July, 1983 LIDS-P-1308
Revised September, 1983

AN EFFICIENT DECOMPOSITION METHOD FOR THE APPROXIMATE EVALUATION OF PRODUCTION LINES WITH FINITE STORAGE SPACE

by Stanley B. Gershwin*

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Abstract

This paper presents a method for the evaluation of performance measures for a class of tandem queuing systems with finite buffers in which blocking and starvation are important phenomena. These systems are difficult to evaluate because of their large state spaces and because they may not be decomposed exactly.

Keywords: 343 inventory levels and throughput in transfer lines, 570 Markov chain model of transfer lines, 721 reliability and storage, 683 decomposition approximation of queuing networks.

* Stanley B. Gershwin
35-310
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

This research has been supported by the U. S. Army Human Engineering Laboratory under contract DAAK11-82-K-0018. I am grateful for the support and encouragement of Dr. Benjamin E. Cummings.

1. INTRODUCTION

Consider the tandem queuing system in Figure 1. It consists of a series of k servers or machines (M_1, M_2, ..., M_k) separated by queues or buffers (B_1, B_2, ..., B_{k-1}). The buffers are each of finite capacity (N_1, N_2, ..., N_{k-1}). The machines are assumed to spend a random amount of time processing each item. If machine M_i spends an unusually long time on a single item, buffer B_{i-1} will tend to accumulate material and buffer B_i will tend to lose material. If this condition persists, B_{i-1} may become full or B_i may become empty. In that case, machine M_{i-1} is <u>blocked</u> and prevented from working or M_{i+1} is <u>starved</u> and also prevented from working.

The purpose of this paper is to present an approximation method for calculating the production rate and the average amounts of material in the buffers for a class of systems of this type. The class includes those in which the service process is deterministic but geometrically unreliable. That is, while a machine is operational and neither starved or blocked, a fixed amount of time is required to process a part. It is assumed that this time is the same for all machines and is taken as the time unit. During a time unit when machine M_i is operational and neither starved nor blocked, it has probability p_i of failing (so that the MTBF, the mean time between failures in <u>working</u> time, is $1/p_i$). After a machine has failed, it is under repair and it has probability r_i of being repaired during a time unit. (Its MTTR, its mean time to repair, is $1/r_i$. This is measured in clock time, not in working time.)

A detailed description of the mathematical model appears in Gershwin and Schick [5]. The model is based on that of Buzacott [2], [3]. The concept of approximate decomposition of tandem queuing models was discussed by Hillier and Boling [6], Takahashi *et al*. [10], Altiok [1], and others. Closely related ideas are discussed by Jafari [8]. Simulation results for models of this type appear in Ho *et al*. [7] and Law [9].

The problem is difficult because of the great dimensionality

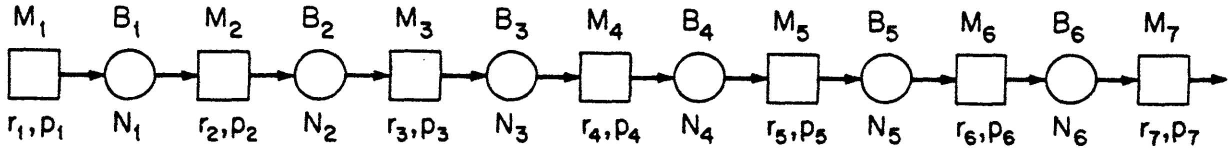

FIGURE 1: The first seven machines and buffers of transfer line L.

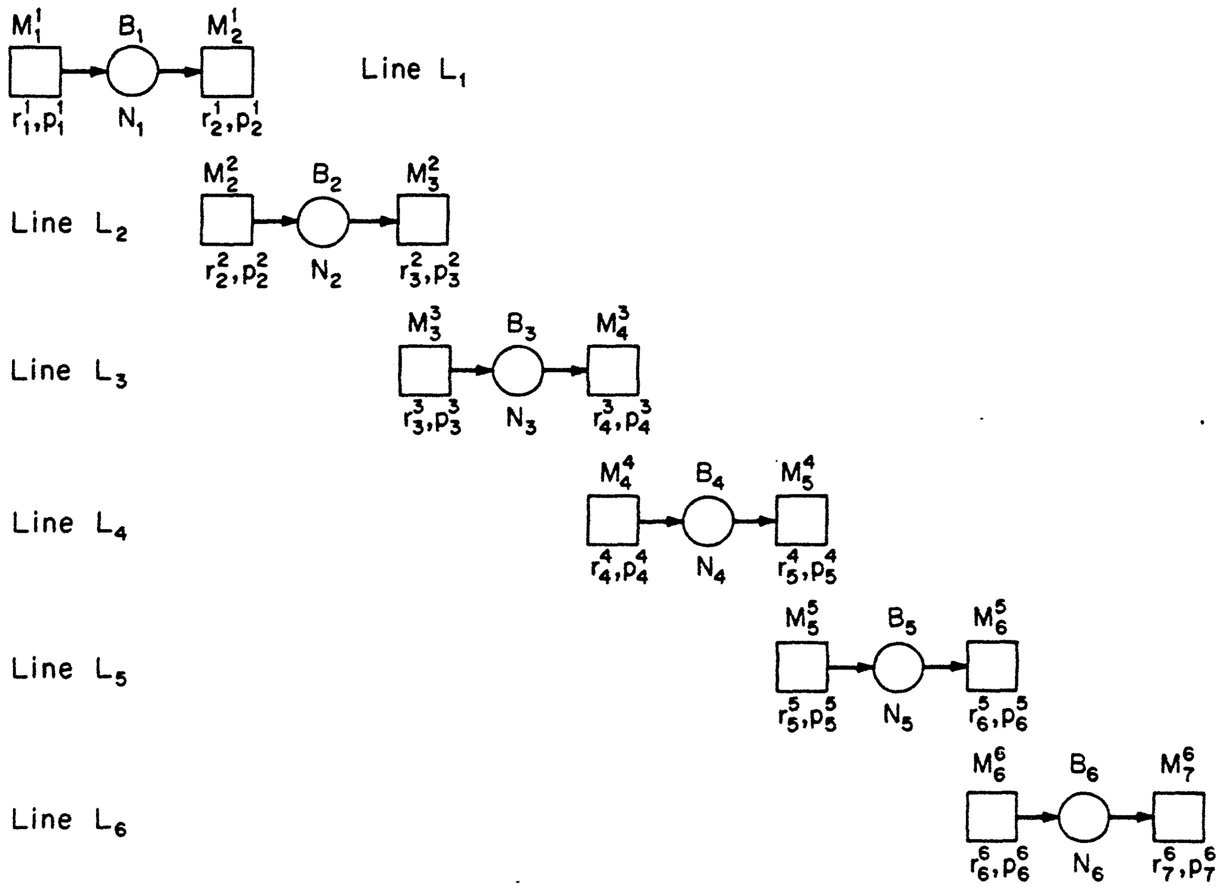

FIGURE 2: A set of two-machine lines.

of the state space. Each machine can be in two states: operational or under repair. Buffer B_i can be in N_i+1 states: $n_i=0, 1,\ldots,N_i$, where n_i is the amount of material in B_i. As a consequence, the Markov chain representation of a 20-machine line with 19 buffers each of capacity 10, for example, has over 6.41×10^{25} states.

2. TRANSFER LINE CHARACTERISTICS

Certain quantities are defined and relationships among them are described in this section. Approximations of them are used below to develop the decomposition method.

Two performance measures of great interest to designers of production lines are production rate E_i (ie, throughput or line efficiency) and average buffer level $\bar{n}_i$ (in-process inventory or work-in-process) at each buffer. The efficiency of machine M_i, in parts per time unit, is

$$E_i = \text{prob} [\ B_{i-1} \text{ not empty } \underline{\text{and}}\ M_i \text{ operational } \underline{\text{and}}\ B_i \text{ not full}\ ]$$

Formulas for these and related quantities for two-machine lines are presented in the Appendix.

Conservation of Flow

Because there is no mechanism for the creation or destruction of material, flow is conserved, or

$$E = E_1 = E_2 = \ldots\ E_k . \tag{1}$$

The Flow Rate-Idle Time Relationship

Define e_i to be the isolated production rate of machine M_i. It is what the production rate of M_i would be if it were never impeded by other machines or buffers. It is given by [3] $e_i = r_i / (r_i + p_i)$ and it represents the fraction of time that M_i is operational. The actual production rate of M_i is less because of blocking or starvation. It is

$$E_i = e_i \, (\text{prob} \, [\, B_{i-1} \text{ not empty and } B_i \text{ not full} \,] \,)$$

which is demonstrated in the Appendix. This expression is approximately

$$E_i = e_i \, (1 - \text{prob} \, [\, B_{i-1} \text{ empty} \,] - \text{prob} \, [\, B_i \text{ full} \,] \,) \tag{2}$$

because it is very unlikely (although not at all impossible) for B_{i-1} to be empty and B_i to be full simultaneously.

3. SOLUTION METHOD

Decomposition

Consider Figure 2, a set of two-machine transfer lines. The buffers of these lines have the same capacities as those of Figure 1. The object is to find the parameters (failure and repair rates r_1^1, p_1^1, r_2^1, p_2^1, r_2^2, p_2^2, etc.) of the machines so that the behavior of the material flow in the buffers of the two-machine lines closely matches that of the flow in the buffers of the long line. That is, the rate of flow into and out of buffer B_i in line L_i approximates that of buffer B_i in the real line. The probability of the buffer of line L_i being empty or full is close to that of the corresponding buffer in the real line being empty or full. The probability of resumption of flow into (and out of) the buffer in line L_i in a time unit after a period during which it was interupted is close to the probability of the corresponding event in the actual line. Finally, the average amount of material in the buffer of line L_i approximates the material level in buffer B_i in the real line under study. In order to find such parameter values, we use the relationships of the previous section as well as others described below.

Machine M_i^i models the part of the line upstream of B_i and M_{i+1}^i models the part of line downstream from B_i. There are four parameters per two-machine line (ie, per buffer in the long line): r_i^i, p_i^i, r_{i+1}^i, p_{i+1}^i. Consequently, 4 equations per buffer, or 4(k-1) conditions, are required to determine them.

Let E(i) be the efficiency or production rate of two-machine line L_i. Then one set of conditions is related to conservation of flow:

$$E(i) = E(1), \quad i=2,...,k-1 \qquad (3)$$

There are k-2 equations here. E(i) is a function of the four unknowns r_i^i, p_i^i, r_{i+1}^i, p_{i+1}^i through the two-machine efficiency formulas in the Appendix.

The second set of conditions follows from (2), the flow rate-idle time relationship. Here we assume that the probability of B_i being empty or full in the original line is closely approximated by the probability of B_i being empty or full in L_i. Consequently,

$$E(i) = e_i\,(1 - p_s(i-1) - p_b(i)), \quad i=2,...k-1 \qquad (4)$$

where $p_s(i-1)$ is the probability of buffer B_{i-1} being empty in the i-1'st two-machine line and $p_b(i)$ is the probability of buffer B_i being full in the i'th line. (The subscripts refer to starvation and blockage.) These quantities are calculated in the Appendix.

Equation (4), after some manipulation, can be written

$$\frac{p_i^{i-1}}{r_i^{i-1}} + \frac{p_i^i}{r_i^i} = \frac{1}{E(i)} + \frac{1}{e_i} - 2,$$

$$i=2,...,k-1. \qquad (5)$$

This is demonstrated in the Appendix.

To characterize the repair rates of the two-machine lines, it is necessary to consider the meaning of failure and repair in those systems. Machine M_i^i in line L_i represents, to buffer B_i, everything upstream of B_i in the long line. Therefore, a failure of M^i_i represents either a failure of machine M_i or the emptying of buffer B_{i-1} (which, in turn, is due to a failure of M_{i-1} or the emptying of B_{i-2}, etc.). The repair of M^i_i is thus the termination of whichever condition was in effect. The probability of repair of M^i_i in any cycle in which it is down is r_i if the actual failure is M_i and it is r_{i-1} or r_{i-2}, etc. if, instead, the "failure" is actually the emptying of B_{i-1}. It is r_{i-1} if B_{i-1} is empty because of the failure of M_{i-1}; it is r_{i-2} if B_{i-1} is empty because M_{i-2} has failed and B_{i-2} has emptied; and so forth.

We assume that the probability of B_{i-1} in the real line being empty, due to all causes, is the same as that of B_{i-1} being empty in the i-1'st two-machine line. In line L_{i-1}, however, B_{i-1} can be empty due only to one cause: the failure of M^{i-1}_{i-1}.

Consequently, the probability of repair of M^i_i is r^{i-1}_{i-1} if the cause of failure is the emptying of B_{i-1} and it is r_i otherwise. This leads to

$$r^i_i = \frac{r^{i-1}_{i-1} \, p_s(i-1) + r_i \, (1 - E(i-1) - p_s(i-1))}{1 - E(i-1)} ,$$

$$i=2,...k-1. \qquad (6)$$

A similiar analysis yields the following equation for the second machine in the i-1'st line:

$$r_i^{i-1} = \frac{r_{i+1}^{i}\, p_b(i) + r_i\,(1 - E(i) - p_b(i))}{1 - E(i)},$$

$$i=2,\ldots,k-1. \qquad (7)$$

Finally, there are boundary conditions:

$$\begin{aligned} r_1^1 &= r_1 \\ r_k^{k-1} &= r_k \\ p_1^1 &= p_1 \\ p_k^{k-1} &= p_k \end{aligned} \qquad (8)$$

There are a total of 4(k-1) equations among (3), (5), (6), (7), and (8) in 4(k-1) unknowns: r_i^i, p_i^i, r_{i+1}^i, p_{i+1}^i, $i=1,\ldots,k-1$.

<u>Numerical Technique</u>

These equations can be thought of as defining a two-point boundary value problem (TPBVP) of the form

$$f(x_{i-1}, x_i) = 0$$

where x_i is a 4-vector of the parameters of line L_i; $x_i = (r_i^i, p_i^i, r_{i+1}^i, p_{i+1}^i)$. The nonlinear function f( ) involves the evaluation of E(i), $p_s(i)$, and $p_b(i)$ by means of the two-machine line formulas of the Appendix.

Satisfactory results have been obtained with a modified shooting method consisting of three nested loops. It is described in detail in [4].

The average buffer levels of the long line are simply those of the two-machine lines when convergence is reached.

4. NUMERICAL RESULTS

For a three-machine line, it is possible to compare the results of this algorithm with exact results by using the method of [5]. A set of 5 cases are compared in [4]. The greatest discrepancy in production rate is 0.5%. The greatest difference in average buffer levels is 7.1%. No more than 70 evaluations of a two-machine line are required for these three-machine cases.

Exact methods are not available for systems of more that three machines and two buffers or for three-machine cases with very large buffers. Consequently, other techniques are required to assess the accuracy of the approximation. They include simulation and qualitative observations. A large number of cases are considered in [4] which cover a wide range of failure probabilities, repair probabilities, and buffer sizes. The results also cover a wide range of production rates and average buffer levels.

There is close agreement between the approximation results and the simulation results. In most cases, production rates and buffer levels agree to within a few percent. This remains true even for large buffer capacities (over 100) and long lines (20 machines.) There is no obvious trend indicating that the accuracy of the approximation decreases as the line length increases.

The number of evaluations of the two-machine line increases with the length of the line. The number of evaluations appears to be less than approximately $2k^3$, where k is the number of machines. As a consequence, the computer time for the analytic approximation method is much less than that of simulation. For example, two 20-machine cases took about 7 and 12 seconds while simulations required from 248 to 262 seconds. The computer time is that of the MIT Honeywell Multics computer.

Three of the cases come from Ho, Eyler, and Chien [7]. Our approximate production rates are in good agreement with their simulation results. Several other cases are taken from Law [9] and again there is close agreement with the simulation results in the literature.

5. CONCLUSIONS AND FURTHER RESEARCH

A new method has been found for the analysis of tandem queuing systems with finite buffers in which blocking is important. Exact and simulation results indicate that the method, while approximate, is quite accurate. Current research is aimed at extending this work in two directions: other service processes, such as reliable and unreliable machines with exponential processing time; and assembly/disassembly networks. Future efforts will be devoted to systems such as Jackson-like networks with blocking.

REFERENCES

[1] T. Altiok, "Approximate Analysis of Exponential Tandem Queues with Blocking," European Journal of Operations Research, Vol. 11, 1982.

[2] J. A. Buzacott, "Markov Chain Analysis of Automatic Transfer Lines with Buffer Stock," Ph.D. Thesis, Department of Engineering Production, University of Birmingham, 1967.

[3] J. A. Buzacott, "Automatic Transfer Lines with Buffer Stocks," International Journal of Production Research, Vol. 6, 1967.

[4] S. B. Gershwin, "An Efficient Decomposition Method for The Approximate Evaluation of Tandem Queues with Finite Storage Space and Blocking, "Massachusetts Institute of Technology Laboratory for Information and Decision Systems Report LIDS-P-1309, October, 1983.

[5] S. B. Gershwin and I. C. Schick, "Modeling and Analysis of Three-Stage Transfer Lines with Unreliable Machines and Finite Buffers," Operations Research, Vol. 31, No.2, pp 354-380, March-April 1983.

[6] F. S. Hillier and R. W. Boling, "The Effect of Some Design Factors on the Efficiency of Production Lines with Variable Operation Times," Journal of Industrial Engineering, Vol. 17, No. 12, December, 1966.

[7] Y. C. Ho, M. A. Eyler, and T. T. Chien, "A Gradient Technique for General Buffer Storage Design in a Production Line," International Journal of Production Research, Vol. 17, No. 6, pp 557-580, 1979.

[8] M. A. Jafari, Ph. D. Thesis proposal, Syracuse University, June, 1982.

[9] S. S. Law, "A Statistical Analysis of System Parameters in Automatic Transfer Lines," International Journal of Production Research, Vol. 19, No. 6, pp 709-724, 1981.

[10] Y. Takahashi, H. Miyahara, and T. Hasegawa, "An Approximation Method for Open Restricted Queuing Networks," Operations Research, Vol. 28, No.3, Part I, May-June 1980.

APPENDIX

<u>Steady-State Probabilities and Performance Measures for Two-Machine Lines</u>

In the following, $p(n, \alpha_1, \alpha_2)$ is the probability that there are n parts in the buffer and that M_i is in state α_i. Here $\alpha_i = 0$ means that the machine is under repair and $\alpha_i = 1$ means that it is operational, ie capable of doing operations on parts (although it may be starved or blocked). These probabilities are taken from [5].

$$p(0,0,0) = 0$$

$$p(0,0,1) = C\,X\,(r_1 + r_2 - r_1 r_2 - r_1 p_2)/(r_1 p_2)$$

$$p(0,1,0) = 0$$

$$p(0,1,1) = 0$$

$$p(1,0,0) = C\,X$$

$$p(1,0,1) = C\,X\,Y_2$$

$$p(1,1,0) = 0$$

$$p(1,1,1) = (C\,X/p_2)(r_1 + r_2 - r_1 r_2 - r_1 p_2)/(p_1 + p_2 - p_1 p_2 - r_1 p_2)$$

$$p(n, \alpha_1, \alpha_2) = C\,X^n\,Y_1^{\alpha_1}\,Y_2^{\alpha_2},\ 2 \le n \le N-2$$

$$p(N-1,0,0) = C\,X^{N-1}$$

$$p(N-1,0,1) = 0$$

$$p(N-1,1,0) = C\,X^{N-1}\,Y_1$$

$$p(N-1,1,1) = (C\,X^{N-1}/p_1)(r_1 + r_2 - r_1 r_2 - p_1 r_2)/(p_1 + p_2 - p_1 p_2 - p_1 r_2)$$

$$p(N,0,0) = 0$$

$$p(N,0,1) = 0$$

$$p(N,1,0) = C\,X^{N-1}(r_1 + r_2 - r_1 r_2 - p_1 r_2)/(p_1 r_2)$$

$$p(N,1,1) = 0$$

where

$$Y_1 = (r_1 + r_2 - r_1 r_2 - r_1 p_2)/(p_1 + p_2 - p_1 p_2 - p_1 r_2)$$

$$Y_2 = (r_1 + r_2 - r_1 r_2 - p_1 r_2)/(p_1 + p_2 - p_1 p_2 - r_1 p_2)$$

$$X = Y_2 / Y_1$$

and C is a normalizing constant. Performance measures are given by

$$E = \sum_{\substack{n \geq 1 \\ \alpha_2 = 1}} p(n, \alpha_1, \alpha_2)$$

$$= \sum_{\substack{n \geq 1 \\ \alpha_2 = 1}} p(n, \alpha_1, \alpha_2)$$

$$p_s = p(0,0,1)$$

$$p_b = p(N,1,0)$$

$$\bar{n} = \sum n \, p(n, \alpha_1, \alpha_2)$$

Proof of the Flow Rate-Idle Time Relationship

This proof follows a similar proof by Gershwin and Berman [11]. By the definition of conditional probability,

$$\text{prob} \, (\alpha_i = 1 \mid n_{i-1} \neq 0, n_i \neq N_i)$$

$$= \frac{E_i}{\text{prob} \, (n_{i-1} \neq 0, n_i \neq N_i)}$$

where production rate E_i is defined verbally in the text and is, in symbols,

$$E_i = \text{prob}\ (\alpha_i = 1,\ n_{i-1} \neq 0,\ n_i \neq N_i).$$

Let

$$D_i = \text{prob}\ (\alpha_i = 0,\ n_{i-1} \neq 0,\ n_i \neq N_i).$$

Then

$$\text{prob}\ (\alpha_i = 1 \mid n_{i-1} \neq 0,\ n_i \neq N_i)$$

$$= \frac{E_i}{E_i + D_i}.$$

Schick and Gershwin [12] observe that

$$r_i D_i = p_i E_i$$

by noting that the left side is the probability of leaving the set of states

$$\{(n_1, n_2, \ldots, n_{k-1}, \alpha_1, \ldots, \alpha_k) \mid$$
$$\alpha_i = 0,\ n_{i-1} \neq 0,\ n_i \neq N_i\}$$

and the right side is the probability of entering that set. Consequently,

$$\text{prob}\ (\alpha_i = 1 \mid n_{i-1} \neq 0,\ n_i \neq N_i)$$

$$= r_i/(r_i + p_i) = e_i$$

and therefore

$$E_i = e_i\ \text{prob}\ (n_{i-1} \neq 0,\ n_i \neq N_i).$$

This result is counter-intuitive because, as a reviewer pointed out, there is no reason to expect that the events of machine failure and adjacent buffers being empty or full are independent. However, failures may occur only while machines are not forced to be idle due to starvation or blockage. Furthermore, B_{i-1} can become empty and B_i can become full only when M_i is operational. Therefore, an idle period can be thought of as a hiatus in which the clock (measuring working time until the next machine state change event) is not running. The fraction of non-idle time that M_i is operational is thus the same as the fraction of time it would be operational if it were not in a system with other machines and buffers.

While it is possible for n_{i-1} to be 0 and n_i to be N_i simultaneously, it is not very likely. The probability of this event is small because such states can only be reached from states in which $n_{i-1} = 1$ and $n_i = N_i - 1$ by means of a transition in which $\alpha_{i-1} = 0$, $\alpha_i = 1$, $\alpha_{i+1} = 0$. The production rate may therefore be approximated by

$$E_i = e_i \; (1 - \text{prob} (n_{i-1} = 0) - \text{prob} (n_i = N_i)).$$

<u>Proof of Equation (5)</u>

In the two-machine case, (2) reduces to

$$E(i) = e_i^i \; (1 - p_b(i))$$

and

$$E(i-1) = e_i^{i-1} \; (1 - p_s(i-1))$$

in which $e_i^i = r_i^i / (r_i^i + p_i^i)$ is the isolated efficiency of machine M_i^i and e_i^{i-1} is the isolated efficiency of machine M_i^{i-1}. Note that these equations are exact, not approximate. They can be written

$$p_b(i) = 1 - E(i) \; / \; e_i^i$$

and

$$p_s(i-1) = 1 - E(i) \; / \; e_i^{i-1} \qquad \text{(since } E(i) = E(i-1)\text{).}$$

Substituting into equation (4),

$$E(i) = e_i \left(E(i)/e_i^{i} + E(i)/e_i^{i-1} - 1 \right).$$

Equation (5) follows after further manipulations using the expressions for the isolated efficiencies in terms of the parameters of the machines.

APPENDIX REFERENCES

[11] S. B. Gershwin and O. Berman, "Analysis of Transfer Lines Consisting of Two Unreliable Machines with Random Processing Times and Finite Storage Buffers," <u>AIIE Transactions</u> Vol. 13, No. 1, March 1981.

[12] I. C. Schick and S. B. Gershwin, "Modelling and Analysis of Unreliable Transfer Lines with Finite Interstage Buffers," Massachusetts Institute of Technology Electronic Systems Laboratory Report ESL-FR-834-6, September, 1978.