition 10, we may use the representation in (22) for u_i in which case it remains to bound from above and below the pair of integrals appearing there. This is done in Appendix C by making use of the sign properties of the higher-order derivatives of the function $f(\cdot)$.

Notice that as $N \to \infty$, the upper bound in Proposition 10 on $\{u_i(\mathbf{K} + \mathbf{e}_i)\}^{-1}$ approaches $\{\gamma_i - 1/(\alpha N)\}/(\beta_i + 1/N)$, the sum of the leading *two* terms of the series for u_i in (43). Also, as $N \to \infty$, the expressions in (52) and (53) both approach 0.

VI. EXPANSIONS FOR THE CASE $\alpha \approx 0$

At the end of Section 5.2, we commented that the expansions given earlier may require a large number of terms in regions where $\alpha \approx 0$. For this case, we here generate a somewhat different and more efficient series. We shall, however, be quite brief in our exposition because the uniform asymptotic expansions to be derived in Section VIII also allow appropriate expansions to be obtained. The ad hoc but direct treatment here is supplementary.

We need notation that is specific to this section. Let

$$\left.\begin{aligned} K_i &= b_i N + d_i\sqrt{N} \\ r_i &= a_i(N + c\sqrt{N}) \end{aligned}\right\} \quad i = 1, 2, \cdots, p. \tag{54}$$

Each of the variables $\{d_i\}$ and c may be either positive or negative. However, as our interest in this section is in $\alpha \approx 0$ and as $\alpha \sim \sum b_i/a_i - 1$ as $N \to \infty$, we require that

$$\sum b_i/a_i = 1. \tag{55}$$

In a computational procedure the above restriction poses no particular problem.

Also, we shall mainly consider the integral

$$I \triangleq \left(\prod r_i^{-K_i}\right) \int_0^\infty e^{-t} \prod (r_i + t)^{K_i} dt. \tag{56}$$

Comparison with (16) shows that the integral is related to the partition function thus:

$$G(\mathbf{K}) = \left(\prod \frac{\rho_{i1}^{K_i}}{K_i!}\right) I. \tag{57}$$

As previously, the computation of the quantity in parentheses is not required to obtain the mean values. Our main result is

Proposition 11: As $N \to \infty$,

$$I \sim \sum_{j=0}^{\infty} c_j/\sqrt{N}^{j-1}, \tag{58}$$

where the sequence $\{c_j\}$ is given below. □

The proof of the proposition is in Appendix D. Here we comment on some features of the sequence:

$$c_j = \int_0^{\infty} e^{-A\nu^2/2-B\nu} H_j(\nu)d\nu, \tag{59}$$

where $A = \sum b_i/a_i^2 > 0$ and $B = c - \sum d_i/a_i$ and $H_j(\nu)$ is a polynomial of degree $3j$ in ν with coefficients that are fairly straightforward to obtain. The key point is that A, B as well as the coefficients of the polynomials are all 0(1), i.e., N does not enter into their definitions. Results given in Section VIII indicate how the coefficients c_j given in (59) may be effectively computed.

We give below the leading three polynomials in the sequence $\{H_j(\nu)\}$:

$$\begin{aligned} H_0(\nu) &= 1, \\ H_1(\nu) &= c - (\textstyle\sum d_i/a_i^2)\nu^2/2 + (\sum b_i/a_i^3)\nu^3/3, \\ H_2(\nu) &= -c(\textstyle\sum d_i/a_i^2)\nu^2/2 + (\sum d_i/a_i^3 + c\sum b_i/a_i^3)\nu^3/3 \\ &\quad + [(\textstyle\sum d_i/a_i)^2 - 2\sum b_i/a_i^4]\nu^4/8 \\ &\quad - (\textstyle\sum d_i/a_i^2)(\sum b_i/a_i^3)\nu^5/6 + (\sum b_i/a_i^3)^2\nu^6/18. \end{aligned} \tag{60}$$

VII. ASYMPTOTIC EXPANSIONS FOR HEAVY TRAFFIC CONDITIONS

Here we obtain asymptotic expansions for the basic integral in (17) and (18) for the case $\alpha > 0$. For reasons similar to those discussed earlier for the case $\alpha < 0$, the expansions to use for $\alpha \approx 0$ are in Sections VI and VIII. Hence, the expansions given below are for exceptionally heavy traffic conditions, where α is not only positive but also not close to 0.

7.1 Laplace's method

The key difference from the treatment in Section 4.1 is the presence of the singularity at $\hat{z}$ (see Fig. 2), which will be assumed to be known. For large N the dominant contribution to the integral

$$I = \int_0^{\infty} e^{-Nf(z)}dz \tag{61}$$

comes from the neighborhood of $\hat{z}$. A Taylor series expansion around $\hat{z}$ gives

$$f(z) - \hat{f}_0 = \sum_{j=2}^{\infty} \hat{f}_j (z - \hat{z})^j, \tag{62}$$

where

$$\hat{f}_j = f^{(j)}(\hat{z})/j!, \qquad j = 0, 1, 2, \cdots. \tag{63}$$

In particular, for $j \geq 2$,

$$\hat{f}_j = (-1)^j \left[\sum_i \beta_i/(\gamma_i + \hat{z})^j \right] \Big/ j. \tag{64}$$

We make the following specific decomposition of I, which is convenient here and even more so in the error analysis to follow with z_2 as in Fig. 2,

$$e^{N\hat{f}_0} I = \int_0^{\hat{z}} e^{-N\{f(z)-\hat{f}_0\}} dz + \int_{\hat{z}}^{z_2} e^{-N\{f(z)-\hat{f}_0\}} dz$$
$$+ \int_{z_2}^{\infty} e^{-N\{f(z)-\hat{f}_0\}} dz. \tag{65}$$

Consider the terms in turn starting with the middle term. If we let

$$u \triangleq f(z) - \hat{f}_0, \qquad z \geq \hat{z}, \tag{66}$$

and use the series in (62) for the right-hand side, then we may reverse the series, as discussed in Appendix A, to obtain

$$z - \hat{z} \triangleq g(u) = \sum_{j=1}^{\infty} g_j u^{j/2} \tag{67}$$

with the coefficient g_j depending only on the coefficients $\hat{f}_2, \hat{f}_3, \cdots, \hat{f}_j$. Now,

$$\int_{\hat{z}}^{z_2} e^{-N\{f(z)-\hat{f}_0\}} dz = \int_0^{-\hat{f}_0} e^{-Nu} g^{(1)}(u) du$$
$$= \int_0^{-\hat{f}_0} e^{-Nu} \sum_{j=0}^{\infty} a_j u^{(j-1)/2} du, \tag{68}$$

where $a_j = (j+1) g_{j+1}/2$. The individual integrals in the sum will be recognized to be incomplete gamma functions.

On returning to (65) and the first term in the right-hand side, we find by an identical argument that

$$\int_0^{\hat{z}} e^{-N(f(z)-\hat{f}_0)}dz = \int_0^{-\hat{f}_0} e^{-Nu} \sum_{j=0}^{\infty} (-1)^j a_j u^{(j-1)/2} du. \tag{69}$$

The two integrals in (68) and (69) may conveniently be combined to give

$$e^{N\hat{f}_0} I = \int_0^{-\hat{f}_0} e^{-Nu} \sum_{j=0}^{\infty} 2a_{2j} u^{j-1/2} du + \int_{z_2}^{\infty} e^{-N(f(z)-\hat{f}_0)} dz. \tag{70}$$

At this stage, the following two approximations are made, with their effects bounded in Section 7.2 in the course of the error analysis: the integration interval in the first term is extended to [0, ∞) and the second term is ignored. Nonetheless, the error analysis shows that

$$I \sim e^{-N\hat{f}_0} \int_0^{\infty} e^{-Nu} \sum_{j=0}^{\infty} 2a_{2j} u^{j-1/2} du, \tag{71}$$

giving Proposition 12.

Proposition 12: As $N \to \infty$,

$$I = \int_0^{\infty} e^{-Nf(z)} dz \sim e^{-N\hat{f}_0} \sum_{j=0}^{\infty} 2\Gamma(j + 1/2) a_{2j}/N^{j+1/2}. \quad \square \tag{72}$$

Recall that $\Gamma(1/2) = \sqrt{\pi}$ and for $j = 1, 2, \cdots,$

$$\Gamma(j + 1/2) = \sqrt{\pi} \prod_{i=1}^{j} (i - 1/2).$$

We give the leading three coefficients:

$$\begin{aligned} a_0 &= (1/2)/\hat{f}_2^{1/2}, \\ a_2 &= (1/16)(15\hat{f}_3^2 - 12\hat{f}_2\hat{f}_4)/\hat{f}_2^{7/2}, \\ a_4 &= (5/256)(-64\hat{f}_2^3\hat{f}_6 + 224\hat{f}_2^2\hat{f}_3\hat{f}_5 + 112\hat{f}_2^2\hat{f}_4^2 \\ &\quad - 504\hat{f}_2\hat{f}_3^2\hat{f}_4 + 231\hat{f}_3^4)/\hat{f}_2^{13/2}. \end{aligned} \tag{73}$$

The procedure for obtaining the asymptotic expansion for the integral

$$I^{(1)} = \int_0^{\infty} z e^{-Nf(z)} dz \tag{74}$$

is similar. Notice

$$(I^{(1)} - \hat{z} I) = e^{-N\hat{f}_0} \int_0^{\infty} (z - \hat{z}) e^{-N(f(z)-\hat{f}_0)} dz; \tag{75}$$

we find it convenient to expand the integral on the right-hand side. The expression analogous to (71) is

$$(I^{(1)} - \hat{z}I) \sim e^{-N\hat{f}_0} \int_0^{\infty} e^{-Nu} \sum_{j=1}^{\infty} 2a_{2j-1}^{(1)} u^{j-1/2} du, \tag{76}$$

where, for $j = 0, 1, 2, \cdots$,

$$a_j^{(1)} = \frac{1}{2} \sum_{k=1}^{j+1} k g_{j-k+2} g_k, \tag{77}$$

which is to be compared with the expression for a_j following (68). The following is a useful formula for efficiently generating the sequence $\{a_j^{(1)}\}$ from $\{a_j\}$, which is needed anyhow for computing I:

$$a_j^{(1)} = 2 \sum_{k=0}^{j} a_{j-k} a_k/(k+1), \qquad j = 0, 1, 2, \cdots. \tag{78}$$

Recognizing the gamma functions in (76) gives Proposition 13.

Proposition 13:

$$I^{(1)} - \hat{z}I \sim e^{-N\hat{f}_0} \sum_{j=1}^{\infty} 2\Gamma\left(j + \frac{1}{2}\right) a_{2j-1}^{(1)}/N^{j+1/2} \tag{79}$$

as $N \rightarrow \infty$. □

The asymptotic expansions in Propositions 12 and 13 may also be combined, as discussed earlier in Section 4.2 to yield an asymptotic expansion for the mean values. In particular, we obtain Proposition 14.

Proposition 14: As $N \rightarrow \infty$,

$$\{u_i(\mathbf{K} + \mathbf{e}_i)\}^{-1}(\beta_i + 1/N) \sim (\gamma_i + \hat{z}) + \frac{A_1}{N} + \frac{A_2}{N^2},$$

where

$$A_1 = -3\hat{f}_3/4\hat{f}_2^2,$$

$$A_2 = (6\hat{f}_2\hat{f}_3\hat{f}_4 - 15\hat{f}_2^2\hat{f}_5/8 - 135\hat{f}_3^3)/\hat{f}_2^5. \square \tag{80}$$

7.2 Error analysis

The analysis to be presented supplements the result in Proposition 12 and the error estimates to be given provide guidelines for the use of the expansions. The broad outline of the analysis have been suggested in Ref. 23.

As in Section 5.2, let R_n denote the error incurred when only n leading terms of the series in Proposition 12 is used, i.e.,

$$R_n = I - e^{-N\hat{f}_0} \sum_{j=0}^{n-1} \Gamma\left(j + \frac{1}{2}\right) 2a_{2j}/N^{j+1/2}. \tag{81}$$

For I we will use the expression given in (70) and decompose the error R_n thus

$$R_n = -\epsilon_{n,1}(N) + \epsilon_{n,2}(N) + \epsilon_3(N), \tag{82}$$

where

$$\epsilon_{n,1}(N) = e^{-N\hat{f}_0} \int_{-\hat{f}_0}^{\infty} e^{-Nu} \left(\sum_{j=0}^{n-1} 2a_{2j}u^{j-1/2} \right) du, \tag{83}$$

$$\epsilon_{n,2}(N) = e^{-N\hat{f}_0} \int_{0}^{-\hat{f}_0} e^{-Nu} \left(\sum_{j=n}^{\infty} 2a_{2j}u^{j-1/2} \right) du, \tag{84}$$

$$\epsilon_3(N) = \int_{z_2}^{\infty} e^{-Nf(z)} dz. \tag{85}$$

Thus, the three terms on the right-hand side of (82) respectively denote components arising from the extension of the integration interval from $[0, -\hat{f}_0]$ to $[0, \infty)$ in (70) and (71), the use of only n leading terms from the infinite series in (71), and the neglect of the second term in the right-hand side of (70). Each component is now bounded.

To bound $\epsilon_{n,1}(N)$, we make use of known bounds on the incomplete gamma function:[23]

$$\int_{-\hat{f}_0}^{\infty} e^{-Nu}u^{j-1/2}du = \Gamma\left(j + \frac{1}{2}, -N\hat{f}_0\right) \Big/ N^{j+1/2}$$

$$\le \frac{e^{N\hat{f}_0}(N|\hat{f}_0|)^{j+1/2}}{N^{j+1/2}\left\{N|\hat{f}_0| - \max\left(j - \frac{1}{2}, 0\right)\right\}} \quad \text{for} \quad n|\hat{f}_0| > \max\left(j - \frac{1}{2}, 0\right).$$

Thus,

$$|\epsilon_{n,1}(N)| \le \frac{2}{N|\hat{f}_0| - \delta_n} \sum_{j=0}^{n-1} |a_{j2}|\,|\hat{f}_0|^{j+1/2}, \tag{86}$$

where $\delta_n = \max(n - 3/2, 0)$.

In bounding $\epsilon_{n,2}(N)$, we will postulate the existence of a finite valued σ_n with the property that

$$\left| \sum_{j=n}^{\infty} 2a_{2j}u^{j-1/2} \right| \le |2a_{2n}|\,u^{n-1/2}e^{\sigma_n u},\ 0 < u < |\hat{f}_0|. \tag{87}$$

This approach fails when σ_n is infinite but the characteristics of the integral and the specific decomposition (65) that has been employed preclude this possibility. Let

$$\sigma_n = \max_{|u|<|\hat{f}_0|} \psi_n(u), \tag{88}$$

where

$$\psi_n(u) = \frac{1}{u} \ln \left| \frac{\sum_{j=n}^{\infty} 2a_{2j}u^{j-1/2}}{2a_{2n}u^{n-1/2}} \right|. \tag{89}$$

Small u is where (see Ref. 23) the danger of unbounded σ_n is usually most manifest. However, as

$$\psi_n(u) \sim \frac{a_{2n+2}}{a_{2n}} + \left(\frac{a_{2n+4}}{a_{2n}} - \frac{a_{2n+2}^2}{a_{2n}^2} \right) u + \cdots \tag{90}$$

when $u \to 0^+$, no problem arises here.

Using (87) in the defining expression for $\epsilon_{n,2}(N)$ in (84)

$$|\epsilon_{n,2}(N)| \le e^{-N\hat{f}_0}|2a_{2n}| \int_0^{|\hat{f}_0|} e^{-(N-\sigma_n)u}\, u^{n-1/2}du$$

$$= e^{-N\hat{f}_0}|2a_{2n}|\Gamma(n + ½)/(N - \sigma_n)^{n+1/2}. \tag{91}$$

The last term to be considered from (82) is $\epsilon_3(N)$. We use the following property.

$$f(z) \ge f'(z_2)(z - z_2), \qquad z \ge z_2, \tag{92}$$

which yields

$$|\epsilon_3(N)| \le \frac{1}{Nf'(z_2)}, \tag{93}$$

a small quantity compared to the right-hand side of (91). This concludes the process of bounding the components of the error term R_n. A corollary is the proof to Proposition 12.

The bound in (91) is the largest component in the error bound. In examining (91), we observe that the condition in which the bound is large is when a_{2n}/N^n is large. Now the expression for a_{2n} contains in the denominator a term $f_2^{3n+1/2}$, as (73) attests. Thus, when f_2^3N is small, we expect the asymptotic expansions in Section 7.1 to be inefficient. We return to this case later in Section 8.2, where this as well as the similar difficulty encountered in Section 5.2—where α was negative and small—is treated in a unified manner.

VIII. UNIFORM EXPANSIONS

This section has two objectives. The first is to show that there is a framework that unifies the expansions in Sections IV, VI, and VII.

This consists of showing that the integrals of interest may each be given by a common expansion valid uniformly for the entire range of values of the system parameters. These expansions turn out to be in parabolic cylinder (or Weber) functions.[13,25,26] The advantage derived is that these classical special functions have extensively documented expansions for the entire range of parameter values.[13,26] Indeed, using these expansions we sketch in Sections 8.3 and 8.4, at the cost of some duplication, derivations of the expansions obtained earlier in Sections IV and VII for $\alpha < 0$ and $\alpha > 0$, respectively. The second objective is to derive a computationally efficient expansion for the case $\alpha \approx 0$, i.e., where the stationary point $\hat{z}$ is very close to the boundary of the integration interval. The error analysis in Sections 5.2 and 7.2 has shown the need for a separate treatment. The expansion that is obtained for this case in Section 8.2 is obtained from an appropriate expansion of the parabolic cylinder functions.

8.1 Uniform expansions in parabolic cylinder functions

Consider the integral

$$I = \int_0^\infty e^{-Nf(z)} dz \tag{94}$$

without restrictions on the parameter α. Following Friedman,[24] consider a change of variables from z to v given by

$$v^2 - 2av = f(z), \tag{95}$$

where a is a parameter of the transformation to be fixed later. The objective of the transformation is that the component of the integrand in braces below

$$I = \int_0^\infty e^{-N(v^2-2av)} \left(\frac{dz}{dv}\right) dv \tag{96}$$

satisfy the dual requirements of boundedness and a convergent power series, as required for an application of Watson's lemma (see Section 4.1 following Proposition 3). [The reader may verify that the simpler transformation $v = f(z)$ violates the boundedness requirement whenever $\alpha > 0$ and $z = \hat{z}$ since $f^{(1)}(\hat{z}) = 0$.] For the transformation in (95),

$$\frac{dz}{dv} = \frac{2(v-a)}{f^{(1)}(z)}. \tag{97}$$

This suggests the selection of the parameter a to be such that $v = a$ when $z = \hat{z}$, with the accompanying indeterminacy and the possibility of boundedness of dz/dv. This key clue does indeed give a unique map of the form in (95) with the desired properties, as summarized below.

Proposition 15: For $z \geq 0$, let

$$v(z) = a + \operatorname{sgn}(z - \hat{z})\sqrt{f(z) - f(\hat{z})}, \tag{98}$$

where the constant a depends on all the system parameters:

$$a = (\operatorname{sgn} \alpha)\sqrt{-f(\hat{z})}. \tag{99}$$

The transformation is monotonic, increasing and maps $[0, \infty)$ to $[0, \infty)$. Also dv/dz is continuous and uniformly bounded. □

The transformation is used to derive for dz/dv a convergent power series $\sum_0^\infty C_j v^j$ in a neighborhood of the origin. This is achieved in three steps:

(*i*) Use (98) to obtain

$$v(z) = A_1 z + A_2 z^2 + A_3 z^3 + \cdots. \tag{100}$$

(*ii*) Reverse the series (see Appendix A) to obtain

$$z(v) = B_1 v + B_2 v^2 + B_3 v^3 + \cdots. \tag{101}$$

(*iii*) Differentiate term by term to obtain

$$\frac{dz}{dv} = C_0 + C_1 v + C_2 v^2 + \cdots, \tag{102}$$

where $C_j = (j + 1)B_{j+1}$.

The reader may verify that the leading terms of the sequence $\{C_j\}$ thus obtained are as follows [recall from (33) the definition $f_j = f^{(j)}(0)/j!$]:

$$C_0 = 2a/\alpha,$$
$$C_1 = \frac{2}{\alpha}\left[\left(\frac{2a}{\alpha}\right)^2 f_2 - 1\right]. \tag{103}$$

The above tacitly assumes that $\hat{z}$ and hence a have been evaluated.

The power series expansion for dz/dv may now be substituted in (96) to yield

$$I = \sum_{j=0}^{\infty} C_j \int_0^\infty e^{-N(v^2 - 2av)} v^j dv. \tag{104}$$

The integrals appearing above are simply related to the parabolic cylinder functions $U(\cdot, \cdot)$; thus: for $j = 0, 1, 2, \cdots$,

$$\int_0^\infty e^{-N(v^2-2av)} v^j dv = \frac{e^{Na^2/2} j!}{(2N)^{(j+1)/2}} U\left(j + \frac{1}{2}, -a\sqrt{2N}\right). \tag{105}$$

Expansions for related integrals such as

$$I^{(1)} = \int_0^\infty z e^{-Nf(z)} dz \tag{106}$$

are only slightly different. Here the term dz/dv is replaced by zdz/dv, which has the power series expansion

$$zdz/dv = \sum_{j=1}^{\infty} C_j^{(1)} v^j, \tag{107}$$

where, see (101),

$$C_j^{(1)} = \sum_{k=0}^{j-1} B_{j-k} C_k. \tag{108}$$

Specifically,

$$C_1^{(1)} = \left(\frac{2a}{\alpha}\right)^2,$$

$$C_2^{(1)} = \frac{3}{\alpha}\left(\frac{2a}{\alpha}\right)\left[\left(\frac{2a}{\alpha}\right)^2 f_2 - 1\right]. \tag{109}$$

The following summarizes the expansions in parabolic cylinder functions of the two integrals of main interest, with the expansion valid for all α.

Proposition 16:

$$I = \int_0^\infty e^{-Nf(z)} dz = \sum_{j=0}^{\infty} C_j \int_0^\infty e^{-N(v^2-2av)} v^j dv$$

$$= e^{Na^2/2} \sum_{j=0}^{\infty} \frac{j! C_j}{(2N)^{(j+1)/2}} U\left(j + \frac{1}{2}, -a\sqrt{2N}\right), \tag{110}$$

$$I^{(1)} = \int_0^\infty z e^{-Nf(z)} dz$$

$$= e^{Na^2/2} \sum_{j=1}^{\infty} \frac{j! C_j^{(1)}}{(2N)^{(j+1)/2}} U\left(j + \frac{1}{2}, -a\sqrt{2N}\right), \tag{111}$$

where the sequences $\{C_j\}$ and $\{C_j^{(1)}\}$ are respectively as obtained by the procedures in (100) through (102) and (108). Specifically the leading terms are as given in (103) and (109). □

We should add that the above expansions are not strictly asymptotic expansions since the parabolic cylinder functions do not satisfy the requirements of asymptotic sequences for certain ranges of the parameter a^2N.[14] The interested reader will find in Ref. 27 a description of the process for obtaining uniform asymptotic expansions of the integrals. However, we have not found it necessary to undertake the additional effort required to obtain the coefficients of the uniform asymptotic expansions. This is because for a^2N small, the case treated below in Section 8.2 and of main interest, the functions $U(j + \frac{1}{2}, -a\sqrt{2N})$, $j \geq 0$, have all the desirable properties that are required of asymptotic sequences.

A noteworthy property of the functions $U(\cdot, \cdot)$ that can be important in computations is that it satisfy the recursion[13]

$$xU(j + \tfrac{1}{2}, x) = U(j - \tfrac{1}{2}, x) - (j + 1)U(j + \tfrac{3}{2}, x). \qquad (112)$$

8.2 Expansions for the case of a^2N small

To motivate the results to be given here, observe that the stationary point of the curve of $f(z)$ (see Fig. 2) is close to 0 when α is small (utilization of CPU is high), since

$$\hat{z} \sim \alpha/(2f_2) \qquad \text{as} \qquad \hat{z} \to 0. \qquad (113)$$

On enquiring how the parameter a behaves for small $\hat{z}$ and α, we find from (99) that

$$a = \frac{\alpha}{2(f_2)^{1/2}} + O(\alpha^2) \qquad \text{as} \qquad \alpha \to 0. \qquad (114)$$

Thus, the case of small α, which we know from Section 5.2 requires special treatment, corresponds to small a.

Small a is also implied by small $\hat{f}_2$ and is therefore also of interest on account of the discussion in Section 7.2. This follows from

$$a = \hat{z}(\hat{f}_2)^{1/2} + O(\hat{z}^{3/2}) \qquad \text{as} \qquad \hat{z} \to 0. \qquad (115)$$

For small a^2N, it is known[13] that for $j \geq 0$,

$$\int_0^\infty e^{-N(v^2-2av)}v^j dv = \frac{e^{Na^2/2}j!}{(2N)^{(j+1)/2}} U\left(j + \frac{1}{2}, -a\sqrt{2N}\right)$$

$$= \frac{\sqrt{\pi}}{N^{(j+1)/2}} [\mu_0^{(j)} + (a\sqrt{N})\mu_1^{(j)} + (a\sqrt{N})^2\mu_2^{(j)} + \cdots], \qquad (116)$$

where

$$\mu_i^{(j)} = \frac{j!}{i!} \frac{2^{i/2-j-1}}{\Gamma(1 + j/2)} (j + 1)(j + 3) \cdots (j + i - 1), \qquad i \text{ even}$$

$$= \frac{j!}{i!} \frac{2^{(i+1)/2-j-1}}{\Gamma((j + 1)/2)} (j + 2)(j + 4) \cdots (j + i - 1), \qquad i \text{ odd}. \qquad (117)$$

In these relations, notice first in (116) the desirable presence of the powers of N in the denominator. Secondly, in connection with the sequence $\{\mu_i^{(j)}\}$, observe that $\mu_{i+2}^{(j)}/\mu_i^{(j)} = 2(j+i+1)/\{(i+1)(i+2)\}$. Thus, for fixed j, the sequence converges rapidly to 0 with increasing i. These observations state that when using the expression (116) in the expansions of Proposition 16, first, it is necessary to compute only up to small values of the index j and, secondly, with $a\sqrt{N}$ small, the computation of the bracketed quantity in (116) also needs very few terms.

The following summarizes the important computational procedure described above.

Proposition 17: For small a^2N,

$$I = \sqrt{\pi} \sum_{j \geq 0} \frac{C_j}{N^{(j+1)/2}} \left[\sum_{i \geq 0} (a\sqrt{N})^i \mu_i^{(j)} \right], \tag{118}$$

$$I^{(1)} = \sqrt{\pi} \sum_{j \geq 1} \frac{C_j^{(1)}}{N^{(j+1)/2}} \left[\sum_{i \geq 0} (a\sqrt{N})^i \mu_i^{(j)} \right], \tag{119}$$

where $\mu_i^{(j)}$ is in (117), and a, $\{C_j\}$, $\{C_j^{(1)}\}$ appear in Proposition 16. □

8.3 Expansions for normal traffic

For normal traffic, $\alpha < 0$ and consequently $a < 0$. If in addition, $\alpha \ll 0$ or, specifically, $a^2N \gg j^2$, then[13]

$$\int_0^\infty e^{-N(v^2-2av)} v^j dv$$

$$\sim \frac{1}{(-2aN)^{j+1}} \left(j! - \frac{(j+2)!}{4(a^2N)} + \frac{(j+4)!}{32(a^2N)^2} \mp \cdots \right). \tag{120}$$

It can be shown after some manipulations, which we omit, that this expansion when substituted in Proposition 16 is identical to the main result of Section IV, namely, the expansion in Proposition 3. The bridging relation is

$$(v-a)^2 = u + a^2, \tag{121}$$

where u is the integration variable in Section IV [see (31)] and v is the similar variable in the uniform expansion [see (98)].

8.4 Expansions for heavy traffic

For heavy traffic, $\alpha > 0$, and therefore $a > 0$. When $a^2N \gg 1$ as well, then[13]

$$\int_0^\infty e^{-N(v2-2av)} v^j dv = e^{Na^2} \frac{\sqrt{\pi}}{\sqrt{N}} a^j \left(1 + \frac{j(j-1)}{4a^2N} + \frac{j(j-1)(j-2)(j-3)}{32a^4N^2} + \cdots \right). \tag{122}$$

Notice the departure from the expressions in (116) and (120) in the absence of N^j in the denominator.

It may again be established, although not in a simple manner, that the main result of Section VII, the expansion in Proposition 12, is also obtained by substituting the above expansion in Proposition 16. The bridging relation is

$$(v - a)^2 = u, \tag{123}$$

where the variable u is as in (66).

IX. COMPUTATIONAL NOTES

The asymptotic expansions of integral representations of various quantities in the closed queueing networks discussed above, have been implemented as a user-oriented interactive package on Digital Equipment Corporation's VAX 11/780 operating under programmers work bench *UNIX** system version 3. The package written in C-language has about 400 C-language statements and occupies about 60 Kbytes of storage.

The number of classes and constituents† that the package can accommodate is so large that in effect no restriction is placed on these parameters. The other main features of the package are enumerated below.

(*i*) The package is user oriented and easy to use. The user is prompted for relevant problem data. As this is supplied, validation and feasibility checks are made and the user informed of any errors.

(*ii*) The output of the package includes all relevant statistics on each class (response time, utilization, etc.) including the percentage error incurred in the expansion. As an option to the last named, the user can display the various terms used in the expansion.

(*iii*) The package is partly adaptive in that it automatically detects the divergence of the asymptotic expansion and truncates the series at the point of divergence.

(*iv*) Numerical stability is enforced by proper choice of N.

Numerous computational experiments were performed to compare the efficacy of our package, called ASYM, to the current version of a popular commercially available package CADS. CADS is marketed by Information Research Associates and a version of it runs on a VAX 11/780 operating under *UNIX* time sharing system. The test problems run on both these packages are real-world problems encountered in performance analysis of a Bell System project.

---

* Trademark of Bell Laboratories.

† In the computer science applications, the class sizes in closed networks $\{K_j\}$ are referred to as degrees of multiprogramming.

The results of the experiments indicate that CADS is unable to solve our moderately sized, closed-queuing-network problems. Two features that accompany the breakdowns are noteworthy. One is numerical instability which manifests itself as overflows or underflows. Recent experience is that rescaling the service rates, an often-mentioned device to combat the problem, does not help in substantially increasing the problem size. This accuracy problem is of course relieved with the use of more powerful floating-point machines like the CDC 6600 (manufactured by Control Data Corporation) or IBM 3033. The other feature is the built-in limitations of the package. For instance, the current version of CADS limits the degree of multiprogramming for any one class to 100. There is also a limit of three classes.

The above discussion is not intended to disparage the usefulness and success of CADS. CADS is extremely powerful in solving *small* queueing-network problems. Packages implementing recursions, as CADS does, and implementing expansions of integrals complement each other and, when integrated, will provide a powerful general-purpose package.

The computational experiments (see Tables I to IV below) yielded the interesting fact that the asymptotic expansions are quite effective (i.e., yield a small percentage of errors) even for *small* problems, provided the right expansion is used. This in conjunction with the linear growth in computation time with increased number of classes

Table I—Results for test problem No. 1 *

Degree of Multiprogramming	Utilization of CPU Given by CADS	Utilization of CPU Given by ASYM
10	0.0417	0.0414
20	0.083	0.0829
30	0.124	0.124
40	0.166	0.165
50	0.207	0.207
60	0.249	0.248
70	0.290	0.289
80	Breakdown	0.331
90	Breakdown	0.372
100	Breakdown	0.413
110	Breakdown	0.618
150	Breakdown	0.618
200	Breakdown	0.819

* Problem specification:
No. of classes = 1
Think time = 240 seconds
Processing time = 1 second

Table II—Results for test problem No. 2*

Degrees of Multiprogramming	Total Utilization of CPU Given by CADS	Total Utilization of CPU Given by ASYM
10/10	0.118	0.119
20/20	0.239	0.23
30/30	0.358	0.35
40/40	0.476	0.48
50/50	0.593	0.60
60/60	Breakdown	0.70
80/60	Breakdown	0.72
90/60	Breakdown	0.97
100/50	Breakdown	0.69
110/50	Breakdown	0.71
140/50	Breakdown	0.79
200/10	Breakdown	0.54
170/40	Breakdown	0.75

* Problem specification:
No. of classes = 2
Think time for class 1 = 450 seconds
Think time for class 2 = 150 seconds
Processing time for class 1 = 1 second
Processing time for class 2 = 1.5 seconds

Table III(a)—Problem specification for test problem 3

Class	Service Rate of Infinite Server for This Class	Service Rate of CPU for This Class	Degree of Multiprogramming for This Class
1	0.0001	6.00	500
2	0.0005	2.00	500
3	0.0006	4.00	500
4	0.0002	0.33	500
5	0.0002	0.60	500
6	0.00005	0.20	500
7	0.00005	1.00	500

Table III(b)—Results of test problem 3 output by ASYM

Class	Response Time (seconds)	Utilization	% Error in Utilization
1	0.95	0.01	0.0025
2	2.85	0.12	0.0025
3	1.42	0.08	0.0025
4	17.21	0.3	0.0025
5	9.48	0.17	0.0025
6	28.46	0.13	0.0025
7	5.70	0.02	0.0025

Table IV(a)—Problem specification for test problem 4

Class	Service Rate of Infinite Server for This Class	Service Rate of CPU for This Class	Degree of Multiprogramming for This Class
1	0.0033	20.0	5
2	0.033	2	5
3	0.0033	4	5
4	0.033	4	5
5	0.033	4	5
6	0.033	6	5
7	0.033	20	5
8	0.00033	0.6	5
9	0.00055	0.6	5
10	0.0033	0.6	5
11	0.00033	0.2	5
12	0.00055	0.2	5
13	0.0003	0.2	5
14	0.033	1	5
15	0.00033	1	5
16	0.00055	1	5
17	0.0003	1	5

Table IV(b)—Results of test problem 4 output by ASYM

Class	Response Time (seconds)	Utilization	% Error in Utilization
1	0.09	0.008	0.05
2	0.834	0.080	0.045
3	0.43	0.004	0.05
4	0.42	0.046	0.048
5	0.42	0.040	0.049
6	0.28	0.027	0.049
7	0.09	0.008	0.05
8	2.85	0.003	0.06
9	2.9	0.005	0.05
10	2.82	0.027	0.05
11	8.530	0.008	0.05
12	8.523	0.013	0.05
13	8.540	0.007	0.05
14	1.63	0.156	0.04
15	1.71	0.002	0.05
16	1.72	0.003	0.05
17	1.71	0.001	0.05

makes the use of the asymptotic expansions attractive even in cases where the recursive implementation does not break down.

Tables I and II display the results of problems run both on CADS and ASYM. Tables III and IV show the results on two large problems that were not admitted by CADS, but were solved with good accuracy by ASYM.

It may be observed from Table I that the results from ASYM and CADS agree rather well, even for small degrees of multiprogramming. In the cases solved by both CADS and ASYM, α was small and N (about 240) large. Observe also that CADS was unable to solve cases with K_1 larger than 70, even though higher values of K_1 correspond to quite low usage of the CPU and are quite interesting.

Table II shows that CADS also broke down on a relatively small problem with 2 classes and 60 customers in each class.

Tables III and IV display the output given by ASYM for the large problems. Table III corresponds to a problem where the total degree of multiprogramming is 3500. For Table IV the total degree of multiprogramming is 85, but there are 17 classes. As shown, the percentage error in both cases is well within an acceptable range.

In conclusion, the computational experiments suggest that the approach based on expansions of integral representations is robust, computationally fast, and to be recommended for a variety of problems, large and small.

X. GENERALIZATIONS: INTEGRALS IN NETWORKS WITH MANY PROCESSORS

This section shows that for a large class of closed Markovian networks, the partition functions possess simple integral representations. This result provides the basis for future work on the computations of the integrals from expansions rather like those given in this paper. The above-mentioned class of networks allows an arbitrary number of service centers with flexibility in operating disciplines. It is in fact the same class of networks shown by Baskett et al., in Ref. 4, that has the product form in the stationary distributions, except that we do not allow the service rate in Type 1 centers to depend on the number in queue. (To some extent this is only for convenience because for some specific and interesting dependencies, we have obtained the integral representations.)

The representation of the partition function that is obtained is as a multiple integral, i.e., as an integral in Euclidean space of dimension q where q is the number of *queuing* centers, which are the centers of Type 1, 2, and 4 in the notation of Ref. 4. However, in spite of the complexity of the partition function, the form of the integrand is remarkable for its simplicity.

As previously, we let the number of classes of constituents be p. We henceforth consistently index the classes by the symbol j; when the index for summation or multiplication is omitted, it is understood that the missing index is j, where $1 \le j \le p$. A total of s service centers are allowed. We will find it natural to distinguish the centers of Types 1, 2, and 4, which have queuing, from the remaining centers of Type 3, which do not. Thus, centers 1 through q will be the queuing centers while $(q + 1)$ through s will be the Type 3 centers, which have also been called think nodes and infinite server nodes.

Let the

$$\text{stationary state probability} = \pi(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_s), \tag{124}$$

$$\mathbf{y}_i \triangleq (n_{1i}, n_{2i}, \cdots, n_{pi}), 1 \le i \le s,$$

$$n_{ji} \triangleq \text{number of class—}j \text{ jobs in center } i.$$

The well-known results on Markovian closed queuing networks with product form solutions may be given in the following form:[1,4]

$$\pi(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \frac{1}{G} \prod_{i=1}^{s} \pi_i(\mathbf{y}_i), \tag{125}$$

$$\text{where } \pi_i(\mathbf{y}_i) = \left(\sum n_{ji}\right)! \prod \left(\frac{\rho_{ji}^{n_{ji}}}{n_{ji}!}\right), \qquad 1 \le i \le q,$$

$$= \prod \left(\frac{\rho_{ji}^{n_{ji}}}{n_{ji}!}\right), \qquad (q+1) \le i \le s. \tag{126}$$

In the above formulas we have taken into account the previously stated assumption, namely, for the first-come-first-served discipline in Type 1 centers the service rate is independent of the number of jobs in queue. Also, in (126),

$$\rho_{ji} = \frac{\text{expected number of visits of class } j \text{ jobs to center } i}{\text{service rate of class } j \text{ jobs in center } i},$$

where the numerator is obtained from the given routing matrix by solving for the eigenvector corresponding to the eigenvalue at 1.

In (125) G is, of course, the partition function and it is explicitly

$$G = \sum_{\mathbf{1}'\mathbf{n}_1 = K_1} \cdots \sum_{\mathbf{1}'\mathbf{n}_p = K_p} \prod_{i=1}^{s} \pi_i(\mathbf{y}_i), \tag{127}$$

where we have written $\mathbf{1}'\mathbf{n}_j$ for $\sum_{i=1}^{s} n_{ji}$ and the condition $\mathbf{1}'\mathbf{n}_j = K_j$ to indicate the conservation of jobs in each class. Thus,

$$G = \sum \cdots \sum \left[\prod_{i=1}^{q} \left((\sum n_{ji})! \prod \frac{\rho_{ji}^{n_{ji}}}{n_{ji}!} \right) \right] \left[\prod_{i=q+1}^{s} \left(\prod \frac{\rho_{ji}^{n_{ji}}}{n_{ji}!} \right) \right]. \tag{128}$$

10.2 Integral representation

Using Euler's integral, see (10),

$$G = \int_0^\infty \cdots \int_0^\infty \exp\left(- \sum_{i=1}^{q} u_i \right) \sum_{1'\mathbf{n}_1 = K_1} \cdots \sum_{1'\mathbf{n}_p = K_p} \left[\prod_{i=1}^{q} \left(\prod \frac{(\rho_{ji} u_i)^{n_{ji}}}{n_{ji}!} \right) \right]$$
$$\cdot \left[\prod_{i=q+1}^{s} \left(\prod \frac{\rho_{ji}^{n_{ji}}}{n_{ji}!} \right) \right] du_1 \cdots du_q. \tag{129}$$

Now by the multinomial theorem,

$$G = (\prod K_j!)^{-1} \int_0^\infty \cdots \int_0^\infty \exp\left(- \sum_{i=1}^{q} u_i \right)$$
$$\cdot \prod \left(\sum_{i=1}^{q} \rho_{ji} u_i + \sum_{i=q+1}^{s} \rho_{ji} \right)^{K_j} du_1 \cdots du_q. \tag{130}$$

It is noteworthy, but not surprising, that the parameters ρ_{ji} for all the Type 3 centers appear lumped together.

We now introduce the large parameter N and define

$$\beta_j = \frac{K_j}{N}, \qquad 1 \le j \le p, \tag{131}$$

exactly as in (12). However, we define

$$\gamma_{ji} = \left(\frac{\rho_{ji}}{\sum_{m=q+1}^{s} \rho_{jm}} \right) N, \qquad 1 \le j \le p, \quad 1 \le i \le q, \tag{132}$$

which is reciprocal to the natural extension of the parameters $\{\gamma_j\}$ defined in (13). In common with Section 3.1, the suggestion in the notation is that in the generic large network $\{\gamma_{ji}\}$ and $\{\beta_j\}$ are $O(1)$. A tacit assumption being made is that all job classes are routed through at least one Type 3 (infinite server) center.

On substituting (131) and (132) in (130) and after a change of variables we obtain a form for the partition function, which is summarized below.

Proposition 18:

$$G = \left(N^q \prod_{j=1}^{p} \left[\left(\sum_{i=q+1}^{s} \rho_{ji} \right)^{K_j} / K_j! \right] \right) \int_0^\infty \cdots \int_0^\infty e^{-Nf(\mathbf{z})} d\mathbf{z}, \tag{133}$$

where

$$\mathbf{z} = [z_1, z_2, \cdots, z_q]',$$

$$f(\mathbf{z}) = \mathbf{1}'\mathbf{z} - \sum_{j=1}^{p} \beta_j \log(1 + \Gamma_j'\mathbf{z}),$$

$$\mathbf{1} = [1, 1, \cdots, 1]',$$

$$\Gamma_j = [\gamma_{j1}, \gamma_{j2}, \cdots, \gamma_{jq}]', \qquad 1 \le j \le p. \qquad \square$$

As before, the term in brackets in (133) will typically not be required to be computed.

Future work will consider the expansions appropriate for the computation of the integral in (133).

APPENDIX A

Notation for Asymptotic Expansion, Series Reversion

Asymptotic expansion: A series $\sum_{j=0}^{\infty} a_j/N^j$ is said to be an asymptotic expansion[14,22,23] of a function $F(N)$ if

$$F(N) - \sum_{j=0}^{n-1} a_j/N^j = 0(N^{-n}) \qquad \text{as} \qquad N \to \infty$$

for every $n = 1, 2, \cdots$. We write

$$F(N) \sim \sum_{j=0}^{\infty} a_j/N^j.$$

The series itself may be either convergent or divergent.

Series reversion: If $u = f(z)$, $u_0 = f(z_0)$, $f'(z_0) \neq 0$, then by Lagrange's expansion[13,25]

$$z = z_0 + \sum_{j=1}^{\infty} \frac{(u - u_0)^j}{j!} \left[\frac{d^{j-1}}{dz^{j-1}} \left(\frac{z - z_0}{f(z) - u_0} \right)^j \right]_{z=z_0}. \tag{134}$$

In particular, if $f(\cdot)$ is specified in a Taylor series, the above expansion identifies the coefficients $\{g_j\}$ in the reversed power series $z - z_0 = \sum_{j=1}^{\infty} g_j(u - u_0)^j$. We specifically identify the leading coefficients of two reversed series that have been used in Section 4.1 and Section 7.1. If

$$u = f_1 z + f_2 z^2 + \cdots, \tag{135}$$

then

$$z = g_1 u + g_2 u^2 + \cdots, \tag{136}$$

where

$$f_1 g_1 = 1,$$

$$f_1^3 g_2 = -f_2,$$
$$f_1^5 g_3 = 2f_2^2 - f_1 f_3,$$
$$f_1^7 g_4 = 5f_1 f_2 f_3 - f_1^2 f_4 - 5f_2^3,$$
$$f_1^9 g_5 = 6f_1^2 f_2 f_4 + 3f_1^2 f_3^2 + 14f_2^4 - f_1^3 f_5 - 21f_1 f_2^2 f_3.$$

Similarly, if

$$u = f_2 z^2 + f_3 z^3 + f_4 z^4 + \cdots, \tag{137}$$

then

$$z = g_1 u^{1/2} + g_2 u + g_3 u^{3/2} + \cdots, \tag{138}$$

where

$$f_2^{1/2} g_1 = 1,$$
$$f_2^2 g_2 = -f_3/2,$$
$$f_2^{7/2} g_3 = (5f_3^2 - 4f_2 f_4)/8,$$
$$f_2^5 g_4 = (3f_2 f_3 f_4 - f_2^2 f_5 - 2f_3^3)/2,$$
$$f_2^{13/2} g_5 = (-64f_2^3 f_6 + 224f_2^2 f_3 f_5 + 112f_2^2 f_4^2 - 504f_2 f_3^2 f_4 + 231f_3^4)/128.$$

APPENDIX B

Proof of Proposition 6

Before giving the proof, it is worthwhile to generate expressions for the leading derivatives of $g(\cdot)$. In the notation of Section 4.1, $u = f(z)$ and $z = g(u)$,

$$g^{(1)}(u) = \frac{1}{f^{(1)}(z)},$$
$$g^{(2)}(u) = -f^{(2)}(z)/(f^{(1)}(z))^3,$$
$$g^{(3)}(u) = [-f^{(3)}(z)f^{(1)}(z) + 3(f^{(2)}(z))^2]/(f^{(1)}(z))^5,$$
$$g^{(4)}(u) = [-f^{(4)}(z)(f^{(1)}(z))^2 + 10f^{(1)}(z)f^{(2)}(z)f^{(3)}(z) - 15(f^{(2)}(z))^3]/(f^{(1)}(z))^7. \tag{139}$$

For notational convenience, let γ_i and ϕ_i denote, in this appendix only, $g^{(i)}(u)$ and $f^{(i)}(z)$ respectively. Recall that $\phi_1 > 0$ and that $(-1)^i\phi_i > 0$ for $i = 2, 3, \cdots$. We will show by induction that $(-1)^n\gamma_n < 0$, $n = 1, 2, 3, \cdots$.

Let the induction hypothesis be the following

$$n \text{ even: } \phi_1^{2n-1}\gamma_n = \cdots - (\phi_1)^{2i}(X) + (\phi_1)^{2i+1}(Y) \cdots, \tag{140}$$

where $0 \le i$; $\max[2i, 2i+1] < 2n - 1$; X and Y are arbitrary products of $\phi_2, \phi_3, \cdots, \phi_n$ with arbitrary positive numerical weight and the property that $X > 0$, $Y < 0$ for all $z \ge 0$.

$$n \text{ odd: } \phi_1^{2n-1}\gamma_n = \cdots + (\phi_1)^{2i}(U) - (\phi_1)^{2i+1}(V) \cdots , \tag{141}$$

where U and V are like X and Y including that $U > 0$, $V < 0$ for all $z \geq 0$.

For the proof, take the case of n even, first. A key observation is that since ϕ_1 does not occur in either X or Y, $dX/dz < 0$ and $dY/dz > 0$. That is, in common with the functions $\phi_2, \phi_3, \cdots$ the derivative has opposite sign from the function. Also, from (140)

$$\gamma_n = \cdots - \frac{X}{\phi_1^{2n-2i-1}} + \frac{Y}{\phi_1^{2n-2i-2}} \cdots . \tag{142}$$

Differentiate with respect to u,

$$\gamma_{n+1} = \cdots - \frac{1}{\phi_1}\left(\frac{-X(2n-2i-1)\phi_2}{\phi_1^{2n-2i}} + \frac{X'}{\phi_1^{2n-2i-1}}\right)$$
$$+ \frac{1}{\phi_1}\left(\frac{-Y(2n-2i-2)\phi_2}{\phi_1^{2n-2i-1}} + \frac{Y'}{\phi_1^{2n-2i-2}}\right) + \cdots .$$

Hence,

$$\phi_1^{2n+1}\gamma_{n+1} = \cdots + (\phi_1)^{2i}(X(2n-2i-1)\phi_2)$$
$$- (\phi_1)^{2i+1}(X' + Y(2n-2i-2)\phi_2) + (\phi_i)^{2i+2}(Y') \cdots \tag{143}$$

which has the form appearing in (141) as part of the hypothesis.

Now take the case of n odd where, from (141),

$$\gamma_n = \cdots + \frac{U}{\phi_1^{2n-2i-1}} - \frac{V}{\phi_1^{2n-2j-2}} \cdots . \tag{144}$$

Differentiate and rearrange to obtain,

$$\phi_1^{2n+1}\gamma_{n+1} = \cdots - (\phi_1)^{2i}(U(2n-2i-1)\phi_2)$$
$$+ (\phi_1)^{2i+1}(U' + V(2n-2j-2)\phi_2) - (\phi_1)^{2i+2}(V') \cdots .$$

As $U' < 0$, $V' > 0$, the form in (140) of the induction hypothesis is satisfied. This concludes the proof.

APPENDIX C

Proof of Proposition 10

In the following we shall need the sign property of the derivatives of $f(\cdot)$ as given in (25) and (26). Also recall $f_1 = f'(0) = -\alpha$

$$\int_0^\infty ze^{-Nf(z)}dz \le \int_0^\infty ze^{-Nf_1 z} = 1(N\alpha)^2 \tag{145}$$

$$\ge \int_0^\infty ze^{-N(f_1 z+f_2 z^2)}dz$$

$$= \frac{\alpha}{4f_2}\sqrt{\frac{\pi}{Nf_2}}\, e^{\alpha^2 N/4f_2}(1 - erf(-\alpha\sqrt{N/4f_2})) + \frac{1}{2Nf_2}. \tag{146}$$

Similarly,

$$\int_0^\infty e^{-Nf(z)}dz \le -1/(\alpha N) \tag{147}$$

$$\ge \sqrt{\frac{\pi}{4Nf_2}}\, e^{\alpha^2 N/4f_2}(1 - erf(-\alpha\sqrt{N/4f_2})). \tag{148}$$

Thus,

$$\frac{\int_0^\infty ze^{-Nf(z)}dz}{\int_0^\infty e^{-Nf(z)}dz} \le \frac{\sqrt{4f_2/\pi}}{\alpha^2 N^{3/2}} \frac{e^{-\alpha^2 N/4f_2}}{[1 - erf(-\alpha\sqrt{N/4f_2})]}$$

$$\le \frac{1 + \sqrt{1 + 8f_2/\alpha^2 N}}{2|\alpha|N}, \tag{149}$$

where, to bound the term dependent on the error function, we have used the left inequality in the following[13]

$$(x + \sqrt{x^2 + 2})^{-1} < e^{x^2}\int_x^\infty e^{-y^2}dy < [x(1 + \sqrt{1 + 4/(\pi x^2)})]^{-1}. \tag{150}$$

We obtain from (146), (147) and (150) the results analogous to (149):

$$\frac{\int_0^\infty ze^{-Nf(z)}dz}{\int_0^\infty e^{-Nf(z)}dz} \ge \frac{|\alpha|}{2f_2}\left(1 - \frac{2}{1 + \sqrt{1 + 16f_2/(\pi\alpha^2 N)}}\right). \tag{151}$$

Equations (149) and (151) taken together with the representation for the utilization given in (22) is the content of Proposition 10.

APPENDIX D

Proof of Proposition 11

To prove the proposition, we begin by substituting (54) in the expression for I in (56) to obtain after a change of variable

$$I = (\sqrt{N} + c) \int_0^\infty e^{-(\sqrt{N}+c)\nu} \prod (1 + \nu/(a_i \sqrt{N}))^{b_i N + d_i \sqrt{N}} \, d\nu. \quad (152)$$

We may write (152) as

$$I = \int_0^\infty e^{-A\nu^2/2 - B\nu} h(\nu, \sqrt{N}) d\nu, \quad (153)$$

where

$$A = \sum b_i/a_i^2, \; B = c - \sum d_i/a_i \quad (154)$$

and

$$h(\nu, \sqrt{N}) = (\sqrt{N} + c) \exp[A\nu^2/2 + B\nu - \nu(\sqrt{N} + c)][\prod(1 + \nu/(a_i \sqrt{N}))^{b_i N + d_i \sqrt{N}}]. \quad (155)$$

The quantity $h(\nu, \sqrt{N})$ has been defined so that when the second bracketed expression $[\cdot]$ is written as $\exp(\log [\cdot])$ and then the log $[\cdot]$ term expanded, a cancellation of the leading terms is effected by multiplication with the exponential term in (155). In this step, notice has to be taken of the constraint in (55) in that $\exp[A\nu^2/2 + B\nu - \nu(\sqrt{N} + c)] = \exp[A\nu^2/2 - \nu(\sqrt{N} \sum b_i/a_i + \sum d_i/a_i)]$. In this manner we obtain,

$$h(\nu, \sqrt{N}) = (\sqrt{N} + c)\exp\left(\sum_{j\geq 1} F_j(\nu)/\sqrt{N}^j\right), \quad (156)$$

where

$$(-1)^j F_j(\nu) = \left(\frac{1}{j+1} \sum \frac{d_i}{a_i^{j+1}}\right)\nu^{j+1} - \left(\frac{1}{j+2} \sum \frac{b_i}{a_i^{j+2}}\right)\nu^{j+2}. \quad (157)$$

In (156) we have an exponential of a power series in $1/\sqrt{N}$ which may equivalently be represented directly as a power series in $1/\sqrt{N}$:

$$h(\nu, \sqrt{N}) = (\sqrt{N} + c) \sum_{j=0}^{\infty} G_j(\nu)/\sqrt{N}^j. \quad (158)$$

For example, $G_0 = 1$, $G_1 = F_1$, $G_2 = F_1^2/2 + F_2$, $G_3 = F_1^3/6 + F_1 F_2 + F_3$. A feature to observe is that $G_j(\nu)$ is a polynomial of degree $3j$ in ν.

For the final form,

$$h(\nu, \sqrt{N}) = \sum_{j=0}^{\infty} H_j(\nu)/\sqrt{N}^{j-1}, \quad (159)$$

where $H_j(\nu) = G_j(\nu) + cG_{j-1}(\nu)$, also a polynomial of degree $3j$ in ν. It is understood that $G_{-1}(\nu) = 0$.

Insertion of (159) in (153) yields Proposition 11, namely

$$I = \sum_{j=0}^{\infty} c_j/\sqrt{N}^{j-1}, \tag{160}$$

where
$$c_j = \int_0^{\infty} e^{-A\nu^2/2-B\nu} H_j(\nu)d\nu. \tag{161}$$

REFERENCES

1. F. P. Kelly, *Reversibility and Stochastic Networks*, New York: John Wiley, 1980, Chapter 3.
2. L. Kleinrock, *Queueing Systems*, New York: John Wiley, 1976, Vol. II, *Computer Applications*, Chapter 4.
3. M. Schwartz, *Computer-Communication Network Design and Analysis*, Englewood Cliffs, N.J.: Prentice-Hall, 1977.
4. F. Baskett et al., "Open, Closed and Mixed Networks of Queues with Different Classes of Customers," J. ACM, *22*, No. 2 (April 1975), pp. 248–60.
5. F. R. Moore, "Computational Model of a Closed Queueing Network with Exponential Servers," IBM J. Res. Dev. (1972), pp. 567–572.
6. J. Buzen, "Queueing Network Models of Multiprogramming," Ph.D. Thesis, Harvard University, 1971.
7. R. R. Muntz and J. Wong, "Asymptotic Properties of Closed Queueing Network Models," Proc. 8th Annual Princeton Conf. Info. Sci. and Systems, March 1974, pp. 348–52.
8. D. P. Gaver and G. S. Shedler, "Approximate Models for Processor Utilization in Multiprogrammed Computer Systems," Siam J. Comput. *2* (1973), pp. 183–92.
9. M. Reiser, "A Queueing Network Analysis of Computer Communication Networks with Flow Control," IEEE Trans. Commun., *COM-27*, No. 8 (August 1979), pp. 1199–1209.
10. S. S. Lam and M. Reiser, "Congestion Control of Store-and-Forward Networks by Input Buffer Limits—An Analysis," IEEE Trans. Commun., *COM-27*, No. 1 (January 1979), pp. 127–34.
11. B. Pittel, "Closed Exponential Networks of Queues with Saturation: The Jackson-Type Stationary Distribution and Its Asymptotic Analysis," Math. Oper. Res., *4*, No. 4 (Nov. 1979), pp. 357–78.
12. E. Arthurs and B. W. Stuck, "A Theoretical Performance Analysis of a Markovian Switching Node," IEEE Trans. Commun., *COM-26*, No. 11 (November 1978), pp. 1779–84.
13. *Handbook of Mathematical Functions*, eds. M. Abramowitz and I. A. Stegun, New York: Dover Pub., 1970, Chapter 19.
14. A. Erdelyi, *Asymptotic Expansions*, New York: Dover Publications, 1956.
15. M. Reiser and S. S. Lavenberg, "Mean Value Analysis of Closed Multichain Queueing Networks," J. Assoc. Comput. Mach., *27* (April 1980), pp. 313–22.
16. S. S. Lavenberg, "Closed Multichain Product Form Queueing Networks with Large Population Sizes," IBM Research Report, RC 8496 (#36973), September 30, 1980.
17. J. P. Buzen, "Computational Algorithms for Closed Queueing Networks with Exponential Servers," Commun. ACM, *16* (1973), pp. 527–31.
18. M. Reiser and H. Kobayashi, "Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms," IBM J. Res. Dev. (May 1975), pp. 285–94.
19. C. H. Sauer and K. M. Chandy, *Computer Systems Performance Modeling, A Primer*, Englewood Cliffs, New Jersey: Prentice Hall, 1981.
20. J. Zahorjan, "An Exact Solution Method for the General Class of Closed Separable Queueing Networks," 1979 Conf. on Simulation, Measurement, and Modeling of Computer Systems, pp. 107–12.
21. G. Balbo, S. C. Bruell, and H. D. Schwetman, "Customer Classes and Closed Network Models—A Solution Technique," *Proc. IFIP Congress 77*, Amsterdam: North Holland, pp. 559–64.

22. E. T. Copson, *Asymptotic Expansions*, Cambridge: Cambridge University Press, 1965, Chapters 5, 6.
23. F. W. J. Olver, *Introduction to Asymptotics and Special Functions*, New York: Academic Press, 1974, Chapter 3.
24. B. Friedman, "Stationary Phase with Neighboring Critical Points," J. Soc. Indust. Appl. Math., *7*, No. 3 (Sept. 1959), pp. 280–289.
25. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge: Cambridge University Press, 1958, Chapter 16.
26. W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, New York: Springer-Verlag, 1966, Chapter 8.
27. N. Bleistein, "Uniform Asymptotic Expansions of Integrals with Stationary Point Near Algebraic Singularity," Commun. Pure Appl. Math., *19*, No. 4 (1966), pp. 353–70.