Variance of the Output in a Deterministic Two-Machine Line

by

Maria Carrascosa
Ingeniero Industrial
University of Valladolid, Spain (1992)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1995

Author ....
Department of Mechanical Engineering
May 23, 1995

Certified by ....
Stanley B. Gershwin
Senior Research Scientist
Thesis Supervisor

Accepted by ....
Ain A. Sonin
Chairman, Graduate Committee

Variance of the Output in a Deterministic Two-Machine Line

by

Maria Carrascosa

Ingeniero Industrial

University of Valladolid, Spain (1992)

Submitted to the Department of Mechanical Engineering
on May 23, 1995, in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

In this thesis, we explore the variability of the output produced by a deterministic processing time two-machine finite buffer line with unreliable machines. We characterize the interruptions of flow in the system and propose a model to predict the variance of the output in the short term. We identify the correlations between consecutive departures from the line as a main factor in determining the variance of production of a two-machine finite buffer line. We derive some conclusions about the behavior of the variance of the output in the long term, and finally, we explore the changes in the variance of the output as a function of time. The tools used for this work are primarily simulations and previous analytical results.

Thesis Supervisor: Stanley B. Gershwin
Title: Senior Research Scientist

Acknowledgments

First of all, I want to thank my thesis advisor Stanley Gershwin for his encouragement during this past year. I would not have survived my research without having his support. And after a year working together, I can almost pronounce his name properly now.

I also want to thank Dr. Burman for giving me the chance (and the shock) of applying our research to a real manufacturing plant and for being a constant source of energy.

I want to thank Asbjoern for his infinite help. I doubt there is a single way of overloading a computer that I have not explored thoroughly, and he has been patient enough to help me out each time.

Very special thanks for Augusto and Jim for the good time we have had together. They have made me laugh so much I was looking forward to going to work.

I want to say I feel lucky and proud I have got to know them all. It was worth coming all the way from Spain to meet them.

Finally, I want to thank Jose and Rosa for the great and crazy year we have spent together. I also take advantage of the opportunity to apologize for boring them to death with my manufacturing talks.

I want to acknowledge the finantial support and helpful assistance provided by "La Caixa" Fellowship Program. Without this funding I would have never come to the States.

Research in this thesis has been partially supported by NSF Grant DDM-9216358.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

A *manufacturing system* or *production line* is the set of elements (e.g., machines, buffers, transportation) used for the transformation of raw material into a product. The performance of a manufacturing system is conditioned by disruptive events (failures of machines) that randomly occur and penalize the performance of the system. There are two undesirable consequences of these failures: the first is a reduction of the *mean* production, the second is the *variability* of the output produced.

Most of the research on production lines made has focused on the effect of the randomness in the mean production rate and inventory levels of a system. There is an important gap in the research literature on the effect of the randomness in the variability of the production. However, randomness is extremely important on a day-to-day basis. It is not unusual to find variations around 30% of the mean production in the weekly production of a factory. Despite the variability, a manufacturing plant needs to assure consistent delivery of the products to its customers to stay in business. Reducing and dealing with variability of the production is a key factor in the success of a company.

The purpose of this thesis is to explore the variability of the output produced by a deterministic processing time two-machine finite buffer line. This is the simplest non-trivial manufacturing system: the processing time is fixed and equal for all the machines and the failure and repair characteristics of the machines are geometrically distributed. This thesis is meant as a first step in developing an understanding of the variability of the output of a production line that can be used later to analyze more complex systems.

1.2 Approach

1.2.1 Tools

Three basic tools have been used in the development of this thesis:

Study of simpler systems We started the study by understanding the variance of a two-machine zero buffer system, and we built on this system to represent the behavior of the more complex finite buffer system.

Simulations The main tool in this study has been the use of simulations. A simulation package is a computer software that reproduces the events that take place in a production line. Several modifications have been introduced in an available software[1] for the specific purposes of this thesis. The simulations have been used to gain intuition about the behavior of the system, to test hypothesis and to confirm the validity of the results obtained.

Previous analytical results Miltenburg (1987) calculated the limit of the variance of the output per time unit when the time tends to infinite ($\lim_{t\to\infty} \frac{\sigma^2(t)}{t}$). We define this value as the *asymptotic variance rate*. This value has been used to explore the long term behavior of the systems studied.

Some results derived by Buzacott and Shanthikumar (1993) about the equivalence of a two-machine finite buffer line and a two-machine zero buffer line in

[1] described in Section 7.1

terms of their interdeparture distribution have been used to confirm some of the work described in this thesis and to formulate hypotheses about the behavior of the system.

1.2.2 Outline

In Chapter 2 we explain the deterministic two-machine finite buffer model that is used for the study of the variability of the output.

In Chapter 3 we derive analytically the variance of the output for a single machine with two failure modes during a time interval. This result approximates very well the variance of the output of a two-machine zero buffer line.

In Chapter 4 we identify the possible reasons that result in an interruption of the output in a two-machine finite buffer line. We determine the parameters that characterize the frequency and length of these interruptions of flow. These results lead to the representation of the original two-machine finite buffer system by a two-machine zero buffer line with the same characterization of the interruptions of flow. We compare this simplification with a similar one performed by Buzacott and Shanthikumar (1993).

In Chapter 5 we derive some conclusions about the variance of the line output over a long interval, and the changes of this asymptotic variance rate as a function of the buffer size. An analytical result derived by Miltenburg (1987) is used to determine this asymptotic variance rate. Systems with machines having the same and different efficiencies are studied separately.

In Chapter 6 we build a model that predicts the variance of output of a deterministic two-machine finite buffer system during a time interval using the results derived in the previous chapters. Comparison with simulation shows that the model predicts the variance accurately only for short time intervals. We identify the correlation structure between consecutive departures of the line as the factor that is not accounted for in the model and that modifies the output pattern of a two-machine finite buffer line.

In Chapter 7 we perform simulations for a wide variety of two-machine finite buffer lines to evaluate the behavior of the variance rate of the output of a system as a function of time. We identify three different time frames, in each of which the variance is shaped by a different mechanism. We make some observations about the difference in the behavior of a system and its reverse.

1.3 Literature Review

This literature review is restricted to the previous work done on the variability of a production line.

Miltenburg (1987) derives a formula for the variance of the output per time unit produced by a line with finite buffers as the time tends to infinity. Miltenburg uses the Markov chain properties of the transfer line to derive this result. The derivation is done for a deterministic processing time model, with geometric failures and repairs. The method is not restricted to a two-stage line, but it is very computationally intensive and this limits the dimensions of the system to which it can be applied.

Ou and Gershwin (1989) obtain a closed form expression for the variance of the lead time of a two-machine finite buffer line. The Laplace-Stiejes transform is used to derive this result from the steady state probability distribution. The derivation is obtained for deterministic, exponential and continuous processing time models with geometric/exponential failures and repairs.

Heindricks (1992) develops analytical expressions for the steady state interdeparture distribution of a line and the correlation structure of this distribution. Heindrick also uses Markov chain properties to derive this result. The derivation is done for exponential processing time, finite buffer line, perfectly reliable machines. Like Miltenburg's result it would be very computationally intensive for more complex systems.

Buzacott and Shanthikumar (1993) provide a formula for the steady state interdeparture distribution of a system. The derivation is done for a deterministic two-machine line, with geometric failures and repairs. They also obtain the characteristics of a two-machine zero buffer line with exactly the same interdeparture distribution.

Gershwin (1993) obtains an expression for the variance of the output of a single deterministic processing time machine with geometric failures and repairs. The variance is calculated as a function of the time period over which the system is observed.

Chapter 2

Deterministic Processing Time Model

2.1 Description

Deterministic Processing Time By *deterministic line* we mean a system consisting of machines with deterministic processing times. These machines fail and are repaired at random, so the system is not actually deterministic.

This model is described in detail in Gershwin (1994). It is based on the model of Buzacott (1967). This model is formed by two machines with a buffer between (Fig. 2-1). The length of time that parts spend in each machine is fixed, known in advance and the same for both machines. This time is taken to be the time unit. All operational machines start their operations at the same instant. A machine is operational when it is not under repair, starved (when there is no material in the preceding buffer), or blocked (when the following buffer is full). The first machine (M_1) is never starved and the last machine (M_2) is never blocked.

Failure and Repair Distribution Failures are considered to be operation dependent. This means that if a machine is not working, due to starvation or blockage, it

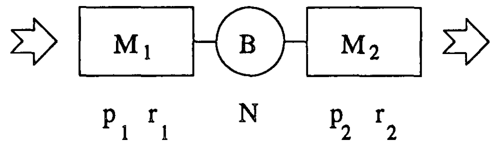

Figure 2-1: Representation of a two-machine line with a buffer

cannot fail. During a time unit when M_i is working, it has probability p_i of failing. Its mean time to fail (MTTF) in *working* time is thus $1/p_i$.

During a time unit when M_i is under repair it has a probability r_i of being repaired. Its mean time to repair (MTTR) is thus $1/r_i$.

Both the failure and repair distributions are assumed to be geometric. One of the main features of this distribution is that it is *memoryless*. This means that the probability of a machine failing (or being repaired) in a time unit is independent of the number of units it has been working (or been down) previously.

Buffer The buffer is a storage element. It is characterized by the total number of parts it can contain (N). Parts pass through the buffer with a transportation delay that is very small compared with service times in the machines (except for the delay caused by other parts in the buffer). A piece spends at least one time unit in the buffer before being transferred from one machine to the next.

Conventions There are two conventions that are made about the operations of the line:

- A machine fails or gets repaired at the beginning of a time unit. This means that a machine that has been repaired must produce at least one part before the next failure. A machine that has failed spends at least one time unit down before being repaired.
- The buffer level changes at the end of each time unit. It depends on the state of the machines at the beginning of the time unit and of the level of the buffer

in the previous time unit.

2.2 Representation of the Two-Machine Line

Markov Chain We use a Markov chain to model the line. The states are

$s = (n, \alpha_1, \alpha_2)$, where

$n = 0, 1, ..., N$ is the number of parts in the buffer

$\alpha_i = 0, 1$ is the state of M_i

0 = DOWN

1 = UP (includes starvation and blockage)

We study the system in steady state. Steady state means that the probability of the system being in a given state does not depend on the conditions in which it started. We determine the probability of finding the system in each of these states.

Steady State Distribution The steady state probability distribution satisfies:

$$\mathbf{p} = \mathbf{p}P \tag{2.1}$$

where $\mathbf{p}$ is the steady state probability vector and P is called the *transition matrix.*

The derivation of the transition matrix and the solution of the steady state probability vector is presented in Gershwin (1994).

Performance Measures Two of the most important performance measures of the system in steady state are the production rate (E) and the average buffer level ($\overline{n}$). They are obtained from the steady state distribution as follows:

Efficiency *The efficiency of a single machine (M_i)* is the percentage of time that it is working when isolated. The efficiency of an isolated machine is:

$$e_i = \frac{r_i}{r_i + p_i} \tag{2.2}$$

The efficiency of a two-machine system is the average number of parts that come out of (or into) the system in a period of time. It is equivalent to the production rate measured in this time unit, and it is calculated as follows:

$$E = \sum_{i=1}^{N} p(i, \alpha_1, 1) = \sum_{i=0}^{N-1} p(i, 1, \alpha_2) \tag{2.3}$$

Two particular two-machine systems are:

$$\text{Zero buffer: } E = \frac{1}{1 + \frac{r_1}{p_1} + \frac{r_2}{p_2}}^{1} \tag{2.4}$$

$$\text{Infinite buffer: } \lim_{N \to \infty} E = \min\,(e_1, e_2) \tag{2.5}$$

Average Buffer Level *Average buffer level* represents the average work in progress in the system. Its formula is:

$$\overline{n} = \sum_{s} n_i \, \mathbf{p}(s) \tag{2.6}$$

Variance of the Output The variance of the output is a way to quantify the discrepancy between the production in a particular period and the average production in steady state. This is a very important measure of the randomness of a system, and very little work has been dome so far to derive it. The work found in the literature use the properties of the Markov chain to derive analytical results.

[1] This formula is an approximation derived by Buzacott (1967). It assumes that the two machines cannot fail at the same time unit. The exact formula taking into account that there might be simultaneous failures is

$$E = \frac{1}{1 + \frac{p_1}{r_1} + \frac{p_2}{r_2} - \frac{p_1 p_2}{r_1 + r_2 - r_1 r_2}}.$$

In this thesis, we try to understand the variance of the output of a system. As we will see, the variance of the output of two-machine zero buffer line is fairly easy to obtain analytically, and it is the existence of a finite buffer that significantly complicates the output process. The difficulty lies on the fact that the buffer accumulates material and in doing so keeps track of the events that have taken place in the system. This complicates the pattern in which pieces leave the production line, and makes it harder to derive any analytical results.

The properties of the interruption of the output are obtained using the characteristics of the steady state distribution presented here. Other tools, especially simulation, are used to gain understanding of the behavior of the system.

Chapter 3

Variance of the Output for Special Cases

3.1 Introduction

In this chapter, we develop an analytical formula for the variance of a deterministic processing time two-machine line with zero buffer size and for a deterministic single machine. In later chapters we will use these results to compare the variance of this system with that of a deterministic two-machine finite buffer line, and to model the latter system.

First, we describe what we mean by the variance of the output. Then, we state why these special cases are of importance to understand the behavior of a system with finite buffer size. Finally, we derive the variance of the output for a deterministic two-machine line with zero buffer size and for a single machine.

3.2 Variance of the Output

In this section, we describe what we mean by the variance of the output over a period of length t.

The output of a production line during a period of length t ($\theta_i(t)$) is the number of parts that an observer, placed after the last machine of the line, (Fig 3-1) sees come out of the line during t consecutive time units. In the case of a deterministic line, the output at a given time unit will be 1 or 0.

The mean ($\overline{\theta(t)}$) and the variance ($\sigma_\theta(t)$) of the output for a period of time t are two important performance measures of the system. If our observer watches the line for N *independent* periods of length t, the mean and variance of the output can be estimated as follows:

$$\overline{\theta(t)} = \frac{1}{N}\sum_{i=1}^{N}\theta_i(t)$$

$$\sigma_\theta(t) = \frac{1}{N-1}\sum_{i=1}^{N}(\theta_i(t) - \overline{\theta(t)})^2$$

As it has been derived in equation (2.3)

$$\begin{aligned} \text{Efficiency} &= \frac{1}{t}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\theta_i(t) = E \\ \text{Mean Production (t)} &= Et \end{aligned}$$

In the next sections we derive the variance of the output as a function of the length of the period $[0, t]$.

$$\text{Variance of the Output(t)} = \lim_{N\to\infty}\frac{1}{N-1}\sum_{i=1}^{N}(\theta_i(t) - \overline{\theta(t)})^2$$

We assume that we assume that the probability of being in a particular state at the start of a period of observation of the system is the steady state probability of that state. For our calculations, we use the steady state probability as the initial condition to solve our system.

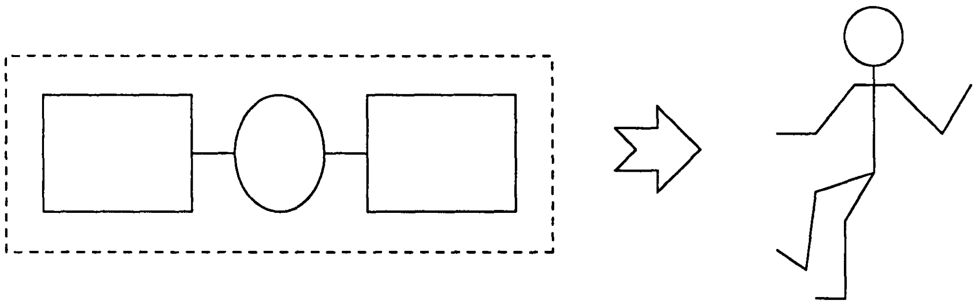

Figure 3-1: Observer watching the output of a production line

3.3 Effect of the Buffer as a Decoupling Element

The purpose of a buffer is to increase the production rate of a system by acting as a decoupling element between the machines. Its effect is to reduce the dependence of each machine's behavior on the other. As long as the buffer is not completely full or empty, a machine is not affected by the failures and repairs of the other one. M_2 is only affected by the performance of M_1 when the buffer empties (*starvation*). In this case, M_2 cannot produce due to lack of raw material. Similarly, M_1 is affected by the performance of M_2 when the buffer fills (*blockage*). In this circumstance, M_1 cannot produce due to lack of space to store the finished parts.

Given two machines, the decoupling effect of the buffer depends on the buffer size. It is reasonable to think that, the bigger the buffer size the greater the decoupling effect, because the probability of a M_2 being starved and M_1 being blocked decrease. Thus, there are two cases where the effect of the buffer is extreme: zero buffer (no decoupling) and infinite buffer (maximum decoupling). For finite buffer sizes, we expect the decoupling effect to be between these two extreme cases.

3.4 Variance of Output of a Deterministic Two-Machine Line with Zero Buffer Size

In this section, we represent the two-machine system with zero buffer size by a simple Markov chain. Using two simplifications, we conclude that this system can be very well approximated by a single machine with two failure modes. As the derivation of the variance of this second system is easier, we will use it to determine the variance of the original system.

3.4.1 Representation of a Two-machine Line with Zero Buffer Size

Conventions We use the conventions explained in Chapter 2. The pieces in the system are represented as being in a buffer. The machines do not hold pieces. They transfer pieces from the upstream to the downstream buffer while performing an operation on them. If a machine fails while performing an operation, the piece it is working on is considered to be in the upstream buffer at the next time unit. If the machine does not fail the piece is in the downstream buffer at the next time unit. As there are two pieces in the system when both machines are working, $N = 2$ is necessary to represent the behavior of a zero buffer system with these conventions. From now on we use the terms zero buffer system and $N = 2$ as equivalent.

Model The behavior of a two-machine line with zero buffers is as follows:

1. When both machines are working, material is coming out of the system.

2. If M_2 fails at time t_f and M_1 keeps working, output stops at time t_f. M_1 is blocked at time $t_f + 1$. If M_2 is repaired at time t_{r2}, there is output at time t_{r2} and M_1 can work (or fail) at time $t_{r2} + 1$.

3. If M_1 fails at time t_f and M_2 keeps working, output stops at time $t_f + 1$. M_2 is starved at time $t_f + 1$. If M_1 is repaired at time t_{r1}, M_2 can work (or fail) at time $t_{r1} + 1$.

4. Both machines can fail at the same time. The way the system resumes normal activity depends on the order of repairs. M_1 is repaired at time t_{r1} and M_2 is repaired at time t_{r2}.

   - $t_{r2} < t_{r1}$: If M_2 is repaired at time t_{r2}, a piece comes out of the system at time t_{r2}. M_2 is starved at time $t_{r2} + 1$. If M_1 is repaired at time t_{r1}, M_2 can work (or fail) at time $t_{r1} + 1$.

   - $t_{r1} < t_{r2}$: If M_1 is repaired at time t_{r1}, M_1 is blocked at time $t_{r1} + 1$. If M_2 is repaired at time t_{r2}, there is output at time t_{r2} and M_1 can work (or fail) at time $t_{r2} + 1$.

   - $t_{r1} = t_{r2} = t_r$: There is output at time t_r.

Steady State Probability Distribution The only non-transient states are $(0, 0, 1)$, $(1, 0, 0)$, $(1, 1, 1)$, and $(2, 1, 0)$. The Markov chain that represents this system is shown in Fig. 3-2.

3.4.2 Simplification : Single Machine with Two Failure Modes

First simplification The state $(1, 0, 0)$ is visited very infrequently (only when M_1 and M_2 fail simultaneously). Therefore, the system that results if we ignore the existence of this state is a good approximation of the system. This approximation results in a simplification of the Markov chain and, as a consequence, a simplification of the derivation of a formula for the variance of the output. In Fig. 3-2 the states and transitions that can be eliminated have been drawn in dashed lines. This simplification can also be explained by saying that $p_1 p_2$ is a second order quantity, very small compared to the rest.

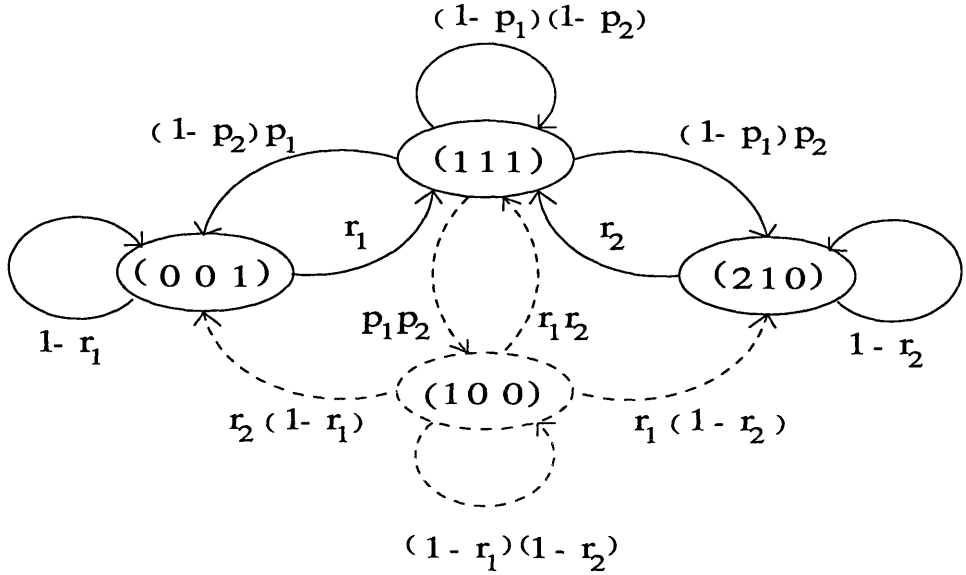

Figure 3-2: Markov chain that represents a two-machine line with zero buffers

Second simplification The system after this simplification is shown in Fig. 3-3. When the system is in state $(1,1,1)$ we assume material is coming out of the system at a rate of one piece per time unit. If the system is in the other states, we assume there is no output. This is not precisely true. The failures and repairs of M_1 are seen by the observer at the end of the line with a *delay* of a unit of time after they take place. If M_1 fails at time t_f and M_2 remains operational, the state of the system at this time unit is $(0,0,1)$ and one part comes out of the system. During the rest of the time M_1 spends in this state, no output emerges. If M_1 is repaired at time t_r the state of the system at this time unit is $(1,1,1)$ but no material comes out of the system.

If we wanted the Markov chain to describe this output process precisely, it should be modified as shown in Fig. 3-4. This discussion is presented in Schick and Gershwin (1978). States $(1,1,1)$ and $(0,0,1)$ in Fig. 3-3 are split into two different states to represent the transitions explained in the previous paragraph. The underlined states represent the productive states (states where material comes out of the system). The

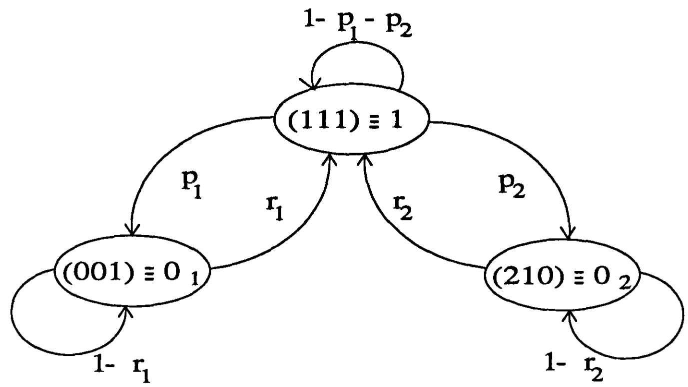

Figure 3-3: Simplified Markov chain of a two-machine line with zero buffer size

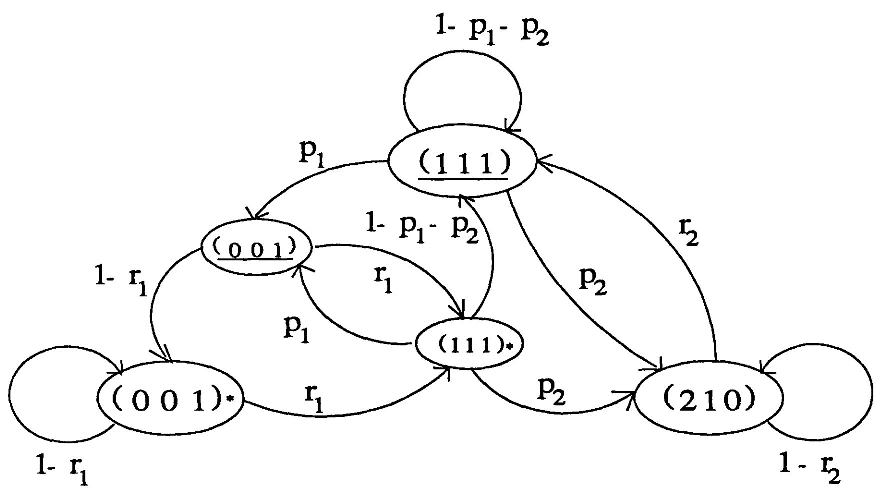

Figure 3-4: Markov chain with productive and non-productive states

probabilities of these states satisfy

$$p(1,1,1) = p\underline{(1,1,1)} + p(1,1,1)^*$$
$$p(0,0,1) = p\underline{(0,0,1)} + p(0,0,1)^*$$

where $p(1,1,1)$ and $p(0,0,1)$ are the steady state probabilities of the system represented in Fig. 3-2. The steady state probabilities are not modified: the system we are describing is the same as the previous one, but we are presenting it in such a way that we can precisely identify the productive and non-productive states. Whenever

M_1 fails, the system spends a time unit in $\underline{(0,0,1)}$ first, and whenever the system leaves state $(0,0,1)^*$ it spends a time unit in $(1,1,1)^*$. So,

$$p\underline{(0,0,1)} = p(1,1,1)^*$$

The mean production rate is the probability of the system being in a productive state. In this case,

$$E = p\underline{(0,0,1)} + \underline{p(1,1,1)} = p(1,1,1)^* + \underline{p(1,1,1)} = p(1,1,1).$$

Therefore, we can conclude that the simplification does not alter the performance of the system in terms of the average production rate. Also, the simplification does not modify significantly the pattern of non-productive and productive periods (and subsequently the variance). As there is a delay of one unit both when M_1 fails and when it gets repaired, one delay compensates the other. Thus, we can conclude that this approximation does not alter significantly the performance measures of the system while it simplifies the process of deriving the variance.

The Markov chain left after these simplifications is identical to the Markov chain that represents the behavior of a single machine with two failure modes[1] (Fig. 3-3). State $(1,1,1)$ is equivalent to state 1 and represents the system is producing pieces at a rate of one per unit time. State $(0,0,1)$ (Failure of M_1) and $(2,1,0)$ (Failure of M_2) are equivalent to states 0_1 (Failure of type 1) and 0_2 (Failure of type 2) respectively. There is no output while the system is in either of these states.

Fig. 3-5 compares the variance of the output of the original two-machine zero buffer system (Fig. 3-2) and of the simplified single machine two failure modes system (Fig. 3-3). Both the variance of the original system and the variance of the simplified system have been calculated analytically [2]. Example 1 graphs show cases were the machines

[1] All through the thesis the term ***single machine with two-failure modes*** refers to two failure modes that cannot happen at the same time

[2] The variance of the simplified system has been calculated using the formula derived in Section 3.5. The variance of the original system has been computed using the same procedure but for the

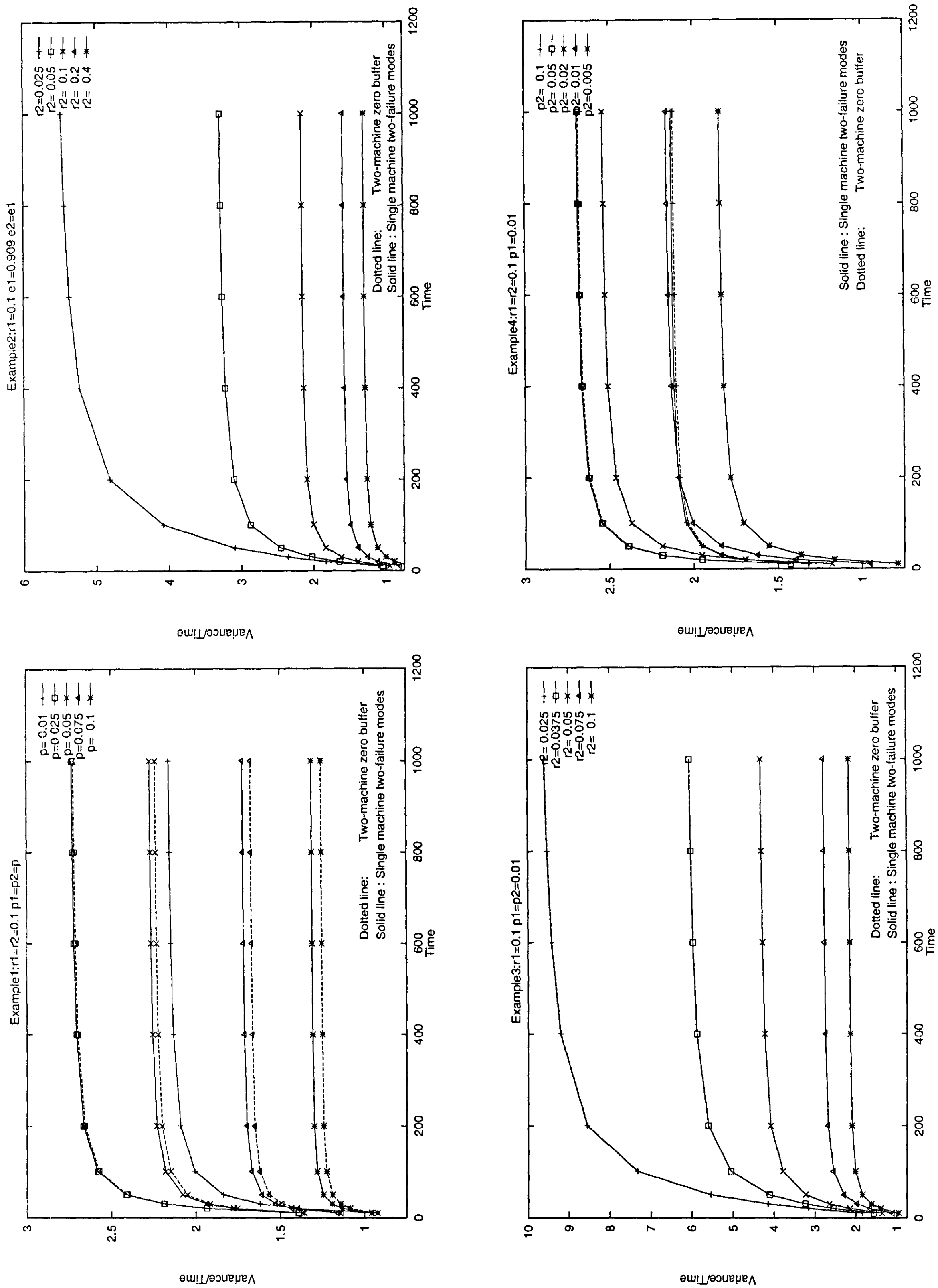

Figure 3-5: Variance of the output: original two-machine zero buffer system vs. simplified single machine with two failure modes

are identical. In Example 2, both machines have the same efficiency. Example 3 shows the effect of modifying r_2 and Example 4 shows the effect of modifying p_2. We can see that it is a very good approximation in all cases. The maximum difference obtained in all the cases tried was less than 3%. We can see that as the machines become more unreliable, the probability that both machines fail at the same time increase, and as a consequence, the simplification works worse.

3.5 Variance of a Single Machine with Two Failure Modes

In the next section, we derive the variance of the output of a single machine with two failures modes, and from now on, we will use this derivation to approximate the variance of the output of a two-machine line with zero buffers.

This derivation is based on the procedure followed by Gershwin (1993) to determine the variance of a single machine with one failure mode.

Model We assume the processing time of the machine is deterministic, that is, it takes a fixed amount of time to complete an operation. This time is the time unit. Both failure modes cannot take place simultaneously. During a time unit when the machine is down, it may either stay down (with probability $1 - r_1$ if it is a failure of type 1 and with probability $1 - r_2$ if it is a failure of type 2), or its repair may be completed (with probability r_1 if it is a failure of type 1 and with probability r_2 if it is a failure of type 2). During a time unit when the machine is operational, it may either perform an operation (with probability $1 - p_1 - p_2$ or it may fail without completing an operation (with probability p_1 the failure is of type 1 and with probability p_2 the failure is of type 2).

original Markov chain. As this system is more complex the variance has been derived numerically using an iterative process.

Steady State Distribution Fig. 3-3 shows the Markov chain that represents this system. In state 1 the machine is producing pieces. In state 0_i, the machine is down with a failure of type i.

The steady state probabilities are:

$$\left[\begin{array}{ccc} \mathbf{p}(1) & \mathbf{p}(0_1) & \mathbf{p}(0_2) \end{array}\right] = \left[\begin{array}{ccc} \mathbf{p}(1) & \mathbf{p}(0_1) & \mathbf{p}(0_2) \end{array}\right] \left[\begin{array}{ccc} 1-p_1-p_2 & p_1 & 0 \\ r_1 & 1-r_1 & p_2 \\ r_2 & 0 & 1-r_2 \end{array}\right]$$

$$\mathbf{p} = \mathbf{p}P \tag{3.1}$$

Probability Distribution of the Output Let $\pi(n, \alpha, t)$ be the probability that the machine is in state α at time t and n pieces have been produced in $[0, t]$. Then,

$$\pi(n, 0_1, t+1) = (1-r_1)\pi(n, 0_1, t) + p_1\pi(n, 1, t) \tag{3.2}$$

$$\pi(n, 0_2, t+1) = (1-r_2)\pi(n, 0_2, t) + p_2\pi(n, 1, t) \tag{3.3}$$

$$\pi(n, 1, t+1) = \sum_{i=1,2} r_i\pi(n-1, 0_i, t) + (1-p_1-p_2)\pi(n-1, 1, t) \tag{3.4}$$

subject to

$$\begin{aligned} \pi(n, \alpha, t) &= 0, t \geq 0, n > 1 & (3.5) \\ \pi(t, 0_i, t) &= 0, t \geq 0, n > 1, \ \forall i & (3.6) \\ \pi(0, 1, t) &= 0, t \geq 0, n > 1 & (3.7) \end{aligned}$$

In addition, as we commented in Section 3.1, we assume that the system starts in steady state. Therefore,

$$\pi(0, 1, 0) = \mathbf{p}_1 = p(1) = \frac{1}{1 + \sum_{i=1,2} \frac{p_i}{r_i}} = C \tag{3.8}$$

$$\pi(0, 0_i, 0) = \mathbf{p}_{0_i} = p(0_i) = \frac{\frac{p_i}{r_i}}{1 + \sum_{i=1,2} \frac{p_i}{r_i}} = \frac{p_i}{r_i} C \quad \forall i \tag{3.9}$$

Formulae for Performance Measures Two performance measures are of interest: the mean and the variance of the number of pieces produced during a specified time interval.

The mean and the variance of the amount of material produced in t time steps are:

$$\begin{aligned} \text{Mean Production :} \quad & \bar{n}(t) = \sum_{n=0}^{t} n(\pi(n, 0_1, t) + \pi(n, 0_2, t) + \pi(n, 1, t)) \\ \text{Variance of the Output:} \quad & \sigma^2(t) = E(n(t)^2) - \bar{n}(t)^2 \end{aligned}$$

where

$$E(n(t)^2) = \sum_{n=0}^{t} n^2(\pi(n, 0_1, t) + \pi(n, 0_2, t) + \pi(n, 1, t))$$

Formulae for the mean and the variance of the production are derived in Appendix A using the difference equations (3.2), (3.3) and (3.4) for $\pi(n, \alpha, t)$. They are

$$\bar{n}(t) = Ct \tag{3.10}$$

$$\begin{aligned} \sigma^2(t) \quad = \quad & (C\frac{r_1^2 p_2(2 - p_2 - r_2) + r_2^2 p_1(2 - p_1 - r_1) - 2p_1 p_2 r_1 r_2}{(p_2 r_1 + r_2 p_1 + r_1 r_2)^3})t \\ & -2C\frac{(2 + b - p_1 - p_2 - r_1 - r_2)((r_2 - r_1)^2 - (p_1 + p_2 - b)^2)}{b(-b + p_1 + p_2 + r_1 + r_2)^3}(1 - \beta_1^t) \\ & -2C\frac{(-2 + b - p_1 - p_2 - r_1 - r_2)((r_2 - r_1)^2 - (p_1 + p_2 - b)^2)}{b(b + p_1 + p_2 + r_1 + r_2)^3}(1 - \beta_2^t) \end{aligned} \tag{3.11}$$

where

$$b^2 = (r_2 - r_1)^2 + (p_1 + p_2)^2 + 2(r_2 - r_1)(p_2 - p_1) \tag{3.12}$$

$$\beta_1 = 1 - \frac{p_1 + p_2 + r_2 + r_1 - b}{2} \tag{3.13}$$

$$\beta_2 = 1 - \frac{p_1 + p_2 + r_2 + r_1 + b}{2} \tag{3.14}$$

Note that $\beta_1, \beta_2 < 1$ and therefore the influence of the terms $(1-\beta_2^t)$, $(1-\beta_1^t)$ in the variance decrease as t increases. For large t,

$$\lim_{t\to\infty} \frac{\sigma^2(t)}{t} = C\frac{r_1^2 p_2(2-p_2-r_2) + r_2^2 p_1(2-p_1-r_1) - 2p_1p_2r_1r_2}{(p_2r_1 + r_2p_1 + r_1r_2)^3}. \tag{3.15}$$

We introduce now the concept of variance rate and asymptotic variance rate:

- *variance rate* is the variance of a period of length t per time unit.

$$\text{variance rate} = \frac{\sigma^2(t)}{t}$$

- *asymptotic variance rate* is the limit of the variance rate when the length t of the period tends to infinity.

$$\text{asymptotic variance rate} = \Delta = \lim_{t\to\infty} \frac{\sigma^2(t)}{t}$$

Then, equation (3.15) can be written

$$\Delta = \lim_{t\to\infty} \frac{\sigma^2(t)}{t} = C\frac{r_1^2 p_2(2-p_2-r_2) + r_2^2 p_1(2-p_1-r_1) - 2p_1p_2r_1r_2}{(p_2r_1 + r_2p_1 + r_1r_2)^3} \tag{3.16}$$

Example Fig. 3-6 shows the comparison between the results obtained by simulation[3] and by analytical derivation. Example 1 graphs show cases were the machines are identical. In Example 2, both machines have the same efficiency. Example 3 shows the effect of modifying r_2 and Example 4 shows the effect of modifying p_2. The shape of the curve variance rate vs. time is common for all zero buffer systems. It is concave and grows asymptotically to reach a limit. This limit has been calculated by

[3]The procedure used in the simulations is described in Section 7.1.

Miltenburg (Section 5.1.1).

3.6 Variance of the Output of a Machine with One Failure Mode

Derivation The formula derived in the previous section can be simplified to obtain the variance of the output of a machine with a single failure mode. For this purpose, we choose parameters for one of the failure modes so that it never takes place.

$$r_1 = 1 \quad p_1 = 0$$
$$r_2 = r \quad p_2 = p$$

The mean and the variance of the output of the production simplify to:

$$\bar{n}(t) = \frac{r}{r+p} t \tag{3.17}$$

$$\begin{aligned} \sigma^2(t) &= \frac{rp}{(r+p)^2}\left(\frac{2}{r+p} - 1\right) t \\ &\quad - \frac{2rp}{(r+p)^4}(1-r-p)(1-(1-r-p)^t) \end{aligned} \tag{3.18}$$

This results agrees with the result derived previously by Gershwin (1993). In this case, the asymptotic variance rate is

$$\Delta_{M_1} = \lim_{t \to \infty} \frac{\sigma^2(t)}{t} = \frac{rp}{(r+p)^2}\left(\frac{2}{r+p} - 1\right) \tag{3.19}$$

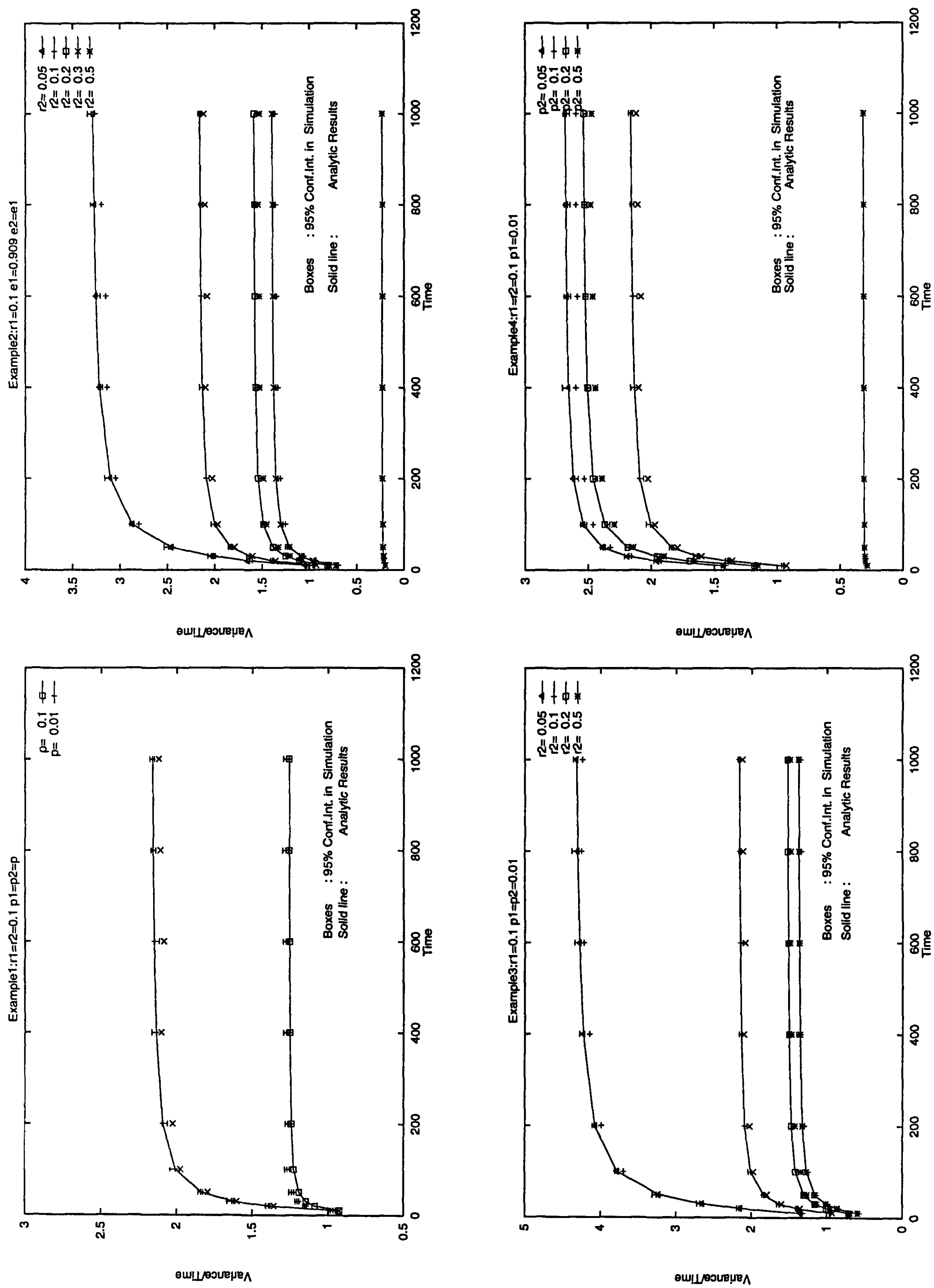

Figure 3-6: Variance of the output of a two-machine zero buffer system: analytical results vs. simulations

3.7 Asymptotic Variance Rate of the Output for a Two-Machine Zero Buffer System

In this section, we derive an expression that relates the asymptotic variance rate of a two-machine zero buffer system ($\Delta_{(M_1,M_2,0)}$) to the asymptotic variance rate of the machines (Δ_{M_1}, Δ_{M_2}) that compose the system.

From (3.19),

$$\begin{aligned}\Delta_{M_i} &= \frac{r_i p_i}{(r_i+p_i)^2}\left(\frac{2}{r_i+p_i}-1\right)\\ &= e_i(1-e_i)\frac{2-r_i-p_i}{r_i+p_i}\\ &= \frac{e_i^2(1-e_i)}{r_i}(2-r_i-p_i) \quad i=1,2\end{aligned}$$

and thus,

$$2-r_i-p_i = \frac{\Delta_{M_i} r_i}{e_i^2(1-e_i)} \quad i=1,2.$$

Substituting this expression in (3.16) and simplifying we get,

$$\Delta_{(M_1,M_2,0)} = \frac{C^4}{r_1 r_2}\left(\frac{\Delta_{M_1}}{e_1(1-e_1)^2}+\frac{\Delta_{M_2}}{e_2(1-e_2)^2}-\frac{2e_1e_2}{(1-e_1)(1-e_2)}\right). \qquad (3.20)$$

Chapter 4

Flow Interruption of a Two-Machine Line with Finite Buffer Size

In this chapter, we identify the circumstances that lead to flow interruption in a two-machine finite buffer system. We obtain parameters that characterize these interruptions in terms of their mean frequency and length. We also evaluate their probability distributions. These results are used to define a two-machine zero buffer system whose failure behavior is the same as that of the original two-machine finite buffer system. Finally, we show the similarities between these results and those obtained by Buzacott and Shanthikumar (1993).

4.1 Types of Interruptions

The system is working if an observer placed at the end of the line sees material coming out. In this case, the system is in a productive state. A period when output is not coming out is defined as an interruption. An interruption of the output takes place when the system enters a non-productive state. There are two different types

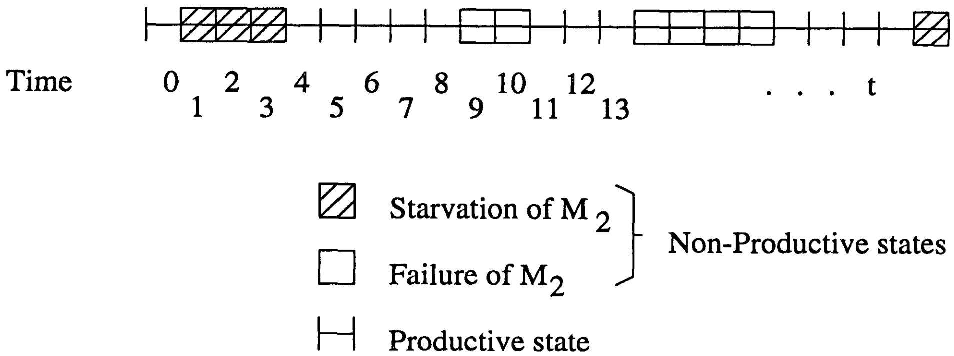

Figure 4-1: Interruptions of flow in the system

of non-productive states, and they lead to two different types of interruptions. These interruptions are failures of M_2 and starvations of M_2, and are explained in detail in the following paragraphs. In Fig. 4-1, we represent the output of the line for a range of time units:

- Productive periods: $t = 0$, $4 \leq t \leq 8$, $11 \leq t \leq 13$.
- Interruption of flow due to starvation of M_2: $1 \leq t \leq 4$.
- Interruption of flow due to failure of M_2: $9 \leq t \leq 10$.

Failure of M_2 Failures of M_2 always affect the output of the system. While M_2 is operational there might or might not be output, but if M_2 is down there is definitely no output. The non-productive states that cause this interruption of flow are $(n, \alpha_1, 0)\ \forall\ n \in [0, N],\ \alpha_1 = (0, 1)$.

Starvation of M_2 M_2 is starved when it is not allowed to work due to lack of material in the buffer. When M_1 is down and M_2 keeps working, the buffer level decreases at a rate of one piece per time unit. The failures of M_1 only interrupt the flow of output when the buffer empties. This means that some failures of M_1 will not cause an interruption of output, either because M_2 keeps working and M_1 gets repaired before the buffer empties, or because M_2 fails before this happens. Some of the interruptions seen by the outside observer will be shorter than the length of the failure of M_1, because the material in the buffer postpones the interruption. The non-productive state that causes this interruption of flow is $(0,0,1)$[1].

An important observation about these failures is that both *cannot happen simultaneously.* M_2 cannot be starved (unable to work due to lack of material) and undergoing repair at the same time. In general, there will be productive periods in between these interruptions. In very rare occasions[2] we may find a starvation of M_2 followed immediately by a failure of M_2.

4.2 Buffer as the Memory of the System

In this section, we explore how the buffer acts as a record of the events of the system and how the interruptions of flow due to starvations provide a way of simplifying the study of a two-machine line over time.

Suppose that we observe the buffer level for a while. If both machines are working or under repair, the buffer level does not change. If M_1 fails and M_2 keeps working, the buffer level decreases at a rate of one part per time unit. If M_2 fails and M_1

[1]As we discuss in Section 4.3.1 this is not exactly true. The first time unit that state $(0,0,1)$ is reached there is output. There is no output in the subsequent time units while the system remains in this state or the time unit after the system leaves the state (and reaches $(1,1,1)$). As a consequence, the time that the system spends in state $(0,0,1)$ is the same as the time that the output is interrupted due to starvation, and the periods of starvation are the periods that the system is in state $(0,0,1)$, with a unit delay. Therefore, it is accurate to identify starvation with state $(0,0,1)$.

[2]The sequence of states have to be:

TIME	STATE	OUTPUT	REASON FOR NO OUTPUT
1	$(1,1,1)$ or $(1,1,1)$	yes	
2	$(0,0,1)$	yes	
3	$(0,0,1)$	no	starvation of M_2
$\cdots$	$(0,0,1)$	no	starvation of M_2
$t_r - 1$	$(0,0,1)$	no	starvation of M_2
t_r	$(1,1,1)$	no	starvation of M_2
$t_r + 1$	$(1,0,0)$ or $(2,1,0)$	no	failure of M_2

M_2 is starved from $t = 3$ to $t = t_r$. At $t = t_r + 1$, M_2 is allowed to work, but it fails, so a failure of M_2 starts after a starvation before any production comes out of the line.

keeps working, the buffer level increases at the same rate. Therefore, the change in buffer level during a period of time reflects the difference in performance between the upstream and the downstream machine during that time.

There is a limit to this role of the buffer. Whenever there is a blockage or a starvation the difference in performance is not reflected accurately. A starvation empties the buffer, and as a consequence the memory is lost. Whatever happens in the system after a starvation does not depend on the history of the system before this starvation took place. This leads to an interesting observation: the behavior of the two-machine line over time can be studied as a series of periods which are independent of each other. These periods are the periods between starvations.

On the other hand, the likelihood of observing a starvation depends on the buffer level of the system at the time of a failure of M_1. As we have seen in Section 4.1, the higher the buffer level, the less likely it is that a starvation occurs. The probability of a starvation taking place depends only on the history of the system after the last starvation.

These observations can be summarized in technical terms: a two-machine line can be studied as a renewal process where starvations are the events that bring the system to its starting state (Gallager, 1995). Though this observation is not further used in this thesis, we believe it is worth mentioning it because it may provide additional tools to study the behavior of the system.

4.3 Interruption of output due to starvation: Equivalent Machine

As we have seen before, there are two kinds of interruptions of the output:

- Failures of M_2: The parameters that define these interruptions are the mean time to fail ($MTTF_2$) and to repair ($MTTR_2$) of M_2.

- Starvation of M_2: In this section, we calculate the parameters that define the interruptions of flow due to starvation of M_2.

We characterize the interruptions of output of the system due to starvations of M_2 by determining the mean time between starvations ($MTTF_*$) and the mean length of a starvation ($MTTR_*$). If the time between starvations and the length of a starvation are geometrically distributed random variables, we can define:

$$r_* = \frac{1}{MTTR_*} = p\,(\text{starvation finishes this period/there was no output last period})$$

$$p_* = \frac{1}{MTTF_*} = p\,(\text{starvation starts this period/there was output last period})$$

In this way, the interruptions of output due to starvations of M_2 are characterized in the same way as the failure and repair behavior of a machine. Therefore, M_1 and the buffer can be substituted by a single machine M_* with parameters p_* and r_*. The two systems are very close equivalents in terms of the failure and repair behavior. This is represented in Fig. 4-2.

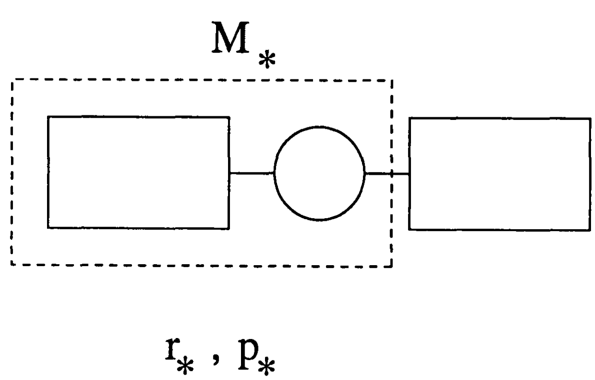

Figure 4-2: Equivalent machine to M_1 and the buffer

This two-machine zero buffer equivalent system will be used to determine the variance of the output of the original. The model developed for this purpose is

TIME	STATE	OUTPUT	REASON FOR NO OUTPUT
-1	$(1,0,1)$ or $(1,1,1)$	yes	
0	$(0,0,1)$	yes	
1	$(0,0,1)$	no	starvation of M_2
$\cdots$	$(0,0,1)$	no	starvation of M_2
$t_r - 1$	$(0,0,1)$	no	starvation of M_2
t_r	$(1,1,1)$	no	starvation of M_2
$t_r + 1$	$(1,0,0)$ or $(2,1,0)$	no	failure of M_2

Table 4.1: Sequence of states that define a starvation in a two-machine finite buffer line

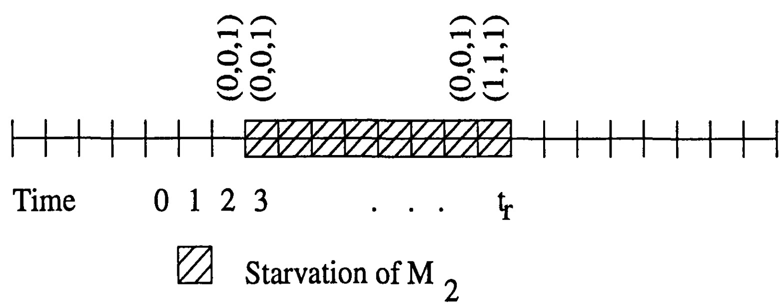

Figure 4-3: Start and end of an starvation

presented in Section 6.1.1.

4.3.1 Length of a Starvation

Start and End of a Starvation The sequence of states that define a starvation is shown in Fig. 4-3 and Table 4.1:

At time $t = 0$, there is output coming from the system. M_1 starts down and does not get repaired during this time unit. M_2 starts up, gets the last piece from the buffer and does not fail while processing it. This piece leaves the system at the end of $t = 0$. Starvation starts at $t = 1$, the time unit after the system first gets to state $(0,0,1)$.

Starvation ends at $t_r + 1$, the time unit after the system first gets to state $(1,1,1)$.

At time $t = t_r$, there is no output coming from the system. M_1 starts down, gets repaired during this time unit and processes a part. M_2 is up, and cannot work because the buffer is empty. The piece processed by M_1 reaches the buffer at the end of the time unit, ending the starvation of M_2.

We have found the states that define the start and the end of an starvation. If we determine the mean time to go from $(0,0,1)$ to $(1,1,1)$, we have the mean time to repair of the starvations. The actual starvation has a *delay* of one time unit, meaning that it starts and ends a unit of time after these states are reached, but this delay has no influence on its length.

Mean Length of a Starvation ($MTTR_*$) We have determined this parameter using two different methods, which give exactly the same results.

Method 1 The easiest way to determine $MTTR_*$ is to think about the characteristics of the repair distribution of M_1. The length of the interval from $(0,0,1)$ to $(1,1,1)$ is the time it takes M_1 to get repaired once in state $(0,0,1)$. As the repair distribution of M_1 is geometric (and thus, memoryless), the probability of getting repaired once in state $(0,0,1)$ is independent of how long M_1 was down before reaching $(0,0,1)$. So, we can conclude

$$MTTR_* = MTTR_1 \tag{4.1}$$

Method 2 There is another more elaborate way of determining $MTTR_*$. It consists of using the transition matrix in equation (2.1) of the original two-machine line to determine the average time it takes to go from state (001) to (111) (which we call $w_{(001)(111)}$). In technical terms, this average time is called the *expected first passage time* between those states. The procedure to obtain this value is described in detail in Appendix B. The result obtained is

$$MTTR_* = MTTR_1 \tag{4.2}$$

TIME	STATE	OUTPUT	REASON FOR NO OUTPUT
-1	$(0,0,1)$	no	starvation of M_2
0	$(1,1,1)$	no	starvation of M_2
1	$(1,0,1)$ or $(1,1,1)$	yes	
...	...	...	
$t_s - 1$	$(1,0,1)$ or $(1,1,1)$	yes	
t_s	$(0,0,1)$	yes	
$t_s + 1$	$(0,0,1)$	no	starvation of M_2

Table 4.2: Sequence of states that define the time between starvations in a two-machine finite buffer line

Repair Time Distribution From Method 1 we can conclude that the repair time distribution is geometric.

4.3.2 Time between Starvations

Start and End of a Period between Starvations The reference states we use to determine the distribution parameters are again $(1,1,1)$ and $(0,0,1)$. The sequence of states that define a period between starvations is shown in Table 4.2:

A period between starvations starts at $t = 1$, one time unit after state $(1,1,1)$ is reached from $(0,0,1)$. This period ends at $t = t_s$, one time unit after state $(0,0,1)$ is first reached. Fig 4-4 shows a period between starvations. During this time, there might be interruptions of flow due to failures of M_2. These interruptions are not taken into account in calculating the time between starvations. As we stated in Section 4.1, a starvation cannot take place while the output has been interrupted due to a failure of M_2. This is equivalent to saying that M_* cannot fail in this case.

Mean time between starvations ($MTTF_*$) There are three different methods of obtaining this parameter, which give us the same result:

Method 1 In Fig. 4-2 the original two-machine system is represented as an equivalent two-machine line with zero buffer size. The efficiency (E) of a two-machine

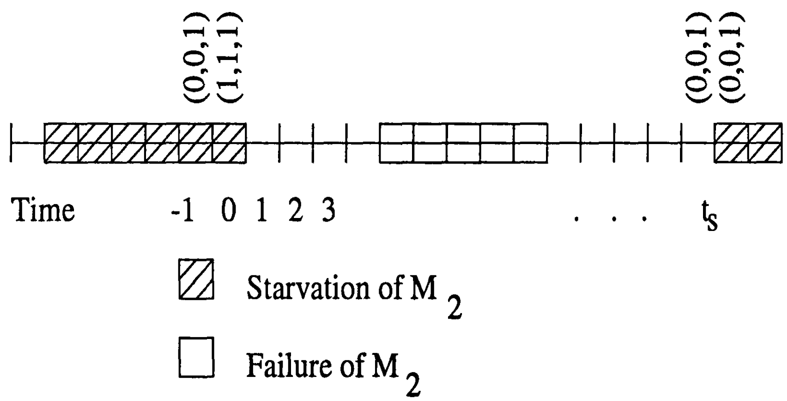

Figure 4-4: Start and end of a period between starvations

line with zero buffer size is

$$E = \frac{1}{1 + \frac{MTTR_2}{MTTF_2} + \frac{MTTR_1}{MTTF_*}}$$

We want the efficiency to be that of the original system, which can be obtained from equation (2.3). All the parameters of the equation are known except $MTTF_*$ and thus,

$$MTTF_* = \frac{1}{r_1}(\frac{Ee_2}{e_2 - E}) \tag{4.3}$$

Method 2 Again this method is more computationally intensive than the previous one. Using the procedure described in Appendix B, we can obtain $w_{(111)(001)}$, which is the mean time it takes the system to go from state $(1,1,1)$ to $(0,0,1)$, or in other words, the mean length of the period between starvations.

However, this value is not yet what we are looking for. The time to fail is defined as *working* time between failures. In the period between starvations, M_2 may fail. If M_2 is down, the system is not working (and does not have the possibility of starving). Therefore, the mean *working* time to fail is the fraction of $w_{(111)(001)}$ that M_2 is working.

$$MTTF_* = w_{(111)(001)} \frac{r_2}{r_2 + p_2} = w_{(111)(001)} \; e_2 \tag{4.4}$$

Method 3 This method is based on the procedure used by Buzacott and Shanthikumar (1993) to determine the mean time to fail in a two-machine line with an intermediate buffer. For this derivation we assume that the *working* mean time between starvations is geometric. Let p be the probability of seeing an interruption of the output in this period, provided that there was output last period. That is,

$$p = p(\text{no output this time unit/there was output last time unit})$$

There are two reasons why the output can be interrupted: failure (p_2) and starvation (p_*) of M_2. Since both cannot happen at the same time,

$$p = p_2 + p_* = \frac{\text{p(failure of } M_2\text{)+ p(starvation of } M_2\text{)}}{\text{p(there was output last period)}}$$

where

$$p_2 = \frac{\text{p(a failure of } M_2 \text{ starts)}}{\text{p(there was output last time unit)}}$$
$$p_* = \frac{\text{p(a starvation of } M_2 \text{ starts)}}{\text{p(there was output last time unit)}}.$$

In steady state,

$$\text{p(a starvation of } M_2 \text{ starts)} = \text{p(a starvation of } M_2 \text{ ends)}$$

so

$$p_* = \frac{\text{p(a starvation of } M_2 \text{ ends)}}{\text{p(there was output last time unit)}}$$

The probability of seeing output during a time unit is the efficiency of the system in steady state

$$\text{p(there was output last time unit)} = E.$$

The probability that a starvation ends during a time unit is the probability of M_2 being starved and M_1 being repaired during that time unit

$$\text{p(a starvation of } M_2 \text{ ends)} = r_1 \mathbf{p}(0,0,1).$$

Combining these expressions, the final result is

$$p_* = \frac{r_1 \mathbf{p}(0,0,1)}{E}$$

and therefore,

$$MTTF_* = \frac{1}{p_*} = \frac{E}{r_1 \mathbf{p}(0,0,1)}. \tag{4.5}$$

Equation (4.3) is equivalent to equation (4.5).

Time between Starvations Distribution Though we have been able to calculate the mean of the time between starvations, the distribution of this variable is not geometric. We have run several simulations[3] to obtain the shape of this distribution. In the example presented we have changed $MTTF_1$ while maintaining the rest of the parameters constant: $r_1 = r_2 = 0.1$, $p_2 = 0.01$, $N = 20$. The results are shown in Fig. 4-5. We have represented the density distribution of the working time between starvations in a semilogarithmic scale. The X-axis has been scaled using $MTTF_1$ to allow several distributions to be plotted in the same graph. If the distributions were geometric, their representation would be straight lines. Though we will assume that the distribution is in fact geometric, this conclusion is not confirmed by the data obtained.

Limit of p_* as the Buffer Size Increases As the buffer size increases, the probability of starvation of M_2 (and thus p_*) decreases. However, the limit p_* reaches as the buffer size grows depends on which machine is less reliable. Using equation (4.3),

[3]The simulation software is described in Section 7.1.

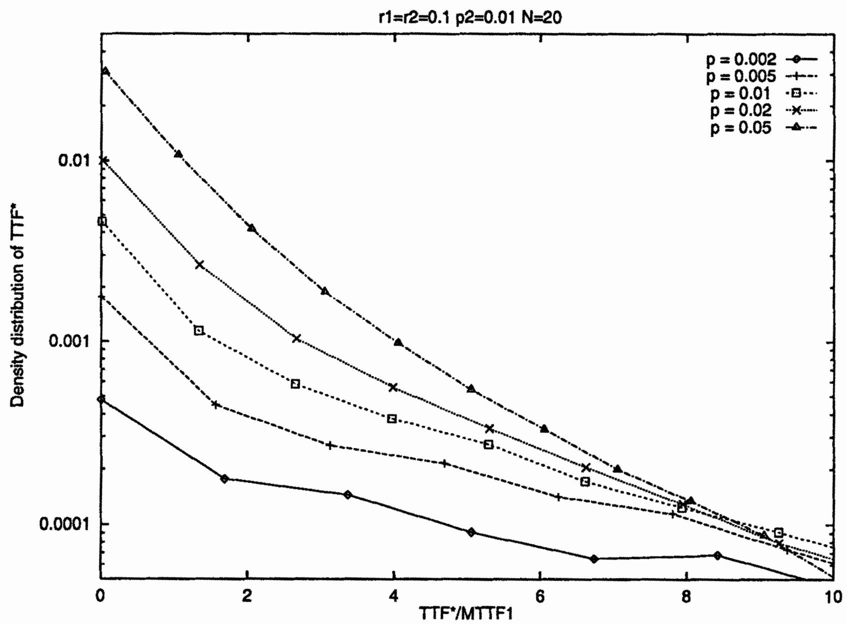

Figure 4-5: Distribution of the working time between starvations: simulation results

$$p_* = r_1 \frac{e_2 - E}{E e_2}$$

and from equation (2.5)

$$\lim_{N \to \infty} E = min(e_1, e_2)$$

so

$$\lim_{N \to \infty} p_* = \begin{cases} 0 & \text{if } e_2 \leq e_1 \\ \frac{r_1(e_2 - e_1)}{e_1 e_2} & \text{if } e_2 > e_1 \end{cases}$$

The fact that the limit of p_* as the buffer size increases has a different value depending on which is the most efficient machine can be explained intuitively. If M_1 is the most efficient machine, the buffer will tend to fill. As the buffer size increases, the amount of material in the buffer in steady state increases and, M_2 does not stop due to lack of material in the buffer. This is equivalent to saying that p_* tends to zero.

However, if the most efficient machine is M_2, the buffer will tend to be empty in steady state regardless of its size. This is because M_2, on average, can produce more material than M_1 can provide. Therefore, the failures of M_1 starve M_2, interrupting the output. The disruptive effect of the failures of M_1 (interruption of the output) decreases as the buffer size increases, but it reaches an asymptotic value greater than zero.

4.4 Another Derivation of the Equivalent Machine

Buzacott and Shanthikumar (1993) analyze a two-machine line system with operation dependent failures. They derive the parameters of a two-machine zero buffer line equivalent to the original system in terms of having the same steady state interdeparture distribution. The parameters of the two-machine zero buffer line obtained are similar to the parameters of the equivalent machine derived in Section 4.3. In this section, we compare the approaches.

4.4.1 Buzacott and Shanthikumar's Model

Assumptions Buzacott and Shanthikumar's model is very similar to Gershwin's model described in Chapter 2. The only differences are:

- Starvation/Blocking behavior: If both machines are up and the buffer level is zero, there is no starvation in Buzacott and Shanthikumar's model and there is in Gershwin's. If both machines are up and the buffer is full, there is no blocking in Buzacott and Shanthikumar's model and there is in Gershwin's.

- Observation instant: In Buzacott and Shanthikumar's model the buffer level in a time unit is defined by the events and the buffer level at the previous time unit. In Gershwin's model the buffer level is defined by the events of that same time unit and the buffer level at the previous time unit.

Buzacott and Shanthikumar's model	$n(t) = f(n(t-1), \alpha_1(t-1), \alpha_2(t-1))$
Gershwin's model	$n(t) = f(n(t-1), \alpha_1(t), \alpha_2(t))$

- Number of pieces in the system: In case of failure, in Buzacott and Shanthikumar's model the pieces being processed are stored in the machine while in Gershwin's model they are considered to be in the upstream buffer. This means that for a buffer of size N, the maximum number of pieces in the system in Buzacott and Shanthikumar's model is $N+2$ while in Gershwin's it is N.

Equivalent Machine for the First Machine and the Buffer The *interdeparture time* is the number of time units between consecutive departures of parts from the line. Buzacott and Shanthikumar conclude that a two-machine line with a finite buffer size has the same interdeparture distribution as a two-machine with zero buffer line, where the first machine has a failure probability p_1', and a repair probability $r_1' \equiv r_1$.

The failure probability p_1' is derived using the same procedure as in Section 4.3.2, Method 3. The result is

$$p_2 + p_1' = p(\text{no output this time unit/there was output last time unit})$$

$$p_1' = p(\text{a starvation of } M_2 \text{ starts/there was output last time unit})$$

$$p_1' = \frac{p_1 p(1,1,0) + (1-r_1)p(0,1,1)}{E}$$

4.4.2 Comparison

Next, we compare the parameters of the equivalent system derived in Section 4.3 with the parameters used by Buzacott and Shanthikumar.

Both procedures represent the two-machine system with a two-machine zero buffer line. They eliminate the buffer and quantify its effect by reducing the number of failures of the system due to failures of M_1. Both p_* and p_1' represent the probability of interrupting the output due to starvation of M_2 in their correspondent systems. The fact that they can both be obtained using the same procedure demonstrates it.

We have seen that the most important assumptions made both in Buzacott and Shanthikumar's and in Gershwin's models are the same. Therefore, the values of p_* and p_1' for a system should be close. Fig. 4-6 compares these values for different systems. As we saw in the description of the models, for a buffer size N, the maximum number of pieces in the system for Gershwin's model is N and for Buzacott and Shanthikumar's model is $N+2$, so the total number of pieces in the system is different. To compensate for this factor, p_1' has been calculated for buffer size $N-2$.

- Cases (a) to (d) have the same M_1, and M_2 has in all cases the same efficiency. The less frequent the events of M_2 are, the closer are p_* and p_1'.
- Cases (I) to (III) show that the difference between p_* and p_1' depends only on the characteristics of M_1 and M_2 and not on the buffer size.
- Cases 1 to 10 are randomly generated systems that are presented to give the reader an idea of the range of the difference between p_* and p_1'.

The difference between p_* and p_1' gets bigger as the frequency of the failures and repairs of the machines increase. This is because the different conventions of the models affect more the behavior as the frequency of the events increase.

We can conclude that by both procedures we obtain a two-machine zero buffer system equivalent to the original two-machine finite buffer system in terms of the interruption of the output. In Section 6.2 we confirm that the interdeparture distribution of the original and the equivalent system are the same. However, we also discover that the variance of the output of the original and the equivalent system are significantly different, and we provide an explanation for this.

Number	r_1	p_1	r_2	p_2	N	p'_1	p_*	$\frac{p'_1 - p_*}{p'_1}$
a	0.95	0.39	0.85	0.15	6	0.260852	0.223400	0.143577
b	0.95	0.39	0.54	0.096	6	0.255330	0.231500	0.093328
c	0.95	0.39	0.27	0.048	6	0.274870	0.261862	0.047323
d	0.95	0.39	0.034	0.006	6	0.361787	0.359620	0.005989
I	0.1	0.01	0.1	0.01	20	0.004897	0.004871	0.005263
II	0.1	0.01	0.1	0.01	50	0.002646	0.002632	0.005263
III	0.1	0.01	0.1	0.01	100	0.001498	0.001491	0.005263
1	0.035	0.002	0.021	0.0021	72	0.017735	0.017711	0.001363
2	0.076	0.0066	0.079	0.0054	29	0.003470	0.003461	0.002767
3	0.057	0.0065	0.055	0.012	89	0.000522	0.000518	0.006363
4	0.1	0.069	0.033	0.0093	15	0.057124	0.056710	0.007245
5	0.068	0.017	0.25	0.036	71	0.007417	0.007356	0.008230
6	0.21	0.038	0.015	0.013	18	0.013620	0.013454	0.012150
7	0.033	0.011	0.22	0.1	233	0.000149	0.000147	0.013950
8	0.093	0.012	0.18	0.11	45	0.000114	0.000110	0.039521
9	0.5	0.18	0.11	0.056	11	0.055353	0.052571	0.050264
10	0.34	0.23	0.81	0.58	8	0.050287	0.039076	0.222941

Figure 4-6: Comparison between the mean time between starvations for Buzacott and Shanthikumar's and Gershwin's two-machine finite buffer lines

4.5 Conclusions

The goal of this chapter was to identify the circumstances that lead to flow interruption in a two-machine finite buffer system and to obtain the parameters that characterize their frequency and length. Two types of interruptions were identified: failures of M_2 and starvations of M_2. Both interruptions cannot happen at the same time.

The frequency of the failures of M_2 are defined by p_2 and their length by r_2. Both distributions are geometric. The frequency of the starvations of M_2 are defined by $p_* = r_1(e_2 - E)/(Ee_2)$ and their length by $r_* \equiv r_1$. The distribution of the length of the starvations is geometric. We assume that the distribution of the frequency of the starvations is geometric, because ,although it is not, it is fairly close to it.

This failure characterization allows us to represent M_1 and the buffer by an equivalent machine M_* whose failures are defined by p_* and r_*. This derivation is analogous to the one performed by Buzacott and Shanthikumar (1993).

Chapter 5

Asymptotic Variance Rate of the Output

There is only one result in the literature that we are aware of that determines analytically the variance of the output of a deterministic two-machine line with a finite buffer size. Miltenburg (1987) derives the asymptotic variance of the output of a production line. In this chapter, we use this result to explore the change in the variance of the output of a two-machine line as function of the buffer size.

5.1 Miltenburg's Asymptotic Variance Rate

5.1.1 Derivation

This derivation is explained in more detail in Miltenburg (1987). The *asymptotic variance rate of a function* f called $\Delta(f)$ is the limit of the variance of this function over a period $[0, t]$ divided by t when t tends to infinity:

$$\text{Asymptotic variance rate} = \Delta(t) = \lim_{t \to \infty} \frac{\sigma^2(f(t))}{t}. \tag{5.1}$$

We restrict our attention to functions defined over the states of a Markov chain. Let f be defined as

$$f(t) = \sum_{k=1}^{t} g(k)$$

where

$$\begin{aligned} g(k) &= \text{value of } g \text{ at the } k\text{th time unit} \\ g(k) &= g_i \text{if the system is in state } i \text{ at time } k \end{aligned}$$

Kemeny and Snell (1976) (corollary 4.6.2) provide a way to calculate $\Delta(f)$ as a function of g:

$$\Delta(f) = \lim_{t\to\infty} \frac{\sigma^2(f(t))}{t} = \lim_{t\to\infty} \frac{1}{t}\sigma^2 \left[\sum_{k=1}^{t} g(k)\right] = \sum_{i,j=1}^{n} g_i c_{ij} g_j \qquad (5.2)$$

where

c_{ij} is the *limiting covariance* of state i and state j

c_{ij} can be calculated using Kemeny and Snell (1976) (theorem 4.6.1):

$$c_{ij} = \begin{cases} \mathbf{p}_i z_{ij} + \mathbf{p}_j z_{ji} - \mathbf{p}_i \mathbf{p}_j & i \neq j \\ \mathbf{p}_i (2z_{ii} - 1 - \mathbf{p}_i) & i = j \end{cases} \qquad (5.3)$$

where

$$\begin{aligned} \mathbf{p}_i &= \text{steady state probability of state } i \\ z_{ij} &= \text{element } (i,j) \text{ of the fundamental matrix } Z, \\ &\quad \text{where} \\ Z &= (I - P + \Phi)^{-1} \\ &\quad I = \text{identity matrix.} \\ &\quad P = \text{transition matrix as defined in (2.1).} \end{aligned}$$

Φ = limiting multi-step transition probability matrix,

where each row is the steady state probability vector $\mathbf{p}$.

Miltenburg derived the asymptotic variance rate of the output of a production line. He defined $g(k)$ as the following function of the state:

$$g_i = \begin{cases} 1 & \text{if } i \text{ is a productive state } (i \in NP) \\ 0 & \text{otherwise } (i \in NP) \end{cases}$$

Then $f(t) = \sum_{k=1}^{t} g(k)$ is the production of the line in an interval $[0, t]$ and $\Delta(f) \equiv \Delta$ is the asymptotic variance rate of the output of the line. By substituting the value of g_i for this particular case in equation (5.2) the asymptotic variance rate of the output (Δ) is calculated as follows:

$$\Delta = \lim_{t \to \infty} \frac{1}{t} \sigma^2 \left[\sum_{k=1}^{t} g(k) \right] = \sum_{i,j \in P} c_{ij} \tag{5.4}$$

5.1.2 Example: Variance of a Single Machine with Two Failure Modes

The purpose of the example is to illustrate Miltenburg's procedure. The asymptotic variance rate obtained by this procedure agrees with the value obtained in Section 3.4, and this derivation validates that result.

The system we are analyzing is a single machine with two failure modes. We assume the two failures cannot happen at the same time and that there is always a productive period between failures. Fig. 5-1 represents the Markov chain transition graph of the system. It has only one productive state (1) and two non-productive states (0_1) and (0_2). Equation (5.2) simplifies to

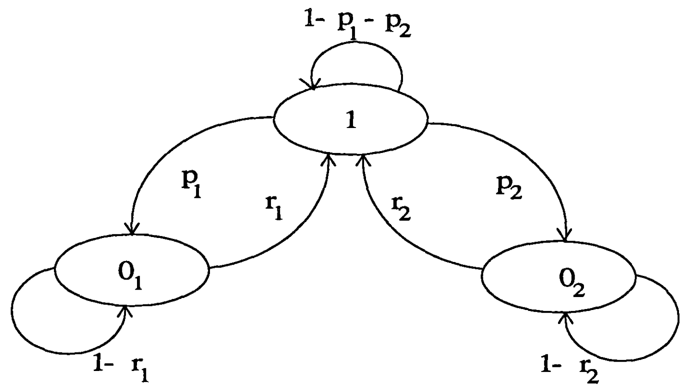

Figure 5-1: Markov chain of a single machine with two failure modes

$$\Delta \;=\; \sum_{i,j=1}^{n} g_i c_{ij} g_j = c_{11} = \mathbf{p}_1(2z_{11} - \mathbf{p}_1 - 1)$$

From the derivation in Section 3.5, using equations (3.1), (3.8) and (3.9),

$$P = \begin{bmatrix} 1-p_1-p_2 & p_1 & 0 \\ r_1 & 1-r_1 & p_2 \\ r_2 & 0 & 1-r_2 \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \\ \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \\ \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{p}(1) & \mathbf{p}(0_1) & \mathbf{p}(0_2) \\ \mathbf{p}(1) & \mathbf{p}(0_1) & \mathbf{p}(0_2) \\ \mathbf{p}(1) & \mathbf{p}(0_1) & \mathbf{p}(0_2) \end{bmatrix} = \frac{1}{1+\frac{r_1}{p_1}+\frac{r_2}{p_2}} \begin{bmatrix} 1 & \frac{p_1}{r_1} & \frac{p_2}{r_2} \\ 1 & \frac{p_1}{r_1} & \frac{p_2}{r_2} \\ 1 & \frac{p_1}{r_1} & \frac{p_2}{r_2} \end{bmatrix}.$$

To calculate c_{11}, we just need z_{11}. Using the definition in equation (5.3), z_{11} turns out to be

$$z_{11} = \frac{p_2 r_1^2(1+r_2) + p_1 r_2^2(1+r_1) + r_1^2 r_2^2}{(p_2 r_1 + p_1 r_2 + r_1 r_2)^2}.$$

Therefore,

$$\Delta = c_{11} = C\frac{r_1^2 p_2(2-p_2-r_2) + r_2^2 p_1(2-p_1-r_1) - 2p_1p_2r_1r_2}{(p_2r_1 + r_2p_1 + r_1r_2)^3}$$

which is the same as equation (3.19).

5.1.3 Limitations of Miltenburg's Asymptotic Variance Rate

There are very simple production lines for which the number of states is very small and Δ can be calculated analytically following the derivation in the previous subsection.

Except for these very simple cases, Miltenburg's procedure to obtain Δ is computationally intensive. The number of states in a system composed of k machines and $k-1$ buffers of size N_i is

$$2^k \prod_{i=1}^{k-1} (N_i + 1).$$

The number of states defines the size of the matrix that must be calculated (and inverted) to calculate Δ. This number grows very quickly as the number of machines or buffer sizes increase and limits the cases in which the method can be used.

5.2 Asymptotic Variance Rate of a Production Line and its Reversed Line

Here, we introduce some notation (some of which has already been used) that we use in the rest of the chapter :

$\sigma^2_{M_i}(t)$ is the variance of the output of M_i if it is working in isolation.

Δ_{M_i} is the asymptotic variance rate of the output of M_i, if it were operating in isolation.

$$\Delta_{M_i} = \lim_{t \to \infty} \frac{1}{t} \sigma^2_{M_i}(t)$$

$\sigma^2_{(M_1,M_2,N)}(t)$ is the variance of the output of the system formed by M_1 and M_2 with a buffer of size N.

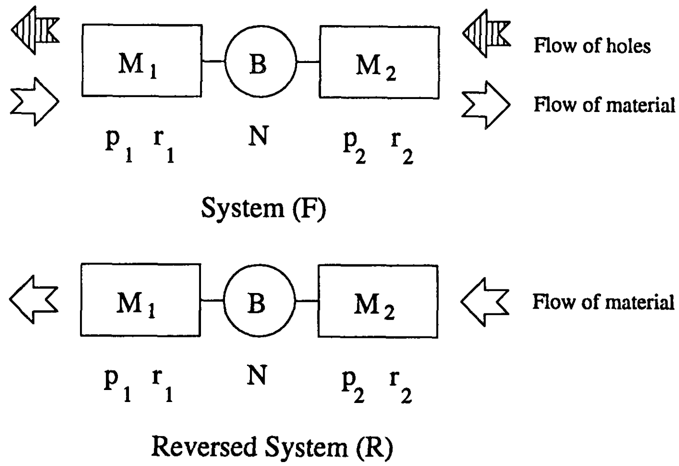

Figure 5-2: Representation of a system and its reverse

$\Delta_{(M_1,M_2,N)}$ is the asymptotic variance rate of the output of the two-machine finite buffer line formed by M_1 and M_2 and a buffer of size N.

$$\Delta_{(M_1,M_2,N)} = \lim_{t\to\infty} \frac{1}{t}\sigma^2_{(M_1,M_2,N)}(t)$$

Two special cases of this notation are:

$\Delta_{(M_1,M_2,0)}$ is the asymptotic variance rate of the output of the two-machine zero buffer line formed by M_1 and M_2.

$\Delta_{(M_1,M_2,\infty)}$ is the value to which $\Delta_{(M_1,M_2,N)}$ converges as N increases:

$$\Delta_{(M_1,M_2,\infty)} = \lim_{N\to\infty} \left(\lim_{t\to\infty} \frac{1}{t}\sigma^2_{(M_1,M_2,N)}(t) \right)$$

N_ϵ is the minimum buffer size that, given ϵ satisfies,

$$\forall\ N' > N_\epsilon, |\Delta_{(M_1,M_2,N')} - \Delta_{(M_1,M_2,\infty)}| < \epsilon.$$

The reverse of a production line is the production line formed by the same machines and buffers but placed in the reverse order (Fig. 5-2).

5.2.1 Finite Buffers

The steady state probability distribution of a system ($F \equiv (M_1, M_2, N)$) and its reverse ($R \equiv (M_2, M_1, N)$) are related as follows:

$$p_F(n, \alpha_1, \alpha_2) = p_R(N - n, \alpha_2, \alpha_1)$$

This result can be derived by exchanging the r_1 and r_2 and p_1 and p_2 in the solution of the steady state probability equations presented in equation (2.1), and which is derived in Gershwin (1994).

A way of interpreting the relationship between a system and its reverse is to consider that, in a system, two processes are taking place simultaneously (Ammar, 1980). One process is the movement of pieces in the direction of the flow and the second process is the movement of holes in the opposite direction. A hole is an empty space in the buffer. At all times, the number of parts plus the number of holes in the buffer is equal to the buffer size N. If a part leaves the system, a hole enters it and if a part enters the system, a hole leaves it. A hole entering the system means that there is output during that time unit.

Let us define for the purpose of this section the following notation:

$\Delta_P^O(M_1, M_2, N)$ or $\Delta_{(M_1,M_2,N)}$ Variance of the output of pieces of a system.

$\Delta_P^I(M_1, M_2, N)$ Variance of the input of pieces of a system.

$\Delta_H^O(M_1, M_2, N)$ Variance of the output of holes of a system.

$\Delta_H^I(M_1, M_2, N)$ Variance of the input of holes of a system.

Asymptotic Variance Rate of the Output of a System (F) and its Reverse(R)

Conjecture: for finite buffers, the asymptotic variance rate of the output of the line ($\Delta_P^O(M_1, M_2, N)$) is the same as the asymptotic variance rate of the output of the reversed line ($\Delta_P^O(M_2, M_1, N)$), or

$$\Delta_P^O(M_1, M_2, N) = \Delta_P^O(M_2, M_1, N) \quad (5.5)$$

This has been observed by computing and comparing both values for a wide variety of examples, and agrees with the values obtained by simulations[1]. Examples are presented in Appendix C.

This conjecture is very important for the study of the variance of the system over time. In Chapter 7, we observe that in general, the variance of a system and its reverse over a short period of time are not the same. However, this conjecture suggests that the behavior of both systems has to converge for some sufficiently large t. We use this asymptotic behavior as a tool to understand the changes of the system over time.

Asymptotic Variance Rate of the Output (O) and the Input (I) of a System

The equivalence between the flow of pieces (P) and the flow of holes (H) in the opposite direction implies:

$$\Delta_P^O(M_2, M_1, N) = \Delta_H^I(M_2, M_1, N) \quad (5.6)$$

$$\Delta_P^I(M_2, M_1, N) = \Delta_H^O(M_2, M_1, N) \quad (5.7)$$

From Fig. 5-2 we can see that the flow of holes in a system is equivalent to the flow of material in its reverse system. Therefore, the asymptotic variance rate of the output of holes in a system is the same as the asymptotic variance rate of the output of pieces in a the reverse one. The same is true for the input of holes and pieces. This can be expressed as:

$$\Delta_P^O(M_1, M_2, N) = \Delta_H^O(M_2, M_1, N) \quad (5.8)$$

[1]Due to the fact that the procedure of calculating the asymptotic variance rate is computationally intensive, the maximum buffer size that we have tried is $N = 400$.

$$\Delta_P^I(M_1, M_2, N) = \Delta_H^I(M_2, M_1, N) \tag{5.9}$$

From equations (5.5), (5.6), (5.7) and (5.8),

$$\Delta_P^I(M_1, M_2, N) = \Delta_P^O(M_1, M_2, N) \tag{5.10}$$

We can conclude that, for a system with finite buffers, the asymptotic variance rate of the output is equal to the asymptotic variance rate of the input.

5.2.2 Infinite Buffers

Though experimentally we have come to the conclusion that

$$\Delta_{(M_1,M_2,N)} = \Delta_{(M_2,M_1,N)}$$

when the buffer size is finite, we cannot extend our conclusions to the infinite buffer case.

In Section 7.3 we see that the variance of a system and its reverse are not the same for small periods of time $[0, t]$, but that as time increases the values of the variance converge. Let t_ϵ^N be the smallest time such that for a given N and ϵ

$$\forall\, t > t_\epsilon^N,\ |\frac{\sigma^2_{(M_1,M_2,N)}}{(}t)t - \frac{\sigma^2_{(M_2,M_1,N)}(t_\epsilon^N)}{t_\epsilon^N}| < \epsilon.$$

The greater the buffer size the longer time it takes for both values to converge. That is, as N increases, t_ϵ^N increases. Miltenburg's derivation does not provide information to define the time t_ϵ^N.

We do not know if the limit of t_ϵ^N as $N \to \infty$ is finite, so we cannot conclude whether the asymptotic variance rate converges in this case.

Another reason that prevents us from deriving conclusions is that if the buffer is infinite either a system or its reversed system never reaches steady state. If the more efficient machine is the first one, the material accumulates in the buffer without limit. Therefore, it is not meaningful to talk about the steady state behavior of the system.

However, though our hypothesis are restricted to the finite buffer cases, the observations of Δ (Section 5.3), make us believe that $\Delta_{(M_2,M_1,N)}$ and $\Delta_{(M_1,M_2,N)}$ approach a limit as N increases. We define this limit as $\Delta_{(M_1,M_2,\infty)}$. The previous argument allows us to state that

$$\Delta_{(M_1,M_2,\infty)} = \Delta_{(M_2,M_1,\infty)}.$$

In Section 7.3 we discuss this a little bit further.

5.3 Asymptotic Variance Rate of the Output as a Function of the Buffer Size

In this section, we analyze the influence of the buffer size in the asymptotic variance rate of the output of a system.

The questions that we want to answer are:

- Are there upper and lower limits for $\Delta_{(M_2,M_1,N)}$?

- How does $\Delta_{(M_2,M_1,N)}$ change as N increases?

We have performed this study by determining $\Delta_{(M_2,M_1,N)}$ as a function of N for a variety of cases. We have used Miltenburg's derivation of the asymptotic variance rate of the output to obtain these results. We discuss them in two different sections: machines with different efficiencies and machines with the same efficiencies. As we mentioned before, we deal with finite buffer sizes so the results are valid for both a system and its reverse.

5.3.1 Machines with Different Efficiencies

Fig. 5-3 and Fig. 5-4 show $\Delta_{(M_2,M_1,N)}$ as a function of N. In Fig. 5-3, M_2 characteristics are $r_2 = 0.1,\ p_2 = 0.1$. In Fig. 5-4, M_2's characteristics are $r_2 = 0.4,\ p_2 = 0.1$. In both figures, M_2 is the less efficient machine. In case (a), $MTTR_1 = MTTR_2$. In case (b), $MTTF_1 = MTTF_2$. In case (c), we compare different M_1 with the same efficiency. The following observations can be made from the graphs:

$\Delta_{(M_1,M_2,0)}$: From the data obtained in these experiments we see that the asymptotic variance rate of the zero buffer system $\Delta_{(M_1,M_2,0)}$ can be greater, smaller or between Δ_{M_1} and Δ_{M_2}.

In Fig. 5-3 case (c), $\Delta_{M_2} = 2.25$. In all the cases represented, $\Delta_{M_2} > \Delta_{(M_1,M_2)}$. We find examples of:

- $\Delta_{M_1} > \Delta_{M_2} > \Delta_{(M_1,M_2,0)}$

  M_1 characteristics are $r_1 = 0.1,\ p_1 = 0.0541,\ \Delta_{M_1} = 2.73$. $\Delta_{(M_1,M_2)} = 1.65$

- $\Delta_{M_2} > \Delta_{M_1} > \Delta_{(M_1,M_2,0)}$

  M_1 characteristics are $r_1 = 0.15,\ p_1 = 0.0811,\ \Delta_{M_1} = 1.74$. $\Delta_{(M_1,M_2)} = 1.43$

- $\Delta_{M_2} > \Delta_{(M_1,M_2,0)} > \Delta_{M_1}$

  M_1 characteristics are $r_1 = 0.225,\ p_1 = 0.122,\ \Delta_{M_1} = 1.09$. $\Delta_{(M_1,M_2)} = 1.28$

In Fig. 5-4 case (c),$\Delta_{M_2} = 0.48$. In all the cases represented, $\Delta_{(M_1,M_2)} > \Delta_{M_2}$. We find examples of:

- $\Delta_{M_1} > \Delta_{(M_1,M_2,0)} > \Delta_{M_2}$

  M_1 characteristics are $r_1 = 0.1,\ p_1 = 0.01,\ \Delta_{M_1} = 1.42$. $\Delta_{(M_1,M_2)} = 1.13$

- $\Delta_{(M_1,M_2,0)} > \Delta_{M_1} > \Delta_{M_2}$

  M_1 characteristics are $r_1 = 0.2,\ p_1 = 0.02,\ \Delta_{M_1} = 0.67$. $\Delta_{(M_1,M_2)} = 0.73$

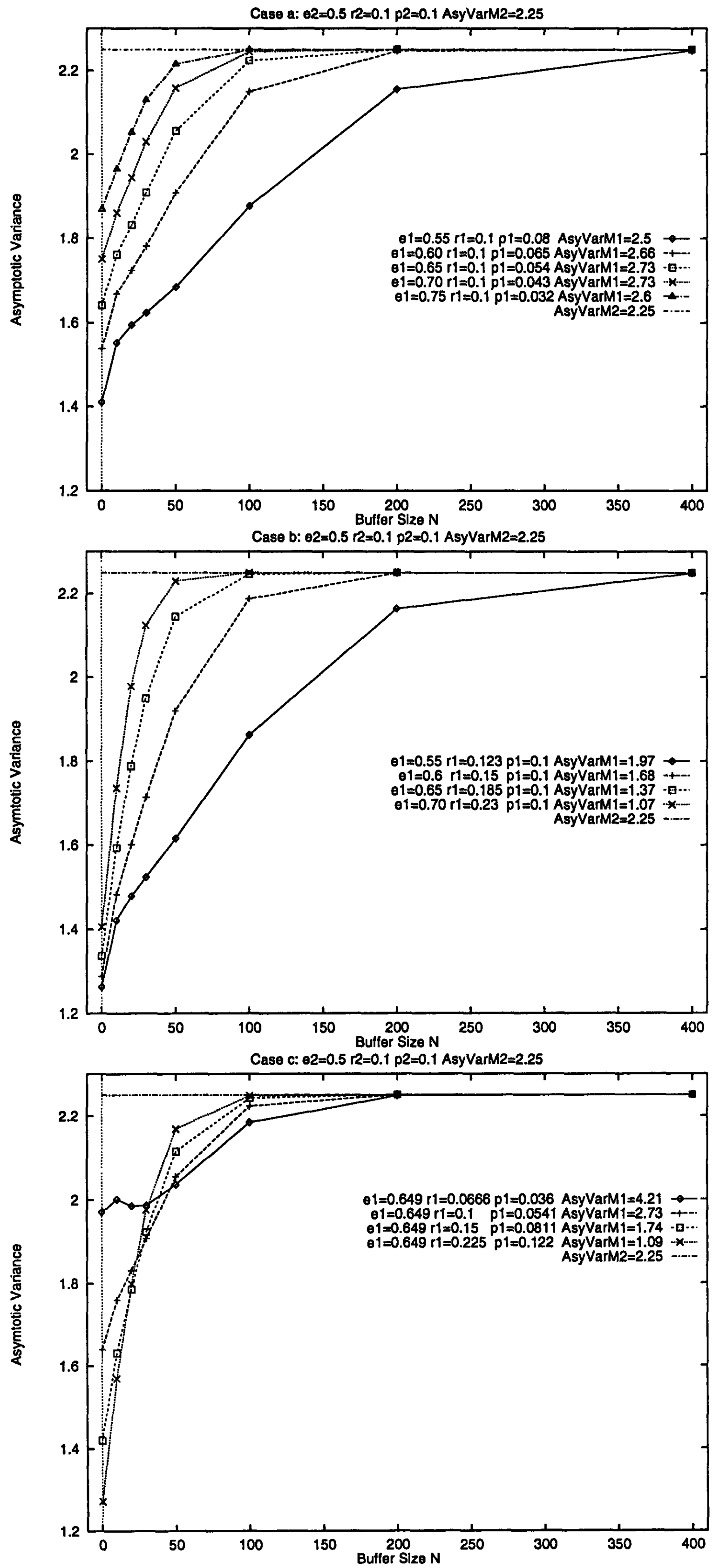

Figure 5-3: Example 1: Asymptotic variance rate of the output of a two-machine line as a function of the buffer size

- $\Delta_{(M_1,M_2,0)} > \Delta_{M_2} > \Delta_{M_1}$

  M_1 characteristics are $r_1 = 0.4,\ p_1 = 0.04,\ \Delta_{M_1} = 0.29.\ \Delta_{(M_1,M_2,0)} = 0.53$

In Section 3.4 we derive the asymptotic variance rate of a single machine with two failure modes. This value very accurately represents the asymptotic variance rate of a two-machine line with no buffers. Equation (3.20) shows the relationship between the asymptotic variance rate of the independent machines and of the zero buffer system.

$\Delta_{(M_1,M_2,\infty)}$: The asymptotic variance rate of the two-machine system approaches the asymptotic variance rate of the less efficient machine as the buffer size increases.

$$\text{If } e_i > e_j, \text{ then } \Delta_{(M_1,M_2,\infty)} = \Delta_{M_j}.$$

This can be seen in all cases of Fig. 5-3 and Fig. 5-4.

Consider a two-machine line with a very big buffer with M_2 less efficient than M_1. As M_1 produces more than M_2, the material accumulates in the buffer and in steady state the buffer level is very high. Therefore, the output of the system depends mostly on the behavior of M_2 because M_2 is almost never starved. Consequently, the asymptotic variance rate of the output of a system as the buffer size increases gets closer to the variance of the production of M_2, that is, the asymptotic variance rate of the production of the less efficient machine. If the system is reversed, the same is true since the asymptotic variance rate of a system and its reversed one coincide.

N_ϵ: We have defined N_ϵ as the minimum buffer size that satisfies

$$\forall\ N' > N_\epsilon, |\Delta_{(M_1,M_2,N')} - \Delta_{(M_1,M_2,\infty)}| < \epsilon.$$

There are two main factors that affect N_ϵ for a given M_2: the difference in efficiencies between the two machines and the frequency of the events of the more reliable machine.

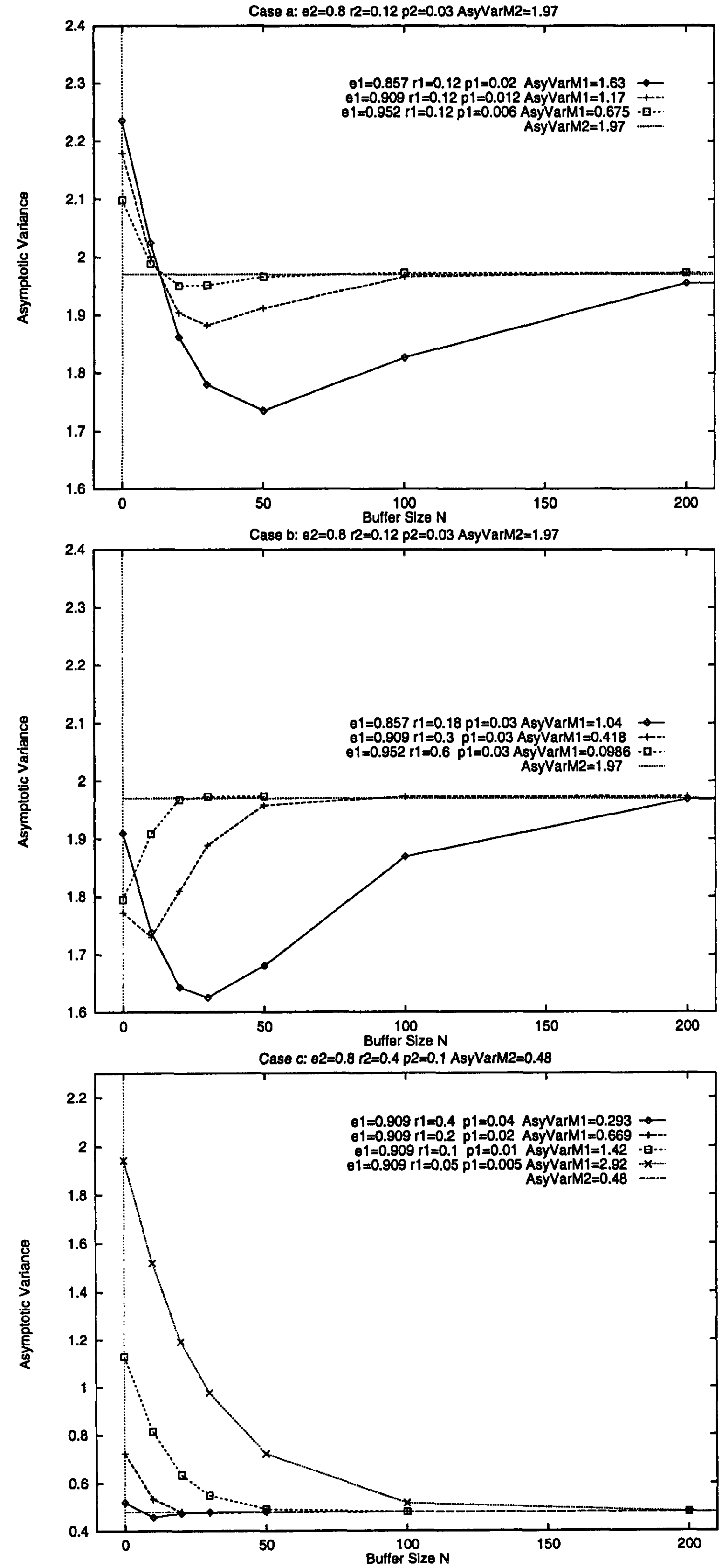

Figure 5-4: Example 2: Asymptotic variance rate of the output of a two-machine line as a function of the buffer size

- *Difference in the efficiency of two machines:* The greater difference there is between the efficiencies of both machines, the smaller N_ϵ. As r_1 increases or p_1 decreases, the efficiency of M_1 increases and N_ϵ decreases.

  This can be seen in cases (a) and (b) of Fig. 5-3 and Fig. 5-4. For example, for an $\epsilon = 0.01$ in case (a) of Fig. 5-3, if $e_1 = 0.75$, $N_\epsilon \approx 100$, if $e_1 = 0.65$, $N_\epsilon \approx 200$ and if $e_1 = 0.55$, $N_\epsilon \approx 400$.

  The effect of the buffer, as we discussed in Section 3.3, is to decouple one machine from the other, to make their performance independent. For the purposes of the discussion, we use the case where the less efficient machine is M_2. The way to decouple M_2 from M_1 is to assure that there is enough material in the buffer so that in case of a failure of M_1, M_2 can keep working during the length of the downtime. $\overline{n}$ depends on the difference in the efficiency of the machines. For a given N, the greater the difference in the efficiencies, the higher $\overline{n}$. Therefore, to isolate M_2, the greater the difference in efficiency between both machines, the smaller N_ϵ needs to be for a given ϵ.

- *Frequency of events:* The frequency at which the events take place in M_1 also affects N_ϵ for a given M_2. The lower the frequency of these events (the lower r_1 and p_1), the higher N_ϵ for a given ϵ.

  This is shown in cases (c) of Fig. 5-3 and Fig. 5-4. For example, in case (c) of Fig. 5-4 in all cases $e_1 = 0.909$. When $r_1 = 0.05$, $N_\epsilon \approx 150$, when $r_1 = 0.1$, $N_\epsilon \approx 50$ and when $r_1 = 0.2$, $N_\epsilon \approx 20$.

  Though the difference in efficiencies between the machines is the most important factor to define N_ϵ, the lower the frequency of these events, the higher is the buffer size needed to isolate M_2 from the failures of M_1.

Upper and lower limits for the $\Delta_{(M_1,M_2,N)}$ In Fig. 5-3, the variance of the output for intermediate buffer sizes is between the variance of the same line with size zero buffer and the variance of the less efficient machine. That is $\Delta_{M_2} > \Delta_{(M_1,M_2,N)} > \Delta_{(M_1,M_2,\infty)}$. However, in Fig. 5-4 there were cases where for some buffer sizes the variance becomes smaller than both $\Delta_{(M_1,M_2,\infty)}$ and Δ_{M_i}, where M_i is the less efficient machine. In cases (a) and (b), we see that this effect appears and becomes more significant as the efficiencies of both machines get closer. We comment this fact further in the discussion of machines with the same efficiency in Section 5.3.2.

5.3.2 Machines with Same Efficiency

When the efficiencies of the machines are closer, N_ϵ becomes larger and $\Delta_{(M_1,M_2,N)}$ may have a minimum for some finite, non-zero N. This behavior has already been described in the previous examples.

Here, we explore further cases where the efficiencies of both machines are the same or very close. Fig. 5-5 and Fig. 5-6 show some examples the systems studied. In Fig. 5-5, we have machines with the same efficiency but with events happening at different frequencies. In Fig. 5-6, we have identical machines and slight differences to study the changes of the system. In Fig. 5-7, the efficiencies of both machines are very similar, but the frequency of the failures and repairs is very different. There are some of observations about the behavior of these systems that are worth mentioning:

$\Delta_{(M_1,M_2,\infty)}$

- *Same efficiencies:* If the machines are identical, $\Delta_{(M,M,\infty)}$ is significantly lower than Δ_M. If both machines have the same efficiency, $\Delta_{(M_1,M_2,\infty)}$ is always lower than $\max(\Delta_{M_1}, \Delta_{M_2})$.

  Fig. 5-5 shows two examples where machines are identical. In case (a) for example, when the parameters of both machines are $r = 0.1$, $p = 0.01$, $\Delta_{(M,M,400)} = 1.01$ which is far from $\Delta_M = 1.42$. In case (b) when the parameters of both machines are $r = p = 0.1$, $\Delta_{(M,M,400)} = 1.50$ which is far from $\Delta_M = 2.25$.

  Fig. also 5-5 shows examples where machines are not identical. In case (b) we have examples where

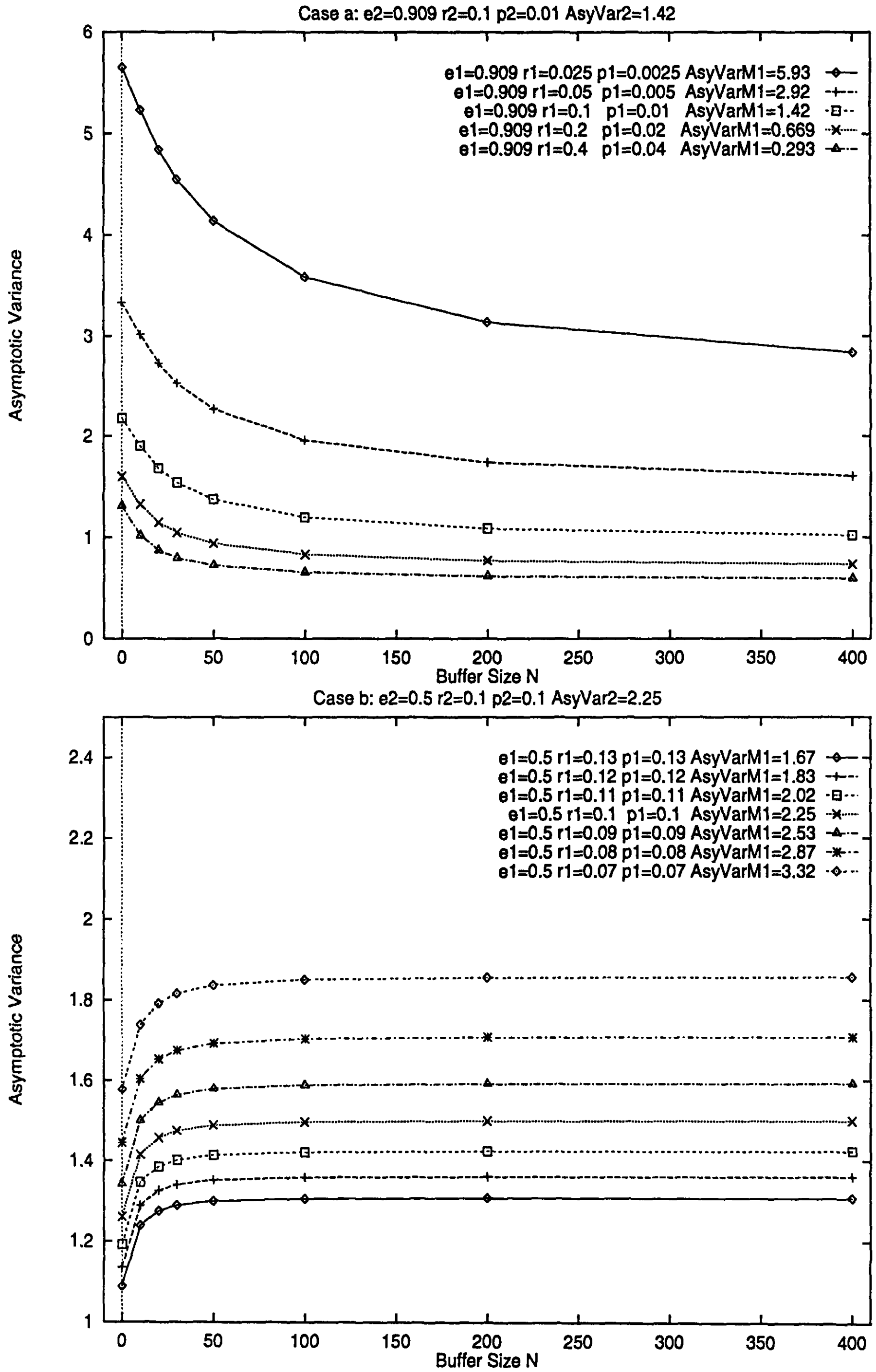

Figure 5-5: Example 3: Asymptotic variance rate of the output as a function of the buffer size

- $\Delta_{M_2} < \Delta_{(M_1,M_2,\infty)} < \Delta_{M_1}$:

  M_1 characteristics: $r_1 = 0.07,\ p_1 = 0.07$.

  $\Delta_{M_2} = 1.42 < \Delta_{(M_1,M_2,\infty)} = 1.856 < \Delta_{M_1} = 3.32$

- $\Delta_{(M_1,M_2,\infty)} < \Delta_{M_1}$ and $\Delta_{(M_1,M_2,\infty)} < \Delta_{M_2}$:

  M_1 characteristics: $r_1 = 0.13,\ p_1 = 0.13$.

  $\Delta_{M_1} = 1.67 > \Delta_{(M_1,M_2,\infty)} = 1.30$ and $\Delta_{M_2} = 1.42 > \Delta_{(M_1,M_2,\infty)} = 1.30$

We understand that both machines having the same efficiency is a special case and that $\Delta_{(M_1,M_2,\infty)}$ is different than in the rest of the cases. However, we have not been able to determine how $\Delta_{(M_1,M_2,\infty)}$ relates to Δ_{M_1} and Δ_{M_2}.

- *Close efficiencies:* When the efficiencies of both machines differ, $\Delta_{(M_1,M_2,\infty)}$ becomes Δ_{M_i} where M_i is the least efficient machine. We have studied the behavior of the system for $N \leq 400$ [2]. If the difference in the efficiency is small, $\Delta_{(M_1,M_2,400)}$ is still far from Δ_{M_i}. This means that N_ϵ is greater than 400. As the differences in the efficiencies increase N_ϵ decreases and eventually becomes smaller than 400.

  Fig. 5-6 explores the changes in the behavior of the system as the efficiencies of the two-machines become different. Case (a) starts with machine characteristics $r = 0.1, p = 0.02$ and modifies r_1 to obtain machines with very close efficiencies to the original one. Case (b) starts with machine characteristics $r = 0.1, p = 0.1$ and modifies p_1. As the difference in efficiencies becomes greater, $\Delta_{(M_1,M_2,N)}$ is closer to Δ_{M_i} when $N = 400$. Note that M_i (the less efficient machine) is not the same for all the curves in each plot.

  Fig. 5-7 shows the changes in the behavior of a system as the efficiencies of the two machines become different. In this case, the frequency of events of the two machines is very different. The behavior of the system in terms of $\Delta_{(M_1,M_2,\infty)}$ is the same as observed in Fig. 5-6: as the differences in efficiencies becomes greater, $\Delta_{(M_1,M_2,400)}$ is closer to Δ_{M_i}.

[2]The derivation of the asymptotic variance rate involves the inversion of a dense matrix of size $4(N-1)$, and this process is very computationally intensive.

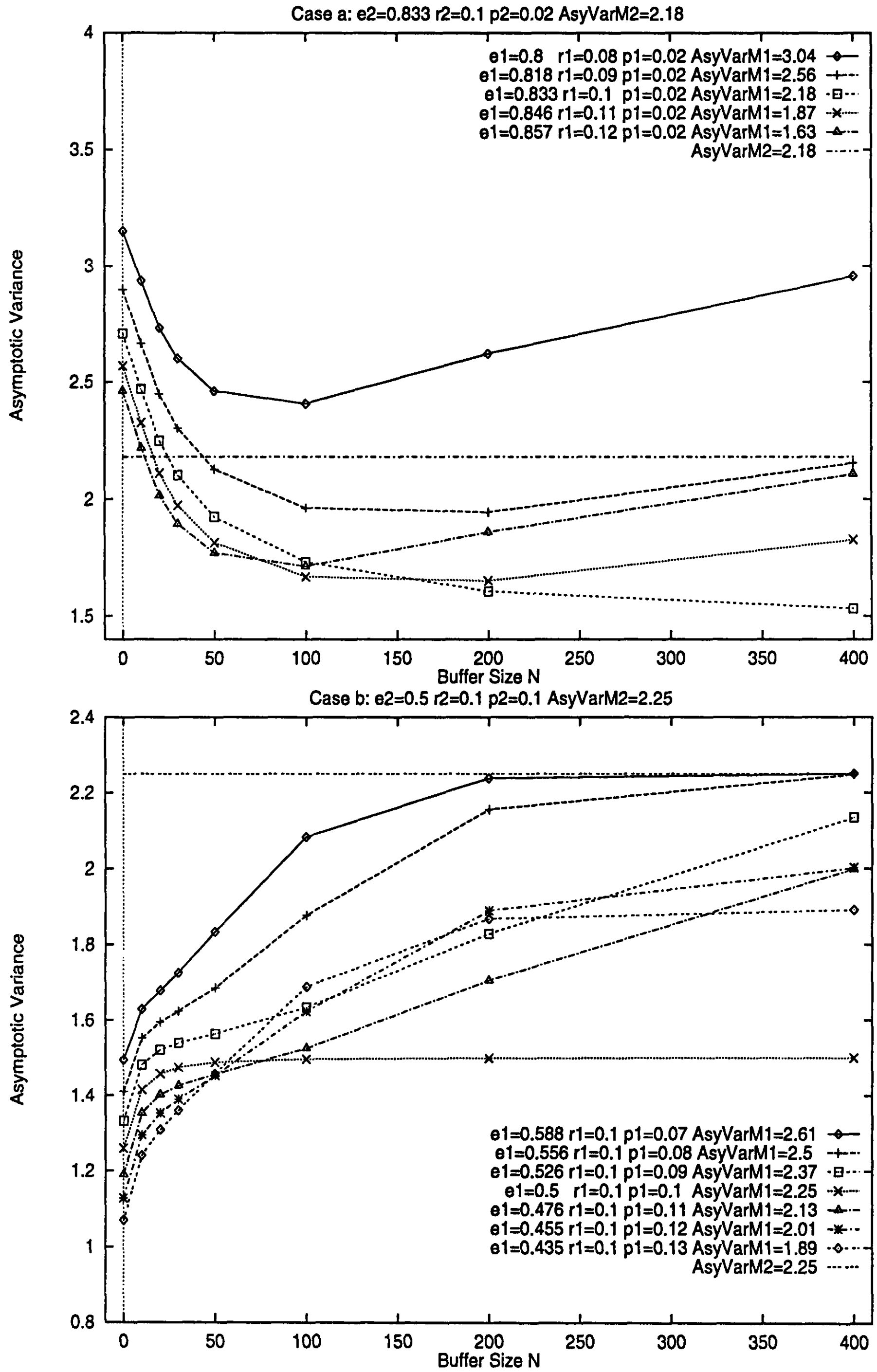

Figure 5-6: Example 4: Asymptotic variance rate of the output as a function of the buffer size

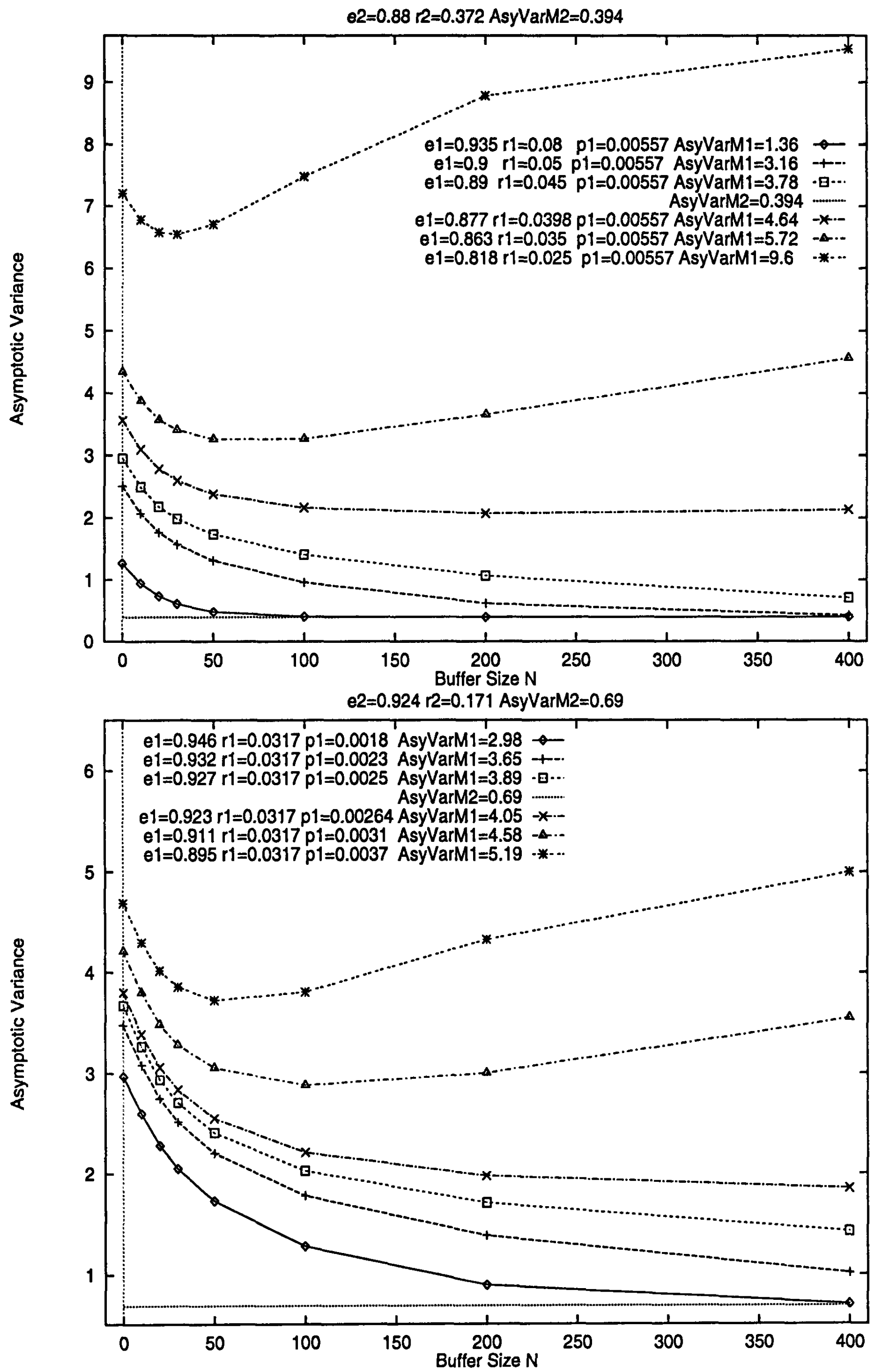

Figure 5-7: Example 5: Asymptotic variance rate of the output as a function of the buffer size

Upper and Lower Limit of the Asymptotic Variance Rate

- *Same efficiencies:* When the efficiencies of both machines are the same (Fig. 5-5), in all the cases observed $\Delta_{(M_1,M_2,N)}$ is between the range $\Delta_{(M_1,M_2)}$ and $\Delta_{(M_1,M_2,\infty)}$.

  It is interesting to compare the behavior in cases (a) and (b) of Fig. 5-5. In case (a), where $e = 0.909$, as N increases, $\Delta_{(M_1,M_2,N)}$ decreases. Comparing the different graphs we can see that as the frequency of the events of M_1 decrease both Δ_{M_1} and $\Delta_{(M_1,M_2,N)}$ increase. In case (b), where $e = 0.5$, as N increases, $\Delta_{(M_1,M_2,N)}$ increases. Comparing the different graphs we can see that as the frequency of the events of M_1, decrease Δ_{M_1} increases but $\Delta_{(M_1,M_2,N)}$ decreases. We cannot explain the reason for this difference in behavior.

- *Close efficiencies:* Fig. 5-7 and Fig. 5-6 show cases where the efficiencies of both machines are close. We observe something that we have already stated when referring to machines with different efficiencies. In some two-machine systems $\Delta_{(M_1,M_2,N)}$ is not between $\Delta_{(M_1,M_2,0)}$ and $\Delta_{(M_1,M_2,\infty)}$ for some N. In these cases there is a finite buffer size N' for which $\Delta_{(M_1,M_2,N')}$ presents a minimum. Again, we cannot explain why this happens.

5.4 Conclusions

From the experiments performed we can obtain the following conclusions regarding the evolution of Δ as the buffer size increases:

- Limit of the asymptotic variance rate ($\Delta_{(M_1,M_2,\infty)}$):

  - Different efficiencies of the machines: as the buffer size increases the asymptotic variance rate of the system becomes closer to the asymptotic variance rate of the least efficient machine.

    If $e_i > e_j$, then $\Delta_{(M_1,M_2,\infty)} = \Delta_{M_j}$.

- Same efficiency: as the buffer size increases the asymptotic variance rate of the system becomes smaller than the greatest asymptotic variance rate of the machines that form the system.

  $\Delta_{(M_1,M_2,\infty)} < \max(\Delta_{M_2}, \Delta_{M_1})$.

- Changes in $\Delta_{(M_1,M_2,\infty)}$ as N increases: This value may not always be within the zero and infinite buffer limits. In all the cases when this happens, it presents a minimum for some value of N'. There are M_1 and M_2 combinations for which

$$\Delta_{(M_1,M_2,N')} = \min(\Delta_{(M_1,M_2,N)}) \ \forall N \in [2, \infty)$$

- N_ϵ: The greater the difference in efficiencies between both machines, the smaller N_ϵ is for a given ϵ. The higher the frequency of the failures and repairs of the most efficient machines, the smaller N_ϵ is.

Chapter 6

Variance of the Output as a Function of Time

6.1 Model for the Variance of the Output

In this section, we propose an analytical model to calculate the variance of the output of a two-machine line with finite buffer size for an interval of time $[0, t]$. We discover that this model only represents the variance of the output of the system accurately for small t. We compare the results of the model with simulations and discuss the reasons why the model may not be accurate for long periods. We discuss the relationship between the variance of a system and its interdeparture distribution using the results derived by Buzacott and Shanthikumar (1993)

6.1.1 Description of the Model

Simplification The failure and repair characteristics of a two-machine line with finite buffer size can be approximated by a single machine with two failure modes. The steps to simplify the system are (Fig. 6-1) are:

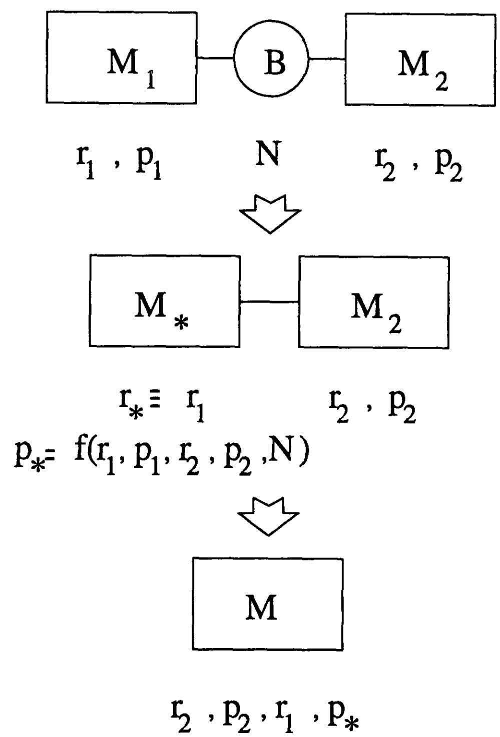

Figure 6-1: Simplification of the two-machine finite buffer line

1. The failure and repair distributions of a two-machine finite buffer line are very close to those of a two-machine zero buffer line (Section 4.3).

2. The production rate and variance of a two-machine zero buffer line is well approximated by a single machine with two failure modes (Section 3.4).

Representation Thus a two-machine finite buffer line can be represented by a single machine, two-failure mode system. A single machine, two-failure mode system is defined by a three-state Markov chain (Fig 6-2). The relationship of the three states with the original two-machine finite buffer system is:

- 1: System is producing pieces, at a rate of one piece per time unit.

- $F2$: System is down due to a failure of M_2.

- $S2$: System is down due to a starvation of M_2.

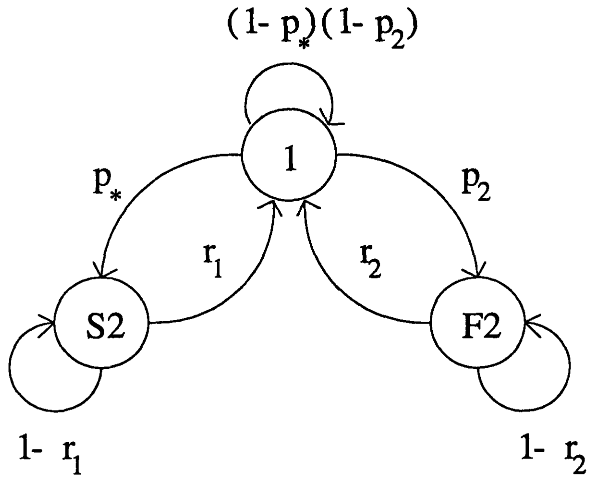

Figure 6-2: Representation of the two-machine system with finite buffer size

During a time unit when the system is operational (State 1), it may either produce a part (with probability $1 - p_1 - p_*$) or it may have an interruption of flow (with probability p_2 the interruption is of type $F2$ and with probability p_* if the interruption is of type $S2$). During a time unit when the system is not producing parts, it may either stay down (with probability $1 - r_1$ if it is an interruption of type $S2$ and with probability $1 - r_2$ if it is an interruption of type $F2$), or the interruption may finish (with probability r_1 if it is an interruption of type $S2$ and with probability r_2 if it is of type $S2$).

Assumptions The assumptions we are making to be able to represent our original system by the single machine two-failure mode system are:

- The distribution of the operational time between starvations is geometric. In Section 4.3.2 we have seen that the simulations performed indicate that the distribution is not geometric. However, we use this distribution to define one of the parameters of M_*, and therefore, we are assuming it is geometrically distributed.

- There is a productive period between failures. This is true most often although as we stated in Section 4.1 there could be a failure of M_2 ($F2$) directly following a starvation ($S2$).

Variance of the Output We have seen that a two-machine finite buffer line can be represented by a single machine with two failure modes, in terms of the failure and repair behavior. In Section 3.5 we derived a formula (3.11) for the variance of the output for an interval of length $[0, t]$ for a single machine with two failure modes. Therefore, we can apply this formula to obtain the variance of the output of the two-machine finite buffer line. The result is

$$\begin{aligned}
\sigma^2(t) = & \\
& (C\frac{r_1^2 p_2(2-p_2-r_2)+r_2^2 p_*(2-p_*-r_1)-2p_*p_2r_1r_2}{(p_2r_1+r_2p_*+r_1r_2)^3})t \\
& -2C\frac{(2+b-p_*-p_2-r_1-r_2)((r_2-r_1)^2-(p_*+p_2-b)^2)}{b(-b+p_*+p_2+r_1+r_2)^3}(1-\beta_1^t) \\
& -2C\frac{(-2+b-p_*-p_2-r_1-r_2)((r_2-r_1)^2-(p_*+p_2-b)^2)}{b(b+p_*+p_2+r_1+r_2)^3}(1-\beta_2^t)
\end{aligned} \tag{6.1}$$

where

$$C = \frac{1}{1+\frac{p_*}{r_1}+\frac{p_2}{r_2}} \tag{6.2}$$

$$b^2 = (r_2-r_1)^2+(p_*+p_2)^2+2(r_2-r_1)(p_2-p_*) \tag{6.3}$$

$$\beta_1 = 1-\frac{p_*+p_2+r_2+r_1-b}{2} \tag{6.4}$$

$$\beta_2 = 1-\frac{p_*+p_2+r_2+r_1+b}{2} \tag{6.5}$$

6.1.2 Comparison with Simulations

We have modeled the variance of a two-machine line with finite buffers by transforming the system into a single machine. Then we have used the variance of this system to predict the variance of the original one. We know that the single machine system represents accurately the behavior of a two-machine zero buffer system (Section 3.4).

The final single machine two-failure mode system obtained in Section 6.1.1 has the same mean production and very close failure and repair characteristics as the original two-machine finite buffer one. Therefore, we would expect that the formula that determines the variance of the output of the second system would work well to predict the variance of the original.

We have performed simulations to verify the results obtained by the model described above[1]. The comparison shows that the model is only accurate for short periods $[0, t]$. Here, we offer an explanation of when and why the model works.

Example Fig. 6-3 compares the results obtained from the model and from simulations for several cases. We use these results to illustrate the discussion.

Discrepancies between the Analytical Model and Simulation As we can see from the examples, the model does not predict the variance of the original system for the entire range of times.

The most important observation about the system is that in the long term the analytical model used does not reproduce some shapes that can be obtained from simulation. In Section 3.5, we see that the variance rate of the output versus time for a single machine two-failure mode system has a concave shape and reaches a limit asymptotically. However, in Fig. 6-3 some of the shapes obtained from simulation for non-zero buffer systems do not display this pattern.

In spite of the discrepancies in the long term, the model predicts accurately the variance of the output for short periods of time. The length of this interval of time has not been quantified in detail but it seems to be smaller than the mean time between starvations.

[1]The procedure used to run the simulations is described in Section 7.1.

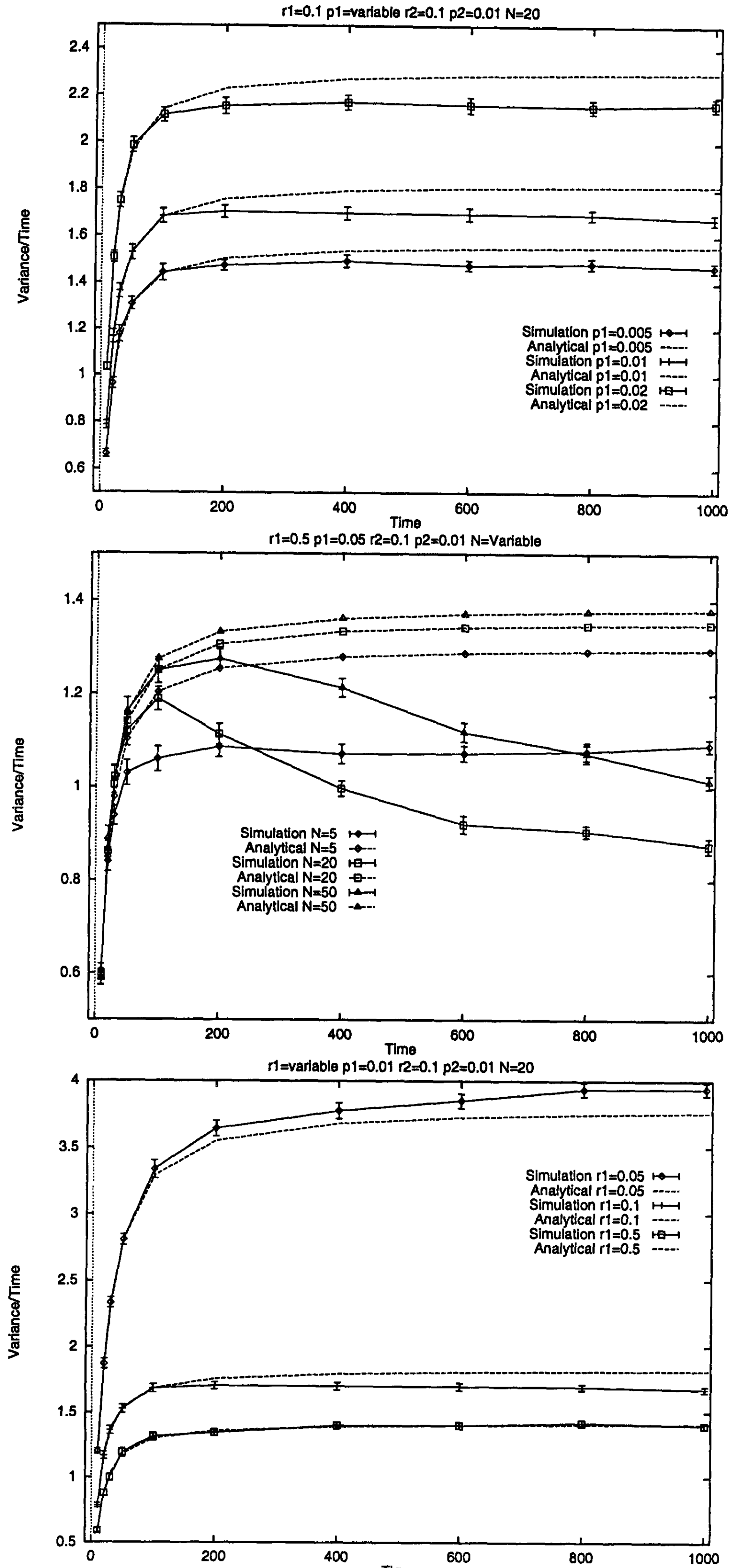

Figure 6-3: Variance of the output of a two-machine finite buffer size: analytical vs. simulation results

6.1.3 Correlation of the Interdeparture Times

The interdeparture time is the number of time units between two consecutive departures of parts from the system. The model derived in Section 6.1.1 reflects accurately the variance of a two-machine zero buffer system, but it is not valid for all the values of t in the case of a finite buffer model. Therefore, there is a difference between the behavior of the finite buffer and the zero buffer systems. We believe that the difference can be explained by the correlations of the interdeparture times that appear in the system as we introduce a buffer.

The ith interdeparture correlation coefficient ($\varphi(i)$) represents the correlation between the departure of the piece that is leaving the system this time unit and the departure of the ith piece that leaves the line after this one. We have also calculated the partial autocorrelation coefficients ($\Phi(i)$) that indicate the correlation between these two departures after their mutual dependency on the intermediate departures is removed.

Determination of the autocorrelation coefficients A more complete derivation is presented in Wei (1993). Let

v be a vector formed by a series of m observations of the interdeparture times $z_1, z_2, ..., z_m$.

$\varphi(i)$ be the correlation coefficient between the tth interdeparture time (z_t) and the ith interdeparture time after it (z_{t+i}).

$\Phi(i)$ be the partial autocorrelation coefficient between the tth interdeparture time (z_t) and the ith interdeparture time after it (z_{t+i}) after their mutual dependency on the intervening variables $z_{t+1}, z_{t+2}, ..., z_{t+k-1}$ has been removed.

The formulae used to determine the correlation coefficients are

$$\varphi(i) = \frac{\sum_{t=1}^{m-k}(z_t - \overline{z})(z_{t+k} - \overline{z})}{\sum_{t=1}^{m}(z_t - \overline{z})^2}$$

and $\Phi(i)$ are obtained recursively as follows

$$\Phi(1) = \phi(1,1) = \varphi(1)$$

For $i > 1$,

$$\begin{aligned}
\Phi(i) &= \phi(i,i) = \frac{\varphi(i) - \sum_{j=1}^{i-1}\varphi(i-j)\phi(i-1,j)}{1 - \sum_{j=1}^{i-1}\varphi(j)\phi(i-1,j)} \\
\phi(i,j) &= \phi(i-1,j) - \Phi(i)\phi(i-1,i-j) \text{ for } j < i.
\end{aligned}$$

Simulation Method The procedure used to obtain the correlation coefficients is as follows: for every system, 100 simulations have been run for a length of 10^5 time units after a warm-up period of 2400 time units[2]. For these values we have obtained the mean of the correlation coefficients and the 95% confidence interval. The number of cases run has been limited by the time-consuming simulation needed. As the buffer size increases or the frequency of the events is lowered, the number of simulations needed to get valid results increase significantly. In all cases we have obtained just the first 10 correlation coefficients.

Comments on the Experiments For a given two-machine system, we have run the experiments for different N. Fig. 6-4 shows $\Phi(i)$ and $\varphi(i)$ for $r_1 = r_2 = p_1 = p_2 = 0.1$. Fig.6-5 shows $\Phi(i)$ for $r_1 = p_1 = 0.5, r_2 = p_2 = 0.1$ and its reversed system.

The following conclusions can be derived from the results obtained:

- $\varphi(i)$ *vs.* $\Phi(i)$ Fig. 6-4 shows that both $\varphi(i)$ and $\Phi(i)$ for a given system are very

[2]The simulator is described in Section 7.1.

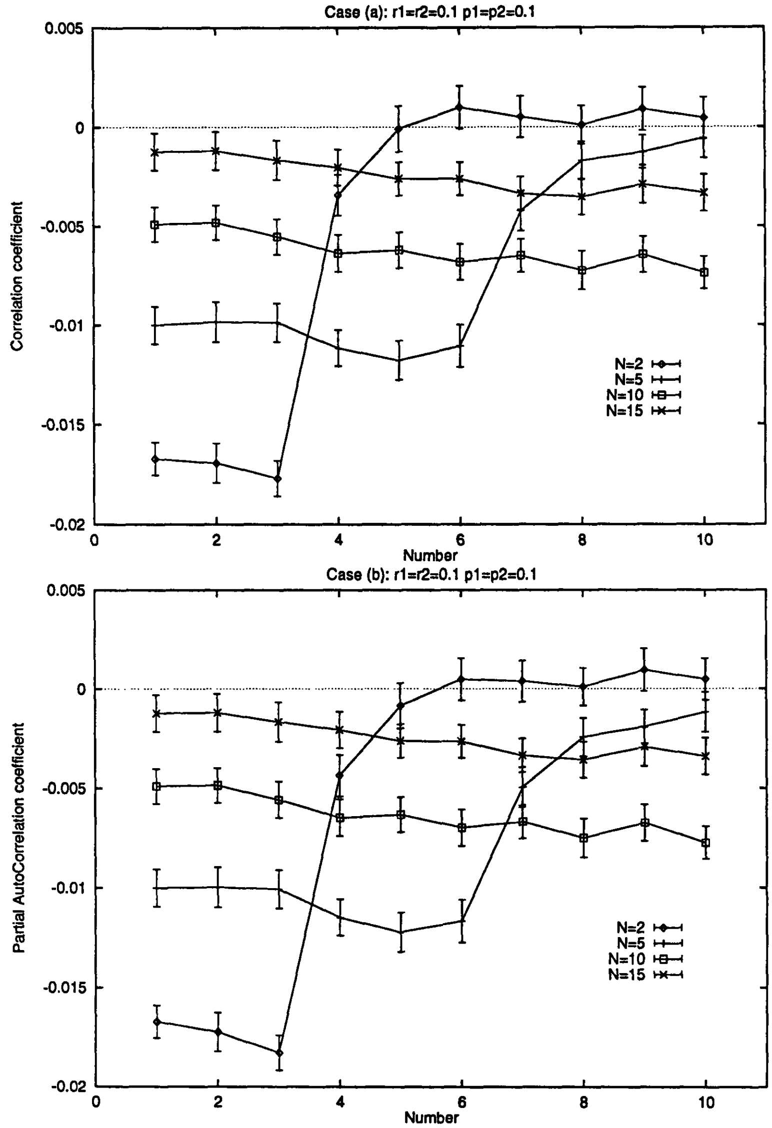

Figure 6-4: Example 1: Correlation coefficients

close. This means that the influence of an event that takes place at this time in the system affects the ith departure from the line directly and that the influence of the intermediate departures does not change this relationship. Therefore, in the rest of the analysis we just concentrate in $\Phi(i)$. The conclusions we propose are also valid for $\varphi(i)$.

- *Changes in $\Phi(i)$ as i increases :* In all the cases, if the buffer size is N, $\Phi(i)$ is significant $\forall\ i \leq N+1$. The buffer size N determines the maximum number of pieces in a system at a given time. The correlation coefficients indicate that the events of the system that take place at this time unit influence the N following departures of the line. A event that takes place now affects the departure times of all the material currently in the system. We can also see that $\Phi(i)$, for $i > N+1$ become increasingly less significant and that this process takes longer as N increases.

- *Negative $\Phi(i)$:* It is also important to mention that $\Phi(i)$ and $\varphi(i)$ are in general negative. This means that after a long interdeparture time there are more likely to be short interdeparture times and vice versa. If there are no failures in the system, the interdeparture time is 1. If there is an interruption, the interdeparture time will be greater and, after the system recovers from the failure, chances are that several pieces will come out with an interdeparture time of 1.

  $\Phi(i)$ becomes more negative as i increases, because longer productive periods are less likely to happen.

- *$\Phi(i)$ as a function of N:* Another interesting observation comparing the values obtained for different N is that the value of the autocorrelation coefficients decreases as the number of significant coefficient increases. This means that the events that take place at this time influence the behavior of the system for a longer period but that their influence is smaller.

- *A system and its reverse:* In Fig. 6-5 we see that if the two machines are different, the correlation structure is different for a system and its reverse. This makes sense because the failure and repair characteristics of the machines are different.

In the examples shown in this section we have seen that the agreement between simulations and the analytical results is very good for the zero-buffer systems and that the agreement becomes worse as the buffer size increases. The number of autocorrelation coefficients that are significant also increase with the buffer size.

This result also agrees with some previous work by Heindrick (1992). Heindrick did some experiments that showed the effect of buffer capacity in the correlation structure of a three-machine line. The model that he used was continuous time, exponential processing time, perfectly reliable machines. The number of significant autocorrelation coefficients was $N+1$, N being the size of the two buffers in the system. Also, as the buffer size increased, the values of these coefficients decreased.

6.2 Interdeparture Distribution vs. Variance of the Output

Interdeparture distribution For the model described in Section 4.4.1, Buzacott and Shanthikumar (1993) calculate the interdeparture distribution. Let $f_{(\overline{d})}(k)$ represent the probability that the time between consecutive departures of a two-machine line $(\overline{d})$ is k. Buzacott and Shanthikumar (1993) prove that the interdeparture distribution for a two-machine finite buffer size system is equal to the interdeparture distribution of a two-machine line with zero buffers where the parameters of the first machine are the parameters of the equivalent machine obtained in Section 4.3. The complete derivation is presented in Buzacott and Shanthikumar (1993).

Let $f_{\hat{d}}(k : p, r)$ be the probability that the time between consecutive departures $(\hat{d})$ is k for a single machine with parameters r and p. The probability of the interdeparture time being 1 $(f_{\hat{d}}(1 : p, r))$ is the probability of the machine not failing during

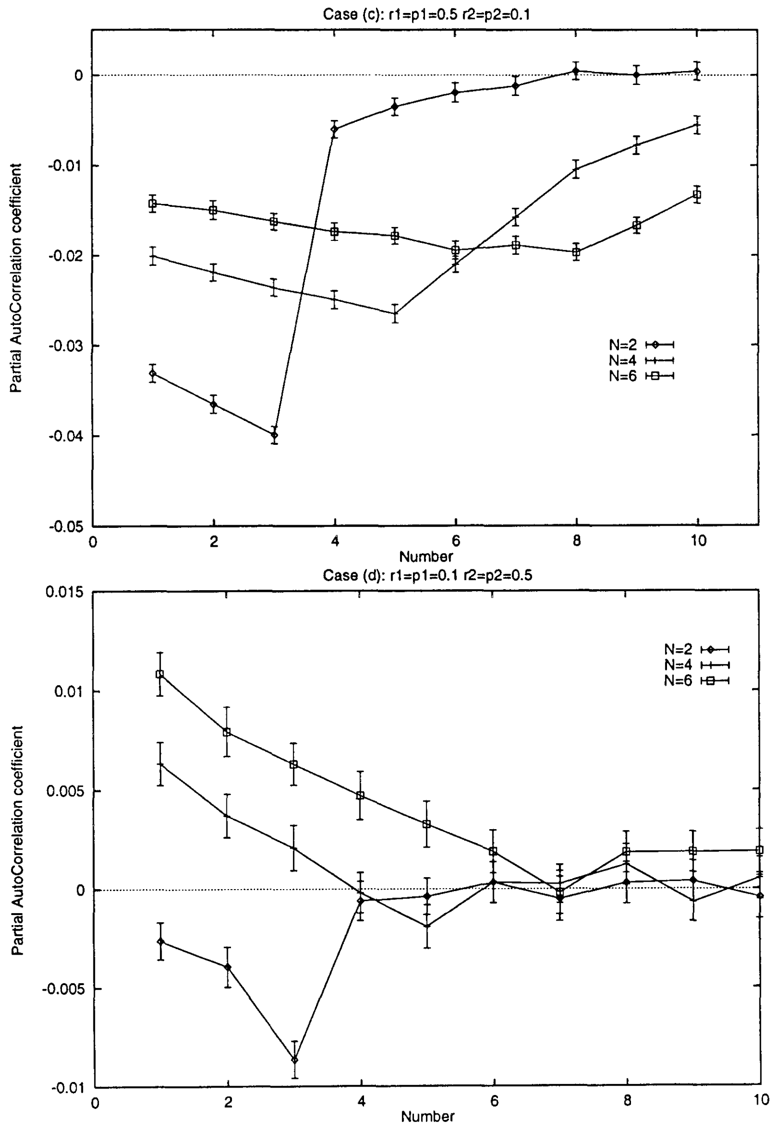

Figure 6-5: Example 2: Correlation coefficients

that time unit $(1-p)$. The probability of the interdeparture time being k time units, $k \geq 2$ ($f_{\hat{d}}(k : p, r)$) is the probability of machine failing and being down for exactly $k-1$ time units.

$$f_{\hat{d}}(k : p, r) = \begin{cases} 1-p & k = 1 \\ rp(1-r)^{k-2} & \forall k \geq 2 \end{cases}$$

Then, $f_{\overline{d}}(k)$ for a two-machine finite buffer size satisfies:

$$\begin{aligned} f_{\overline{d}}(k) &= f_{\hat{d}}(k : p_2, r_2) + f_{\hat{d}}(k : p_1', r_1) - \\ & \quad f_{\hat{d}}(k : p_1' p_2, r_1 + r_2 - r_1 r_2) \end{aligned} \tag{6.6}$$

This expression describes the probability of the interdeparture time being k as the algebraic sum of the probabilities of the interdeparture time being k for three single machines:

- M_2: Machine parameters are r_2 and p_2.
- Equivalent machine to M_1 and the buffer: Machine parameters are r_1 and p_1'.
- Combined machine: Machine parameters are $r_2 + r_1 - r_1 r_2$ and $p_1' p_2$. This machine represents the situation where both machines fail during the same time unit. The term representing the behavior of this combined machine is subtracted from the previous ones.

This expression also describes the interdeparture distribution for a two-machine zero-buffer line with machine M_1 parameters being r_1, p_1' and M_2 parameters being r_2, p_2.

Fig. 6-6 shows the interdeparture density distribution for three different two-machine finite buffer lines and their equivalent zero buffer systems. As Buzacott and Shanthikumar (1993) calculated, the two distributions are identical and agree with equation (6.6) presented above. These graphs have been obtained by simulating the behavior of both systems for 10^7 time units. The simulations and the determination of the parameters of the equivalent machine have been calculated using Gershwin's model. The closeness of these results with the interdeparture distribution calculated by Buzacott and Shanthikumar provide another argument to support the hypothesis that the most important assumptions in both models are very similar.

6.3 Conclusion

Buzacott and Shanthikumar (1993) and we derived a two-machine zero buffer system as a way of representing the original system. We tried to use the simplified system as a way to derive the variance of the original system. Buzacott and Shanthikumar use it to derive the interdeparture distribution of the original system.

Though the derivation of the interdeparture distribution is accurate, it cannot be used to estimate the variance of the original system. The simulations performed in Chapters 6 and 7 prove that the variance of the two systems differ. The correlations between interdeparture times modify the output pattern of the system. Thus, though the steady state interdeparture distribution is the same, the sequence of interdepartures differ. This makes a significant difference in the variance of the output produced during a period of time.

However, the model derived in this chapter predicts the variance of the system works in the short term. Therefore, we can conclude that for short periods of time the correlation structure does not alter significantly the variance of the system. In Chapter 7 we further discuss the changes in the variance of the system over time and we identify different time frames for its evaluation.

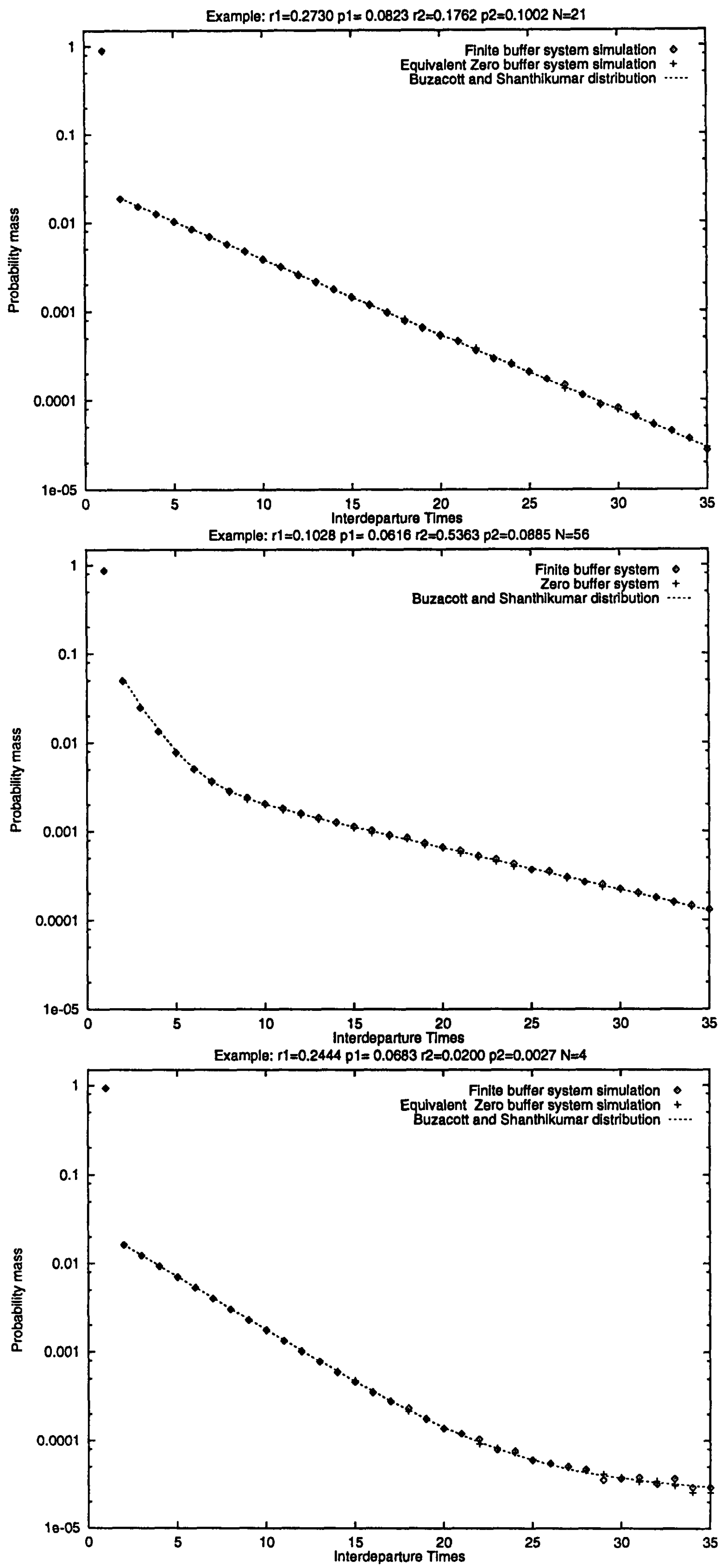

Figure 6-6: Interdeparture distribution for a two-machine finite system and its equivalent zero-buffer system

Chapter 7

Variance of the Output as a Function of Time: Simulation Results

The purpose of this chapter is to present some observations about the variability of the output of a production line. These results have been drawn by analyzing simulation experiments. In this chapter, we focus on understanding the behavior of the variance for different periods of time $[0, t]$ as the time increases. We have observed here are three different time frames where the variance of the output of the system is determined by different factors. Studying simultaneously the behavior of a system and its reverse is a powerful tool to determine these time frames and to derive conclusions about the behavior.

7.1 Simulation Procedure

A simulation software has been used to run the experiments shown in this thesis[1].

[1]This program was written by Asbjoern M. Bonvik. It uses almost all the conventions defined in Gershwin's model (1993). The only difference in assumptions is that in the simulation the time unit just after a machine is blocked/starved it is allowed to fail, and this is not allowed to happen

The procedure used to calculate the variance of a system over a specific time t is as follows: we perform $N = 50$ replications of the simulation. Each of these replications gives us a value for the variance of the production ($\sigma_i^2(t)$). From the Law of Large Numbers (Drake, 1967), when N is large enough, the mean of these variances follows a normal distribution, so we can use the $\sigma_i^2(t)$ samples obtained to approximate the variance and a confidence interval around this value:

$$\text{Variance estimate} \qquad \sigma^2(t) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2(t)$$

$$\text{95\% Confidence Interval} \qquad \sigma^2(t) \pm 1.96\sqrt{\frac{\sum_{i=1}^{N}(\sigma_i^2(t) - \sigma^2(t))^2}{N}}$$

Each replication runs for 800 periods of time of length t, after a warm-up period of 2400 time units. These periods are independent of each other as they are separated by 1000 time units. The variance of the production during these periods is recorded as $\sigma_i^2(t)$. This way of performing the replications assures us that, overall, the probability of the system being in a particular state at the beginning of a period is the steady state probability of that state.

7.2 Time Frames for the Study of the Variance

From the results of the simulations performed, and comparing the performance of a system and its reverse, we have come to the conclusion that the study of the variance of a system can be subdivided into three time regions:

- Short Term: This period of time is not long enough for the interdeparture correlation structure to affect the output pattern. Though the correlation coefficients are significant, the number of transitions between interruptions of flow and productive periods is not big enough for them to affect the variance significantly. This is the reason why the model described in Section 6.1.1 predicts accurately

in the model. This is a very minor difference, so the results obtained from simulation can be used to interpret the behavior of the system.

the variance of the output for this time frame in spite of not accounting for correlation. In general, the variance of the output depends more significantly on the performance of M_2. If both machines are different, the variance of a system and its reverse is different throughout this period.

- Medium Term: The start of this period is defined by the fact that the correlations start affecting the output behavior of the system. It finishes when the variance of the systems gets close to its asymptotic value. In the case of a system formed by two different machines, the variance of the two systems is different throughout this period.

- Long Term: This period of time is characterized by the fact that the variance of the system is linear over time and Miltenburg's asymptotic variance rate describes the output behavior of the system. The time it takes a system and its reverse to get close to the asymptotic value may be different. When both systems reach the long term, the variance of both is the same.

Example of a system and its reverse: short, medium and long term behavior We use Fig. 7-1 as an example to explore the behavior of a system and its reverse as a function of time. The characteristics of the system are $r_1 = 0.5,\ p_1 = 0.05,\ r_2 = 0.1,\ p_2 = 0.01,\ N = 20$ and $\overline{n} = 10$.

If we observe the system for a short period of time, the output of the system depends more heavily on the performance of M_2 than on the performance of M_1. For example, if we want to calculate $\sigma^2(10)$, it is mostly a function of the performance of M_2 during this time, because the maximum number of pieces that can be produced is 10, and on average that many are in the buffer. There are moments where the number of parts in the buffer is less than 10, and then the output might be affected by the performance of M_1, but $\sigma^2(10)$ is determined mostly by M_2. The number of transitions between productive and non-productive periods is very small. Therefore, the correlation structure does not significantly alter the variance.

As the time frame increases, the influence of M_1 becomes greater. If we want

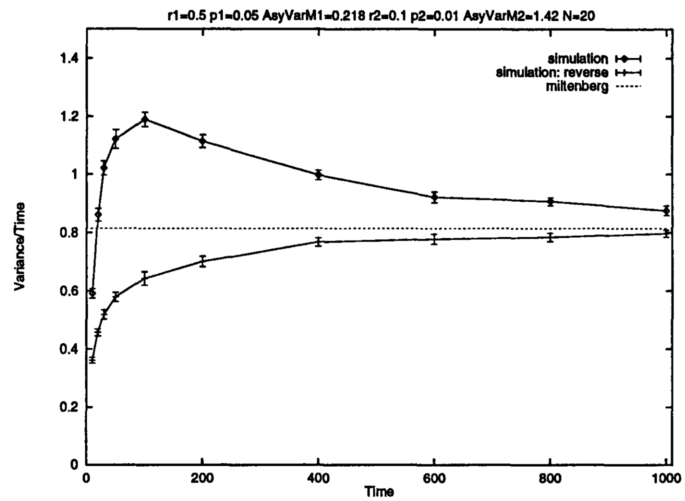

Figure 7-1: Time frames in the variance of a system

to calculate $\sigma^2(50)$ it depends on both M_2 and M_1. The reason is that most of the output has to be processed by M_1 during this period before it can be used by M_2. This period is long enough that the interdeparture correlation structure starts to affect significantly the variance of the period.

If the time frame is long enough, the variance depends equally on the performance of both M_1 and M_2, and the system reaches the asymptotic value.

This argument explains why the behavior of a system and its reverse is different in the short and medium term. The output is determined more heavily by the behavior of the last machine, and, if the two machines have different characteristics, it makes sense that the variances are different. It also explains why the behaviors converge in the long term, as the output in this longer time frame equally depends on the performance of both machines.

An interesting observation that we made when describing the behavior of the system in the long term, is that the time it takes a system and its reverse to get to the long term behavior may be different. Fig. 7-1 shows that the reversed system gets quicker to Miltenburg's asymptotic value than the original one.

Boundaries The boundaries between these time frames have not been derived. However, two different tools have been used to approximately define these boundaries once a simulation has been performed:

- Model for the variance of the output: The time frame when this system works corresponds to the short term behavior of the system. Therefore, the time where the results start diverging from simulation results indicates that the correlation structure starts affecting the variance of the output.
- A system and its reversed: In the case of a system formed by different machines, the study of the variability of a system and its reverse can be used to determine the time where the system is starting the long term behavior. The information that this provides is equivalent to comparing the variance obtained from simulation with Miltenburg's asymptotic variance rate.

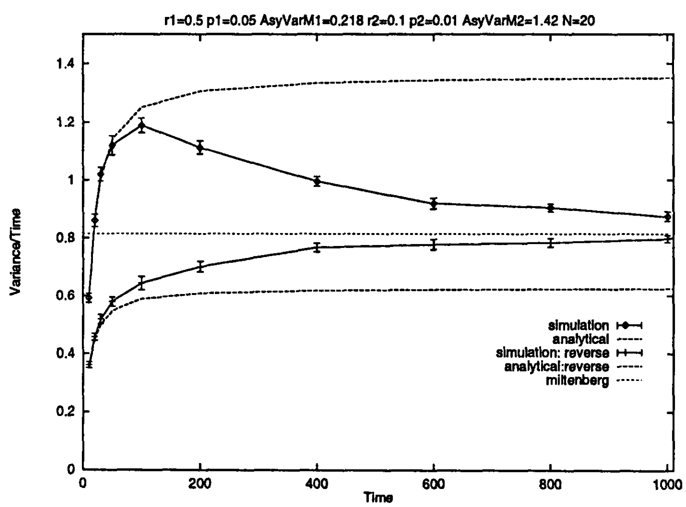

Figure 7-2: Boundaries of the time frames in the variance of a system

Fig. 7-2 shows the variance rate of the system $r_1 = 0.5$, $p_1 = 0.05$, $r_2 = 0.1$, $p_2 = 0.01$, $N = 20$ and its reversed one. Comparison with Miltenburg's asymptotic variance rate indicates that the long term behavior starts at $t \ll 1000$ for the system and at $t \approx 1000$ for its reversed one. Comparison with the results obtained from the analytical model indicates that the short term behavior of the system ends at $t \approx 40$.

7.3 Effect of the Buffer Size in the Variance

In Section 3.3 we discussed the effect of the buffer as a decoupling element. When the buffer is neither full or empty, the machines can work without their behavior being affected by the performance of the other machine. As the buffer size increases, the probability of starvation and blockage decrease and thus, the performance of a machine is less influenced by the presence of the other. This idea is confirmed by the results obtained from simulations.

Another interesting observation from the simulations is that the time it takes a system to reach the long term state is longer as the buffer size increases. In this section we present some examples of this effect and give an explanation of this phenomena.

Machines with the Same Efficiencies Fig. 7-3 illustrates the changes in the variance for a system where $r_1 = 0.5,\ p_1 = 0.05,\ r_2 = 0.1,\ p_2 = 0.01$ as $N = 2, 5, 20$ and 50. As the two machines have the same efficiency, $\overline{n} = N/2$.

In the case of the zero-buffer ($N = 2$) system, the variance of the output of a system and its reverse is the same, because the failure of any machine immediately affects the output of the line. As the buffer size increases, the material in the buffer makes the variance of the output in the short term more dependent on the performance of the last machine. The failures of the last machine affect the output of the line immediately, but the failures of the first machine affect the output depending on the amount of material in the buffer. If the characteristics of the machines are different, the short term behavior of the systems differ. The bigger N, the more material in the buffer and the variance of a system and its reverse diverge more in the short term. Due to the same reason, the time it takes the variance of both systems to converge increases as N increases.

Fig. 7-4 illustrates the changes in the variance for a system with identical machines $r = 0.1,\ p = 0.1$ for $N = 2, 20$ and 50. In this case, the system coincides with its reverse. As N increases, the time it takes the system to converge to its asymptotic

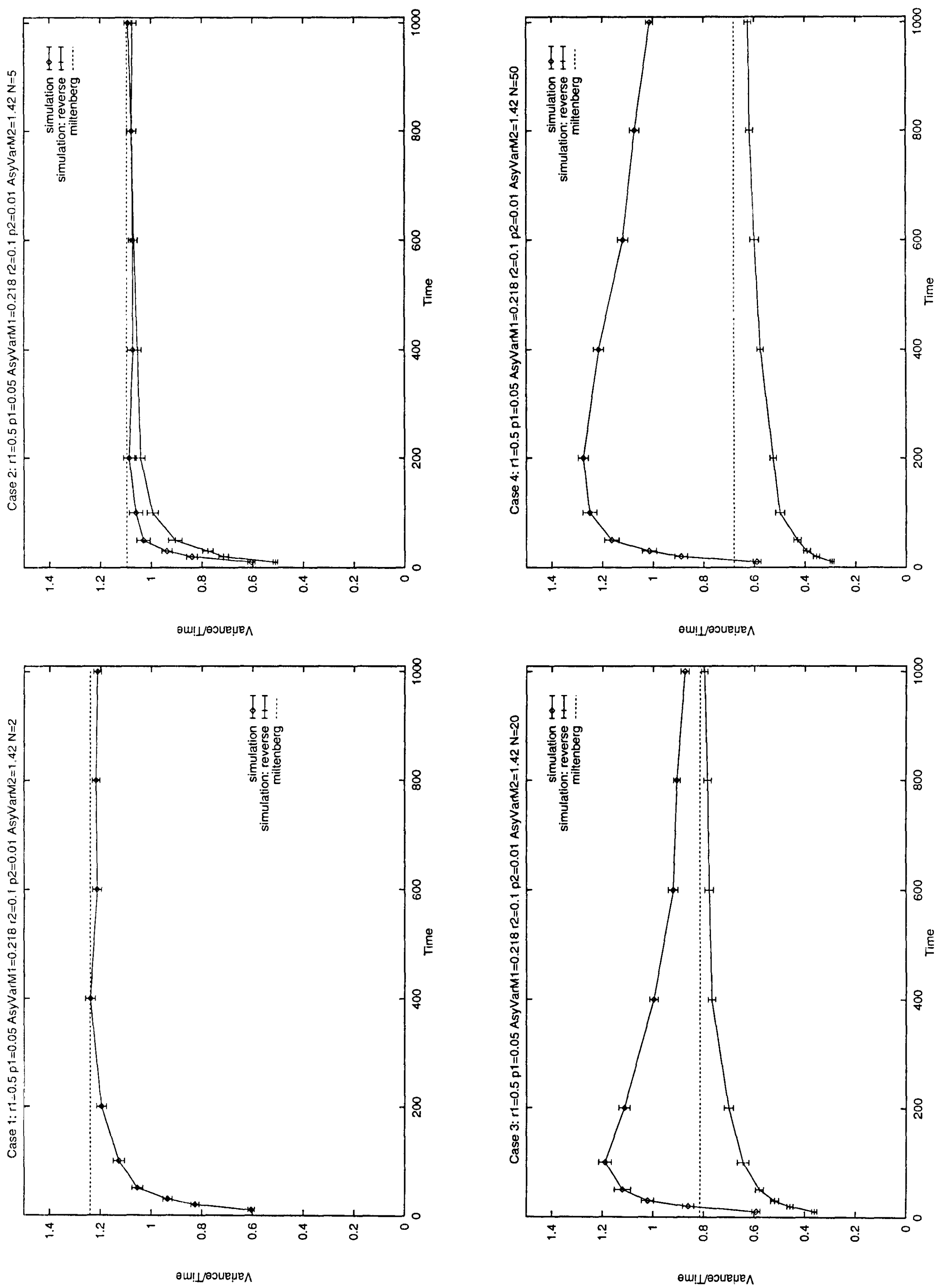

Figure 7-3: Example 1: Evolution over time of the variance of a system

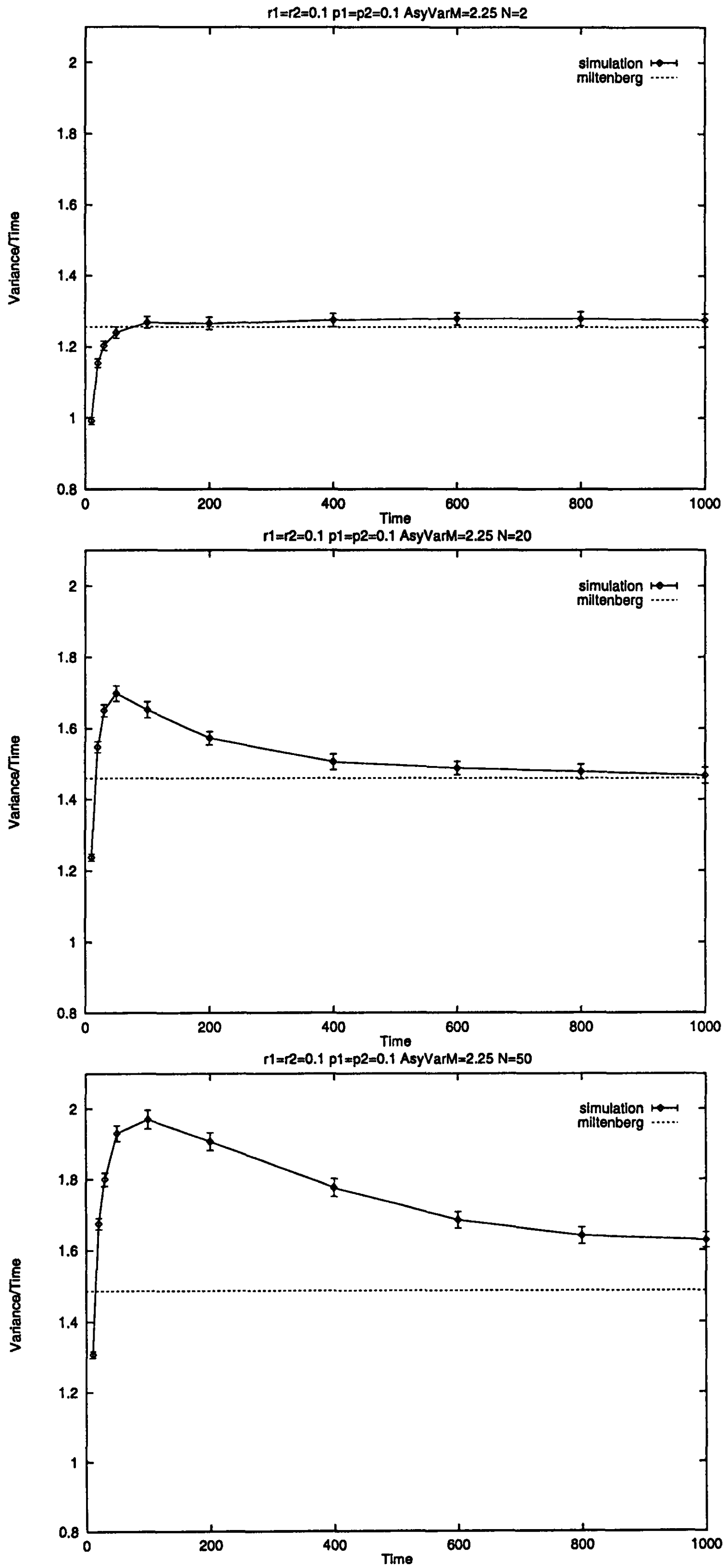

Figure 7-4: Example 2: Evolution over time of the variance of a system

value increases substantially.

$N \to \infty$ These examples can also be used to explore the behavior when $N \to \infty$. Fig. 7-5 shows the variance of M_2 (as if it was working in isolation) with the variance of the system for $N = 50$ for the case of $r = 0.1$, $p = 0.1$.

As N increases, the behavior of the system in the short term gets closer to the behavior of the single machine. This agrees with the previous argument that the material in the buffer makes the behavior of the system be that of M_2 for short periods of time. We know that as N increases the system takes longer to reach the asymptotic variance rate.

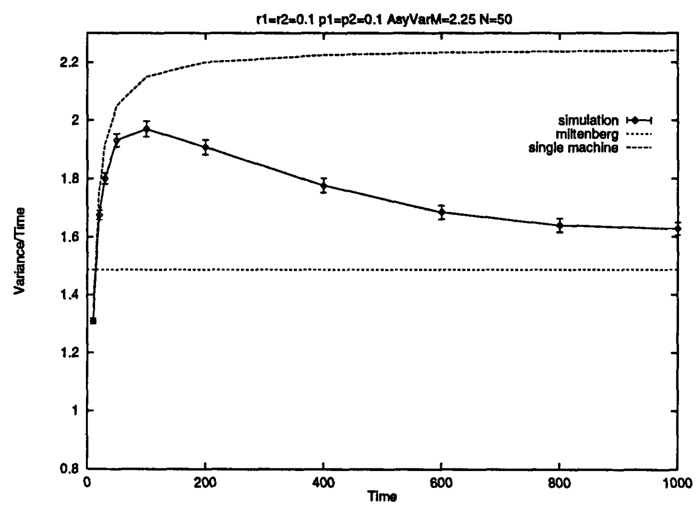

Figure 7-5: Same efficiency: Evolution of the variance as $N \to \infty$

In Fig. 5-5, we calculated $\Delta_{(M_1,M_2,N)}$ for this production line, up to $N = 400$. $\Delta_{(M_1,M_2,N)}$ seemed to reach a limit $\Delta_{(M_1,M_2,\infty)} \approx 1.5$. Is this value going to be the limit as $N \to \infty$ or is the system going to behave as the single machine M_2?. As we stated in Section 5.2.2 the steady state has no meaning when the buffer is infinite if the machines have the same efficiency because the system never reaches steady state. Therefore, we cannot provide an answer to this question. However, if the buffer is finite and big enough, the system eventually gets to $\Delta_{(M_1,M_2,\infty)}$, though the bigger N, the longer it will take.

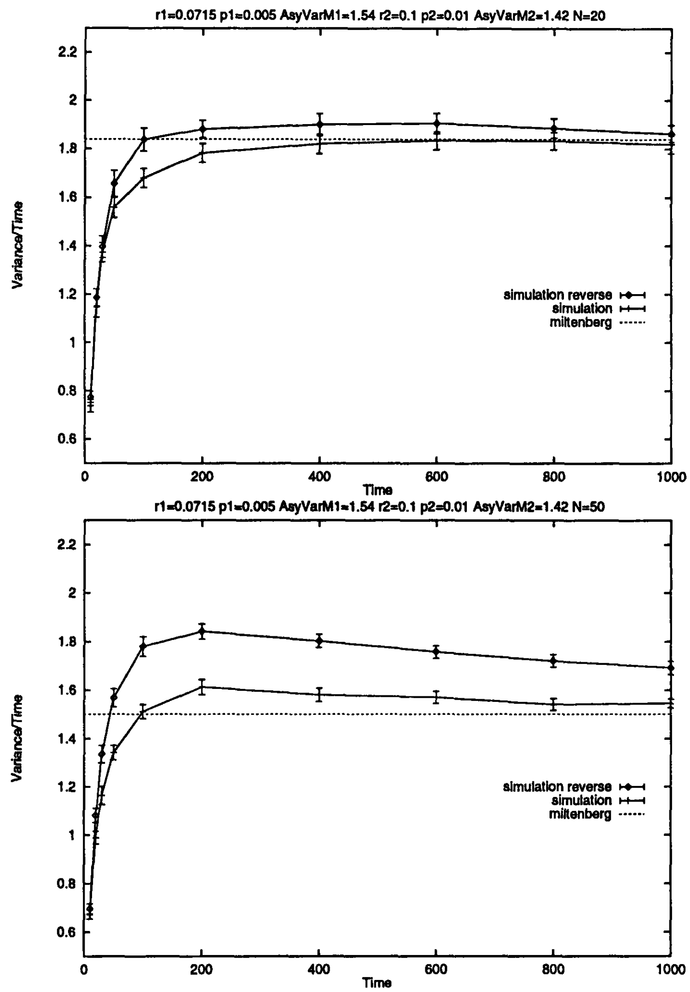

Figure 7-6: Example 3: Evolution over time of the variance of a system

Machines with Different Efficiencies The explanation provided for production lines formed by machines with the same efficiencies is valid for lines with machines of different efficiencies. As N increases the behavior of a system and its reversed line diverge in the short term. Fig. 7-6 presents the system $r_1 = 0.0715$, $p_1 = 0.005$, $r_2 = 0.1$, $p_2 = 0.01$ and Fig. 7-7 the system $r_1 = 0.2$, $p_1 = 0.05$, $r_2 = 0.1$, $p_2 = 0.01$.

There is one difference the case of machines with different efficiencies compared with the case of machines with the same efficiencies. When the same machines have the same efficiency, as N increases both a system and its reverse take a longer time to get to the asymptotic behavior. This is not true in the case of different efficiencies:

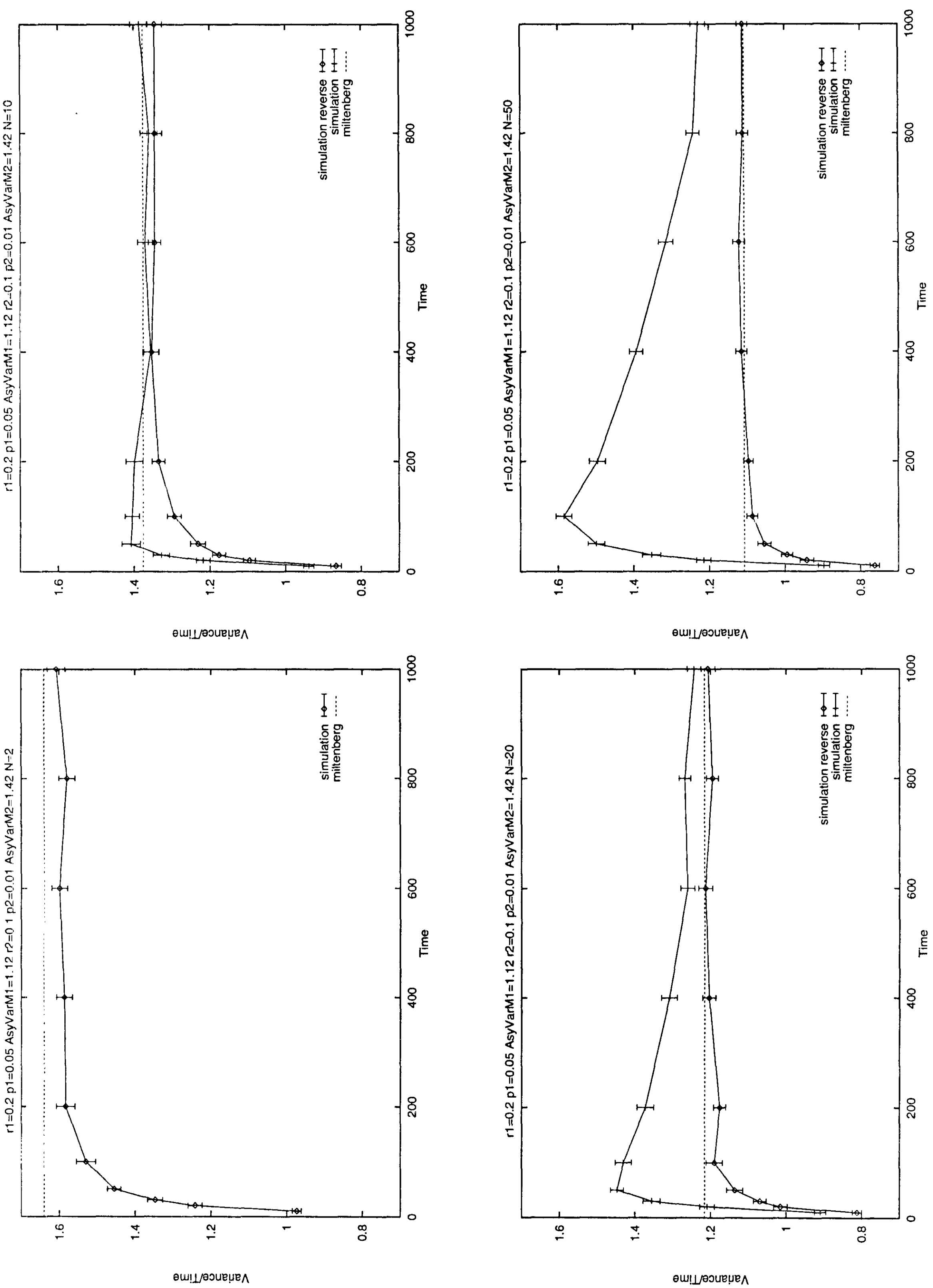

Figure 7-7: Example 4: Evolution over time of the variance of a system

- If M_2 is less efficient than M_1, $\overline{n} > N/2$, and as N gets bigger, M_2 is protected by the buffer level from almost all the failures of M_1. As N increases, the variance of the system gets closer to the variance of M_2. Therefore, the short term and medium term length decrease as N increases.

- In the reversed system, $\overline{n} < N/2$, and though the failures of the less reliable machine affect the performance of the system less as N increases, they always affect the variance of the output. As N increases, the amount of material in the buffer makes this effect happen later in time so this system exhibits a longer short and medium term behavior.

The bigger the difference in efficiency between the machines, the more significant this behavior. Fig. 7-8 shows this evolution for $r_1 = 0.2,\ p_1 = 0.05,\ r_2 = 0.1,\ p_1 = 0.01$. When the less reliable machine ($r = 0.2,\ p = 0.05$) is M_2, as N increases the variance of the system gets increasingly closer to the variance of M_2, and it takes less time for the system to get to the asymptotic variance rate. When the more reliable machine ($r = 0.1,\ p = 0.01$) is M_2, the system takes longer to get to the asymptotic variance rate as N increases.

$N \to \infty$ As in the case of machines with the same efficiency, we can explore the changes in the variance of the system as $N \to \infty$.

- If the less efficient machine is M_2, the behavior of the system approaches that of M_2 as N increases.

- If the less efficient machine is M_1, the system takes longer to converge to the asymptotic value as N increases. However, if N is finite and big enough, it will eventually get to $\Delta(M_1, M_2, \infty)$. If the buffer is infinite the system never reaches steady state.

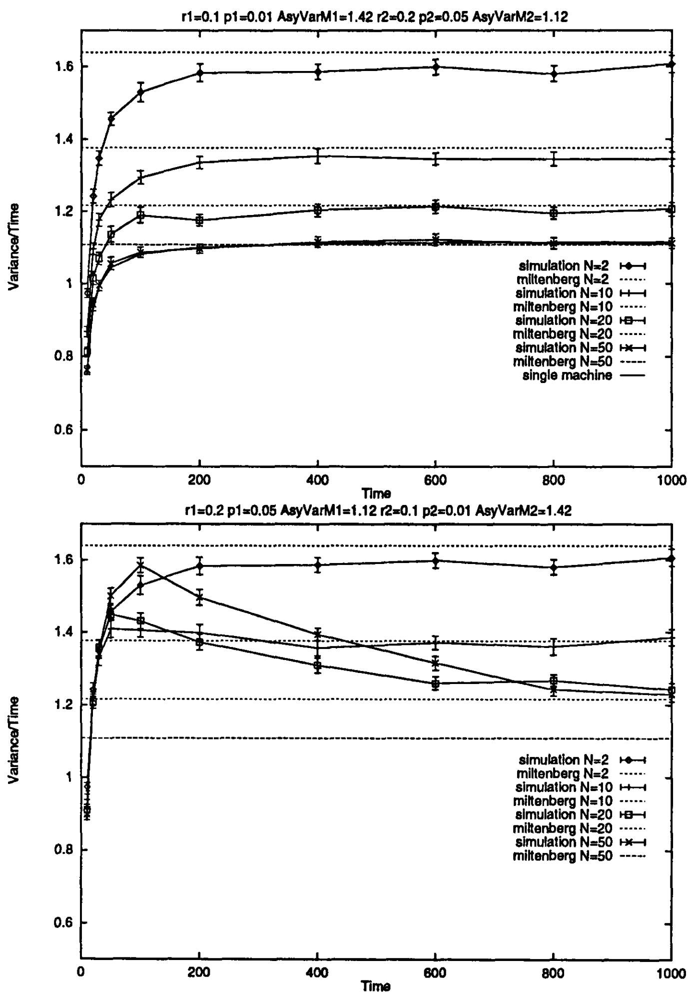

Figure 7-8: Different efficiencies: Evolution of the variance as $N \to \infty$

7.4 Graph of the Variance Rate as a Function of Time

In Section 3.4 we show that the plots of the variance rates as a function of time for a two-machine zero buffer system always presents the same shape. It is concave and asymptotically converges to Miltenburg's asymptotic variance rate.

In many of the graphs shown in this chapter we observe a different pattern for over t: the graph may present a maximum. In this cases there is a period of length t_{max} for which the variance rate is greater than for any other value. In a system whose graph presents this shape the range of t for which the variance rate is bigger than the asymptotic value becomes wider as N increases. Unfortunately, we cannot provide an general explanation of when and why this happens. This is an important and interesting question that requires further study.

7.5 Conclusions

In this chapter we focus on understanding the changes in the variance of a system as a function of time. We also observe the changes in the variance as the buffer size of the system increases. The tool used for this analysis is simulation. The conclusions derived are:

- There are three different time frames for the study of the variance:
  - Short term: The variance of the output can be explained without taking into account the correlation structure.
  - Medium term: The correlation structure affects the output pattern.
  - Long term: The variance of the output is linear with time.
- Effect of the buffer size in the variance: In the case of different machines, the bigger the buffer size, the more different the behavior of a system and its reverse

in the short and medium term.

  - $e_{M_2} < e_{M_1}$: As the buffer size increases, the variance of the system becomes closer to the variance of M_2. It also takes less time for the system to get to the long term behavior.

  - $e_{M_2} \geq e_{M_1}$: As the buffer size increases, the variance of the system takes longer to get to the asymptotic value. In the case of infinite buffers, the system never reaches steady state.

- Shape of the variance rate as a function of time: We cannot fully understand the cases where the function presents a maximum for a finite t.

Chapter 8

Conclusions

8.1 Results

We believe that this work is an important step towards a better understanding of the variability of the output of a manufacturing system. Very little work has been done in this area, and this is a most important issue in the day to day life of a production plant.

The specific results derived in this thesis can be summarized as follows:

- Flow interruption characterization of a two-machine finite buffer size system: the parameters that define the mean frequency and length of these interruptions were derived.

- Determination of the difference in the output pattern of a two-machine zero buffer line and a two-machine finite buffer line: a two-machine finite buffer system has the same interoutput distribution as an equivalent two-machine zero-buffer one. However, the variance of the systems is different. The reason appears to be that the correlation structure of the interdeparture times differ.

- Observations about the asymptotic variance rate of the output:
  - Machines with different efficiencies: The asymptotic variance rate of the output converges to the asymptotic variance rate of the less efficient machine as the buffer size increases.
  - Machines with the same efficiencies: The asymptotic variance rate of the output converges to a value that is smaller than the greatest asymptotic variance rate of the machines that form the system.
  - A system and its reverse: The asymptotic variance rate of both systems coincide.

Another set of results derived deal with the determination of the variance of the output for finite periods of time. This is an area which, as far as we know, has not been explored before, except in a limited way, by Gershwin (1993). The work done here includes:

- Aa analytical derivation of the variance of the output over $[0, t]$ of a single machine with two-failure modes for all t: this result closely approximates the variance of a two-machine zero-buffer line.

- Modeling of the variance of the output of a two-machine finite buffer system in the short term.

- Observations on the variance of the output as a function of time: determination of the existence of three different time frames in the variance of a system.
  - Short term: The number of changes between productive and non-productive periods is not big enough for the interdeparture correlations to affect the output pattern.
  - Medium term: Correlations significantly affect the output pattern.
  - Long term: The variance of the output of the system depends linearly on the time.

8.2 Limitations

We think that this contribution is useful in advancing towards the development of a more complete model of this phenomenon. However, there are still many questions that remain unanswered at this point:

- Asymptotic variance rate: We do not fully understand how the asymptotic variance rate changes as the buffer size increases, especially when the asymptotic variance rate presents a minimum for a finite buffer size. We do not have an explanation for the value of the asymptotic variance rate as the buffer size increases in the cases where the two machines have the same efficiency.
- Variance rate of the output as a function of time:
  - Boundaries of the time frames: we do not have a way of determining the length of the time frames.
  - Representation of the system for medium term: the shape of the curve is not fully understood.

8.3 Future Research

Further development of this work would lead to the derivation of a model that predicts the variance of the output of a two-machine line in the whole time range. A preceding step for the derivation of this model would probably be the analytical determination of the correlation structure in the system. Another very useful result would be to determine the lengths of the different time frames of the variance of the output of a system. This information should be part of a model that predicts the behavior of the system in the medium and long term.

In the longer term, the objective would be to extend the results obtained for the two-machine systems to longer lines. We envision this process as the development of a decomposition method in which the long line is analyzed as a sequence of combined two-machine systems. At this point the research would be in a mature enough stage to be used effectively in the design and control of a manufacturing system.

Appendix A

Derivation of Variance of a Single Machine with Two Failure Modes

In this appendix, we derive the equation for the variance of the production of a deterministic machine with two failure modes during a time interval $[0, t]$. This derivation follows closely Gershwin's derivation (1993) of the variance of the output of a deterministic single machine.

$\pi(n, \alpha, t)$ is defined as the probability of producing n parts in the interval $[0, t]$ being α the state of the machine at time t. This probability distribution is introduced in Section 3.5, and we use it for the derivation of a set of intermediate results.

A.1 Lemma 1

Let

$$U(t) = \sum_{n=0}^{t} \pi(n, 1, t) \tag{A.1}$$

be the average production in an interval of length t and

$$D_i(t) = \sum_{n=0}^{t} \pi(n, 0_i, t) \quad i = 1, 2 \tag{A.2}$$

be the average fraction of time that the machine is down in an interval of length t due to failures of type i. Then,

$$
\begin{aligned}
U(t) &= \frac{1}{1+\sum_{j=1,2}\frac{p_j}{r_j}} = C && \text{(A.3)}\\
D_i(t) &= \frac{\frac{p_i}{r_i}}{1+\sum_{j=1,2}\frac{p_j}{r_j}} = \frac{p_i}{r_i}C \quad i=1,2 && \text{(A.4)}
\end{aligned}
$$

Proof:

$D_i(t+1)$ and $U(t+1)$ can be expressed as a function of $D_i(t)$ and $U(t)$ by using equations (3.2), (3.3), (3.4).

$$
\begin{aligned}
: D_i(t+1) &= \sum_{n=0}^{t+1} \pi(n, 0_i, t+1)\\
&= \sum_{n=0}^{t+1}((1-r_i)\pi(n,0_i,t) + p_i\pi(n,1,t))\\
&= (1-r_i)D_i(t) + p_iU(t) \quad i=1,2 && \text{(A.5)}
\end{aligned}
$$

$$
\begin{aligned}
U(t+1) &= \sum_{n=0}^{t+1} \pi(n,1,t+1)\\
&= \sum_{n=0}^{t+1}(\sum_{i=1}^{2} r_i\pi(n-1,0_i,t) + (1-p_1+p_2)\pi(n,1,t))\\
&= (r_1D_1(t) - p_1U(t)) + (r_2D_2(t) - p_2U(t)) + U(t) && \text{(A.6)}
\end{aligned}
$$

From equations (A.5) and (A.6),

$$
\begin{aligned}
&r_2D_2(t+1) - p_2U(t+1) =\\
&\quad r_2((1-r_2)D_2(t) + p_2U(t)) - p_2(r_1D_1(t) - p_1U(t) + r_2D_2(t) - p_2U(t) + U(t)) =\\
&\quad (r_2D_2(t) - p_2U(t))(1-r_2-p_2) - p_2(r_1D_1(t) - p_1U(t))
\end{aligned}
$$

and similarly,

$$
r_1D_1(t+1) - p_1U(t+1) =
$$

$$(r_1 D_1(t) - p_1 U(t))(1 - r_1 - p_1) - p_1(r_2 D_2(t) - p_2 U(t)).$$

Since from equations (3.8) and (3.9),

$$r_1 D_1(0) - p_1 U(0) = 0$$

$$r_2 D_2(0) - p_2 U(0) = 0$$

we have

$$r_1 D_1(t) - p_1 U(t) = 0 \tag{A.7}$$

$$r_2 D_2(t) - p_2 U(t) = 0. \tag{A.8}$$

Substituting equations (A.7) and (A.8) in (A.6),

$$U(t+1) = U(t) = \cdots = U(0)$$

and plugging this into equations (A.7) and (A.8),

$$D_i(t+1) = \frac{p_i}{r_i} U(t+1) = \cdots = \frac{p_i}{r_i} U(0) = D_i(0) \quad i = 1, 2.$$

Finally, using the definition of $U(t)$ and $D_i(t)$, and equation (3.1)

$$U(0) = \pi(0,1,0) = p(1) = \frac{1}{1 + \sum_{j=1,2} \frac{p_j}{r_j}} = C$$

$$D_i(0) = \pi(0, 0_i, 0) = p(0_i) = \frac{\frac{p_i}{r_i}}{1 + \sum_{j=1,2} \frac{p_j}{r_j}} = \frac{p_i}{r_i} C \quad i = 1, 2$$

Therefore, Lemma 1 is proved.

A.2 Lemma 2

Let

$$E_\alpha(t) = \sum_{n=0}^{t} n\pi(n,\alpha,t) \tag{A.9}$$

be the average number of parts made in an interval of length t given that at time t the machine is in state i. Then,

$$\begin{aligned}
E_1(t) &= \frac{4r_1r_2C}{(p_1+p_2+r_1+r_2)^2-b^2)}t \\
&+\frac{C(b-p_1-p_2+r_2-r_1)(-b+p_1+p_2+r_2-r_1)}{b(p_1+p_2+r_1+r_2-b)^2}(1-\beta_1^t) \\
&+\frac{C(-b-p_1-p_2+r_2-r_1)(-b-p_1-p_2+r_2-r_1)}{b(p_1+p_2+r_1+r_2+b)^2}(1-\beta_2^t)
\end{aligned}$$

$$\begin{aligned}
E_{0_2}(t) &= \frac{4r_1p_2C}{((p_1+p_2+r_1+r_2)^2-b^2)}t \\
&-\frac{2p_2C(b-p_1-p_2-r_2+r_1)}{b(p_1+p_2+r_1+r_2-b)^2)}(1-\beta_1^t) \\
&-\frac{2p_2C(-b-p_1-p_2-r_2+r_1)}{b(p_1+p_2+r_1+r_2+b)^2)}(1-\beta_2^t)
\end{aligned}$$

$$\begin{aligned}
E_{0_1}(t) &= \frac{4p_1r_2C}{((p_1+p_2+r_1+r_2)^2-b^2)}t \\
&-\frac{C(4p_2r_2-(-b-p_1-p_2+r_2+r_1)(-b+p_1+p_2-r_2+r_1))}{2br_1(p_1+p_2+r_1+r_2-b)^2)}(1-\beta_1^t) \\
&-\frac{C(4p_2r_2-(b-p_1-p_2+r_2+r_1)(b+p_1+p_2-r_2+r_1))}{2br_1(p_1+p_2+r_1+r_2+b)^2)}(1-\beta_2^t)
\end{aligned}$$

where C is described in equation (A.3) and b, β_1*,* β_2 *are*

$$b^2 = (r_2-r_1)^2 + (p_1+p_2)^2 + 2(r_2-r_1)(p_2-p_1) \tag{A.10}$$

$$\beta_1 = 1 - \frac{p_1+p_2+r_2+r_1-b}{2} \tag{A.11}$$

$$\beta_2 = 1 - \frac{p_1+p_2+r_2+r_1+b}{2} \tag{A.12}$$

Proof:

Each $E_i(t+1)$ for $i = 1, 0_1, 0_2$ can be written as a function of $E_1(t)$, $E_{0_2}(t)$, $E_{0_1}(t)$ using the equations (3.2), (3.3), (3.4). This generates the following set of difference equations:

$$
\begin{aligned}
E_1(t+1) &= \sum_{n=0}^{t+1} n\pi(n,1,t+1) = \sum_{n=0}^{t}(n+1)\pi(n,1,t+1) \\
&= \sum_{n=0}^{t}(n+1)r_1\pi(n,0_1,t) + \sum_{n=0}^{t}(n+1)r_2\pi(n,0_2,t) \\
&\quad + \sum_{n=0}^{t}(n+1)(1-p_1-p_2)\pi(n,1,t) \\
&= r_1E_{0_1}(t) + r_2E_{0_2}(t) + (1-p_1-p_2)E_1(t) \\
&\quad +(r_2D_2(t) - p_2U(t)) + (r_1D_1(t) - p_1U(t)) + U(t) \\
&= r_1E_{0_1}(t) + r_2E_{0_2}(t) + (1-p_1-p_2)E_1(t) + C \\
E_{0_i}(t+1) &= \sum_{n=0}^{t+1} n\pi(n,0_i,t+1) = \sum_{n=0}^{t} n\pi(n,0_i,t+1) \\
&= \sum_{n=0}^{t} n(1-r_i)\pi(n,0_i,t) + \sum_{n=0}^{t} np_i\pi(n,1,t) \\
&= (1-r_i)E_{0_i} + p_iE_1(t) \quad i = 1,2.
\end{aligned}
$$

$$
\begin{bmatrix} E_{0_1}(t+1) \\ E_{0_2}(t+1) \\ E_1(t+1) \end{bmatrix} = \begin{bmatrix} 1-r_1 & 0 & p_1 \\ 0 & 1-r_2 & p_2 \\ r_1 & r_2 & 1-p_2-p_1 \end{bmatrix} \begin{bmatrix} E_{0_1}(t) \\ E_{0_2}(t) \\ E_1(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ C \end{bmatrix} \qquad \text{(A.13)}
$$

$$E(t+1) = AE(t) + u$$

The initial conditions are:

$$
E(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \text{(A.14)}
$$

The solution of this particular system has the form (Luenberger (1979)) :

$$E(t) = \left[\sum_{i=0}^{t-1} A^i \right] u$$

Defining Λ as the matrix of the eigenvalues of A and M as the matrix of the corresponding eigenvectors,

$$E(t) = M \left[\sum_{i=0}^{t-1} \Lambda^i \right] M^{-1} u \tag{A.15}$$

These matrices are:

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \beta_1 & 0 \\ 0 & 0 & \beta_2 \end{bmatrix}$$

$$M = \begin{bmatrix} \frac{p_1}{r_1} & \frac{4p_2r_2r_1-((b+r_1)^2-(p_1+p_2-r_2)^2)}{2r_1(b+r_1+p_1+p_2-r_2)} & \frac{4p_2r_2r_1-(-b+r_2)^2-(p_1+p_2-r_2)^2)}{2r_1(-b+r_2+p_1+p_2-r_2)} \\ \frac{p_2}{r_2} & \frac{-2p_2}{b+p_1+p_2+r_1-r_2} & \frac{-2p_2}{-b+p_1+p_2+r_1-r_2} \\ 1 & 1 & 1 \end{bmatrix}$$

Performing the operations described in equation (A.15), we get the expressions in Lemma 2.

As a consequence of this lemma we can obtain an expression for the average production rate in an interval of length t $\bar{n}(t)$. By definition,

$$\begin{aligned} \bar{n}(t) &= \sum_{n=0}^{t} n(\pi(n, 0_1, t) + \pi(n, 0_2, t) + \pi(n, 1, t) \\ &= E_{0_1}(t) + E_{0_2}(t) + E_1(t) \end{aligned} \tag{A.16}$$

Equation (A.16) simplifies to

$$\bar{n}(t) = Ct$$

A.3 Theorem 1

The variance of the number of parts produced during $[0,t]$ *is given by*

$$\begin{aligned}
\sigma^2(t) &= (C\frac{r_1^2p_2(2-p_2-r_2)+r_2^2p_1(2-p_1-r_1)-2p_1p_2r_1r_2}{(p_2r_1+r_2p_1+r_1r_2)^3})t \qquad (A.17)\\
&\quad -2C\frac{(2+b-p_1-p_2-r_1-r_2)((r_2-r_1)^2-(p_1+p_2-b)^2)}{b(-b+p_1+p_2+r_1+r_2)^3}(1-\beta_1^t)\\
&\quad -2C\frac{(-2+b-p_1-p_2-r_1-r_2)((r_2-r_1)^2-(p_1+p_2-b)^2)}{b(b+p_1+p_2+r_1+r_2)^3}(1-\beta_2^t)
\end{aligned}$$

where C, b, β_1 *and* β_2 *are defined in equations(A.3), (A.10), (A.11), (A.12).*
Proof:

The variance is given by

$$\sigma^2(t) = E(n(t)^2) - (\bar{n}(t))^2$$

According to equation (A.16) the second term is

$$\overline{n}(t)^2 = C^2t^2$$

Defining

$$S(t) = E((n(t)^2) = \sum_{n=0}^{t} n^2(\pi(n,1,t)+\pi(n,0_1,t)+\pi(n,0_2,t))$$

then,

$$\begin{aligned}
S(t+1) &= \sum_{n=0}^{t+1} n^2\pi(n,1,t+1) + \sum_{n=0}^{t+1} n^2\pi(n,0_1,t) + \sum_{n=0}^{t+1} n^2\pi(n,0_2,t+1)\\
&= \sum_{n=0}^{t} n^2((1-r_1)\pi(n,0_1,t)+p_1\pi(n,1,t))\\
&\quad +\sum_{n=0}^{t} n^2((1-r_2)\pi(n,0_2,t)+p_2\pi(n,1,t))\\
&\quad +\sum_{n=0}^{t} (n+1)^2(r_1\pi(n,0_1,t)+r_2\pi(n,0_2,t))+(1-p_1+p_2)\pi(n,0_2,t))
\end{aligned}$$

$$
\begin{aligned}
= & \sum_{n=0}^{t} n^2\pi(n,1,t) + \sum_{n=0}^{t} n^2\pi(n,0_1,t) + \sum_{n=0}^{t} n^2\pi(n,0_2,t) \\
& +2r_1\sum_{n=0}^{t} n\pi(n,0_1,t) + 2r_2\sum_{n=0}^{t} n\pi(n,0_2,t) + 2(1-p_1-p_2)\sum_{n=0}^{t} n\pi(n,1,t) \\
& +r_1\sum_{n=0}^{t} \pi(n,0_1,t) + r_2\sum_{n=0}^{t} \pi(n,0_2,t) + (1-p_1-p_2)\sum_{n=0}^{t} \pi(n,1,t)
\end{aligned}
$$

Using equation (A.7) and (A.8) together with the definitions in Lemma 1 and Lemma 2 to simplify the expression,

$$
S(t+1) = S(t) + 2r_1E_{0_1}(t) + 2r_2E_{0_2}(t) + 2(1-p_1-p_2)E_1(t) + U(t)
$$

This is a difference equation. Following the same pattern of derivation recursively we obtain,

$$
S(t) = S(0) + 2r_1\sum_{i=0}^{t-1} E_{0_1}(i) + 2r_2\sum_{i=0}^{t-1} E_{0_2}(i) + 2(1-p_1-p_2)\sum_{i=0}^{t-1} E_1(i) + \sum_{i=0}^{t-1} U(i). \quad \text{(A.18)}
$$

The initial condition is

$$
S(0) = 0
$$

The solution to equation (A.18) is

$$
\begin{aligned}
S(t) = & \left(C\frac{r_1^2p_2(2-p_2-r_2) + r_2^2p_1(2-p_1-r_1) - 2p_1p_2r_1r_2}{(p_2r_1+r_2p_1+r_1r_2)^3}\right)t \\
& -2C\frac{(2+b-p_1-p_2-r_1-r_2)((r_2-r_1)^2-(p_1+p_2-b)^2)}{b(-b+p_1+p_2+r_1+r_2)^3}(1-\beta_1^t) \\
& -2C\frac{(-2+b-p_1-p_2-r_1-r_2)((r_2-r_1)^2-(p_1+p_2-b)^2)}{b(b+p_1+p_2+r_1+r_2)^3}(1-\beta_2^t)
\end{aligned}
$$

and the theorem is proved.

Appendix B

Expected First Passage Time

In this appendix, we describe the result used to derive the expected first passage time between two states of a Markov chain. This technique is used in Section 4.3 to calculate the mean time it takes the system to go from one state to another. The result is applied to obtain the parameters of the machine equivalent to M_1 and the buffer in a two-machine line with finite buffer size. A complete description of the derivation of the first passage time can be found in Gallager (1995).

Definition *The expected first passage time* is the expected number of time units, when the system is in some initial state (k), before some other final state (j) is entered. The steady state probability distribution is defined by equation (2.1):

$$\mathbf{p} = \mathbf{p}P$$

where $\mathbf{p}$ is the state vector and P is the transition matrix.

Procedure Let w_{kj} be the expected number of steps to reach state j starting in state $k \neq j$ ($w_{jj} = 0$). w_{kj} includes the first step plus the number of steps from whatever state is entered next (which is 0 if state j is entered next). Therefore,

$$w_{kj} = 1 + \sum_{s \neq j} P_{ks} w_{sj},\ \forall\ k \neq j$$

Therefore, for a final state j, we obtain the number of steps needed to reach it from any other state $(w_{kj}, \forall k \neq j)$ by solving the system

$$\begin{bmatrix} w_{1j} \\ \dots \\ w_{j-1j} \\ w_{j+1j} \\ \dots \\ w_{sj} \end{bmatrix} = \begin{bmatrix} 1 \\ \dots \\ 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} + \begin{bmatrix} P_{11} & \dots & P_{1j-1} & P_{1j+1} & \dots & P_{1s} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_{j-11} & \dots & P_{j-1j-1} & P_{j-1j+1} & \dots & P_{j-1s} \\ P_{j+11} & \dots & P_{j+1j-1} & P_{j+1j+1} & \dots & P_{j+1s} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_{s1} & \dots & P_{sj-1} & P_{sj+1} & \dots & P_{ss} \end{bmatrix} \begin{bmatrix} w_1 \\ \dots \\ w_{j-1} \\ w_{j+1} \\ \dots \\ w_s \end{bmatrix}$$

or

$$[I - P'_j]w_j = u$$

where

$$\begin{aligned} w_j &= \text{vector of } w_{kj}, \forall k \neq j \\ & \quad w_{kj} \text{ is the expected number of steps to get from } k \text{ to } j \\ u &= \text{vector of 1s} \\ P'_j &= \text{transition matrix } P \text{ after eliminating row and column } j \\ &= \begin{bmatrix} P_{11} & \dots & P_{1j-1} & P_{1j+1} & \dots & P_{1s} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_{j-11} & \dots & P_{j-1j-1} & P_{j-1j+1} & \dots & P_{j-1s} \\ P_{j+11} & \dots & P_{j+1j-1} & P_{j+1j+1} & \dots & P_{j+1s} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_{s1} & \dots & P_{sj-1} & P_{sj+1} & \dots & P_{ss} \end{bmatrix} \end{aligned}$$

This system has a unique solution vector w_j. If we are just interested in the expected first passage time between states i and j, w_{ij} is the only component of the vector w_j we need to calculate.

Application In Section 4.3, we want to calculate the expected time to go from state $(0,0,1)$ to $(1,1,1)$ ($w_{(001)(111)}$) and the expected time to go from state $(1,1,1)$ to $(0,0,1)$ ($w_{(111)(001)}$).

$w_{(001)(111)}$: We need to calculate component $w_{(001)}$ of vector $w_{(111)}$. In this case,

$$[I - P'_{(111)}]w_{(111)} = u$$

where $P'_{(111)}$ is the transition matrix P of the system after eliminating row and column $(1,1,1)$.

$w_{(111)(001)}$: We need to calculate component $w_{(111)}$ of vector $w_{(001)}$. In this case,

$$[I - P'_{(001)}]w_{(001)} = u$$

where $P'_{(001)}$ is the transition matrix P of the system after eliminating row and column $(0,0,1)$.

Appendix C

Miltenburg's Asymptotic Variance Rate

In this appendix, Table. C.1 we present a sample of the results obtained for a wide range of cases. In all the examples tried

$$\Delta_{(M_1,M_2,N)} = \Delta_{(M_1,M_2,N)}$$

r_1	p_1	r_2	p_2	N	$\Delta_{(M_1,M_2,N)} = \Delta_{(M_1,M_2,N)}$
0.1	0.01	0.0738	0.00529	10	2.05171
0.1	0.01	0.0738	0.00529	50	1.48795
0.1	0.01	0.397	0.0207	20	1.1913
0.1	0.01	0.397	0.0207	100	1.41533
0.12	0.03	0.0309	0.00212	30	2.67672
0.12	0.03	0.0309	0.00212	50	2.36885
0.12	0.03	0.0936	0.00667	100	1.97053
0.12	0.03	0.0936	0.00661	200	1.97333
0.24	0.1	0.329	0.0239	10	0.995129
0.24	0.1	0.329	0.0239	20	1.0119
0.24	0.1	0.4	0.0404	30	1.01348
0.24	0.1	0.4	0.0404	50	1.01364
0.05	0.033	0.0454	0.00296	50	5.47464
0.05	0.033	0.0454	0.00296	100	5.52314
0.05	0.033	0.0205	0.937	10	5.35317
0.05	0.033	0.0205	0.937	20	5.49812
0.1	0.1	0.259	0.155	100	2.24477
0.1	0.1	0.259	0.155	400	2.25

Table C.1: Asymptotic variance rate for a system and its reverse.

Bibliography

[1] G. Miltenburg (1987), "Variance of the Number of Units Produced on a Transfer Line with Buffer Inventories during a Period of Length T," *Naval Research Logistics* Vol.34, pp. 811-822.

[2] J. Ou and S. Gershwin (1989), "The Variance of the Lead Time of a Two-Machine Transfer Line with a Finite Buffer," Laboratoty for Manufacturing and Productivity, M.I.T. LMP-89-028.

[3] K. Hendricks (1992), "The Output Process of Serial Production Lines of Exponential Machines with Finite Buffers," *Operations Research* Vol. 40, No. 6. November-December 1992, pp. 1139-1147.

[4] J. Buzacott (1967), "Automatic Transfer Lines with Buffer Stocks," *International Journal of Production Research*, Vol.5, No. 3, pp. 183-200.

[5] J. Buzacott and J. Shanthikumar (1993), *"Stochastic Models of Manufacturing Systems,"* Prentice Hall, Englewood Cliffs, New Jersey 07632, pp. 244-263.

[6] S. Gershwin (1993), "Variance of the Output of a Tandem Production System," *Queuing Networks with Finite Capacity. Proceedings of the Second International Workshop held in Research Triangle Park, North Carolina, May 28-29, 1992.* Edited by R. Onvural and I. Akyldiz, Elsevier, 1993.

[7] S. Gershwin (1994), *"Manufacturing Systems Analysis,"* Prentice Hall, Englewood Cliffs, New Jersey 07632, pp. 69-93.

[8] W. Wei (1993), *"Time Series Analysis: Univariate and Multivariate Methods,"* Addison-Wesley Publishing Company, Inc, pp. 349-361.

[9] J. Kemeny and J. Snell (1976), *"Finite Markov Chains"*, Van Nostrand, New York, pp. 69-90.

[10] A. Drake (1967), *"Fundamentals of Applied Probability Theory,"* McGraw-Hill Classic Textbook Reissue Series Publishing Company, pp 206-207.

[11] M. Ammar (1980), "Modelling and Analysis of Unreliable Manufacturing Assembly Networks with Finite Storages," PhD thesis, Laboratory for Information and Decision Systems, M.I.T. LIDS-TH-1004, pp. 49-76.

[12] I. Schick and S. Gershwin (1978), "Modelling and Analysis of Unreliable Transfer Lines with Finite Interstage Buffers," Electronic Systems Laboratory, M.I.T. ESL-FR-834-6, pp. 174-191.

[13] R. Gallager (1995), *"Discrete Stochastic Processes,"* Kluwer Academic Publishers Inc.

[14] D. Luenberger (1979), *"Introduction to Dynamic Systems - Theory, Models and Applications,"* Wiley, pp. 26-38.