OR Spectrum manuscript No.
(will be inserted by the editor)

Integrated Quality and Quantity Modeling of a Production Line

Jongyoon Kim, Stanley B. Gershwin

Department of Mechanical Engineering, Massachusetts Institute of Technology
Cambridge, Massachusetts 02139-4307

Received: December, 2003 / Revised version: May, 2004

Abstract During the past three decades, the success of the Toyota Production System has spurred much research in manufacturing systems engineering. Productivity and quality have been extensively studied, but there is little research in their intersection. The goal of this paper is to analyze how production system design, quality, and productivity are inter-related in small production systems. We develop a new Markov process model for machines with both quality and operational failures, and we identify important differences between types of quality failures. We also develop models for two-machine systems, with infinite buffers, buffers of size zero, and finite buffers. We calculate total production rate, effective production rate (ie, the production rate of good parts), and yield. Numerical studies using these models show that when the first machine has quality failures and the inspection occurs only at the second machine, there are cases in which the effective production rate *increases* as buffer sizes increase, and there are cases in which the effective production rate *decreases* for larger buffers. We propose extensions to larger systems.

Key words Quality, Productivity, Manufacturing System Design

1 Introduction

1.1 Motivation

During the past three decades, the success of the Toyota Production System has spurred much research in manufacturing systems design. Numerous research papers have tried to explore the relationship between production

---

Send offprint requests to: `gershwin@mit.edu`

system design and productivity, so that they can show ways to design factories to produce more products on time with less resources (such as people, material, and space). On the other hand, topics in quality research have captured the attention of practitioners and researchers since the early 1980s. The recent popularity of Statistical Quality Control (SQC), Total Quality Management (TQM), and Six Sigma have demonstrated the importance of quality.

These two fields, productivity and quality, have been extensively studied and reported separately both in the manufacturing systems research literature and the practitioner literature, but there is little research in their intersection. The need for such work was recently described by authors from the GM Corporation based on their experience [13]. All manufacturers must satisfy these two requirements (high productivity and high quality) at the same time to maintain their competitiveness.

Toyota Production System advocates admonish factory designers to combine inspections with operations. In the Toyota Production System, the machines are designed to detect abnormalities and to stop automatically whenever they occur. Also, operators are equipped with means of stopping the production flow whenever they note anything suspicious. (They call this practice *jidoka*.) Toyota Production System advocates argue that mechanical and human jidoka prevent the waste that would result from producing a series of defective items. Therefore jidoka is a means to improve quality and increase productivity at the same time [23], [24]. But this statement is arguable: quality failures are often those in which the quality of each part is independent of the others. This is the case when the defect takes place due to common (or chance or random) causes of variations [16]. In this case, there is no reason to stop a machine that has made a bad part because there is no reason to believe that stopping it will reduce the number of bad parts in the future. In this case, therefore, stopping the operation does not influence quality but it does reduce productivity. On the other hand, when quality failures are those in which once a bad part is produced, all subsequent parts will be bad until the machine is repaired (due to special or assignable or systematic causes of variations) [16], catching bad parts and stopping the machine as soon as possible is the best way to maintain high quality and productivity.

Non-stock or *lean* production is another popular buzzword in manufacturing systems engineering. Some lean manufacturing professionals advocate reducing inventory on the factory floor since the reduction of work-in-process (WIP) reveals the problems in the production lines [3]. Thus, it can help improve product quality. It is true in some sense: less inventory reduces the time between making a defect and identifying the defect. But it is also true that productivity would diminish significantly without stock [5]. Since there is a tradeoff, there must be optimal stock levels that are specific to each manufacturing environment. In fact, Toyota recently changed their view on inventory and are trying to re-adjust their inventory levels [9].

What is missing in discussions of factory design, quality, and productivity is a quantitative model to show how they are inter-related. Most of the arguments about this are based on anecdotal evidence or qualitative reasoning that lack a sound scientific quantitative foundation. The research described here tries to establish such a foundation to investigate how production system design and operation influence productivity and product quality by developing conceptual and computational models of two-machine-one-buffer systems and performing numerical experiments.

1.2 Background

1.2.1 Quality models There are two extreme kinds of quality failures based on the characteristics of variations that cause the failures. In the quality literature, these variations are called *common* (or chance or random) cause variations and *assignable* (or special or unusual) cause variations [18].

Figure 1 shows the types of quality failures and variations. Common cause failures are those in which the quality of each part is independent of the others. Such failures occur often when an operation is sensitive to external perturbations like defects in raw material or when the operation uses a new technology that is difficult to control. This is inherent in the design of the process. Such failures can be represented by independent Bernoulli random variables, in which a binary random variable, which indicates whether or not the part is good, is chosen each time a part is operated on. A good part is produced with probability π, and a bad part is produced with probability $1 - \pi$. The occurrence of a bad part implies nothing about the quality of future parts, so no permanent changes can have occurred in the machine. For the sake of clarity, we call this a *Bernoulli-type quality failure.* Most of the quantitative literature on inspection allocation assumes this kind of quality failure [21]. In this case, if bad parts are destined to be scrapped, it is useful to catch them as soon as possible because the longer before they are scrapped, the more they consume the capacity of downstream machines. However, there is no reason to stop a machine that has produced a bad part due to this kind of failure.

The quality failures due to assignable cause variations are those in which a quality failure only happens after a change occurs in the machine. In that case, it is very likely that once a bad part is produced, all subsequent parts will be bad until the machine is repaired. Here, there is much more incentive to catch defective parts and stop the machine quickly. In addition to minimizing the waste of downstream capacity, this strategy minimizes the further production of defective parts. For this kind of quality failure, there is no inherent measure of yield because the fractions of parts that are good and bad depend on how soon bad parts are detected and how quickly the machine is stopped for repair. In this paper, we call this a *persistent-type quality failure.* Most quantitative studies in Statistical Quality Control (SQC) are dedicated to finding efficient inspection policies (sampling interval, sample size, control limits, and others) to detect this type of quality failure [26].

In reality, failures are mixtures of Bernoulli-type quality failures and persistent-type quality failures. It can be argue that the quality strategy of the Toyota Production System [17], in which machines are stopped as soon as a bad part is detected, is implicitly based on the assumption of the persistent-type quality failure. In this paper, we focus on persistent failures.

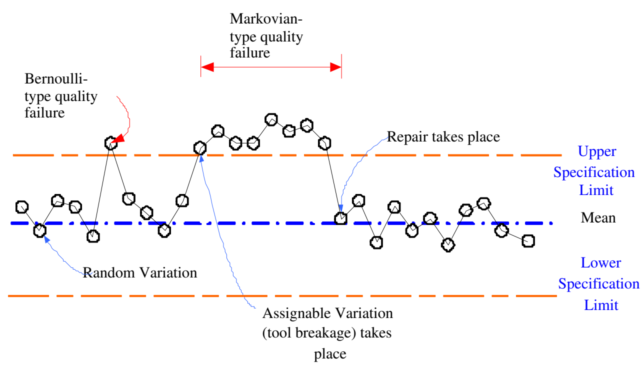

Fig. 1 Types of Quality Failures

1.2.2 System yield *System yield* is defined here as the fraction of input to a system that is transformed into output of acceptable quality. This is an important metric because customers observe the quality of products only after all the manufacturing processes are done and the products are shipped. The system yield is a complex function of how the factory is designed and operated, as well as of the characteristics of the machines. Some of influencing factors include individual operation yields, inspection strategies, operation policies, buffer sizes, and other factors. Comprehensive approaches are needed to manage system yield effectively. This research aims to develop mathematical models to show how the system yield is influenced by these factors.

1.2.3 Quality improvement policy System yield is a complex function of various factors such as inspection, individual operation yields, buffer size, operation policies, and others. There are many ways to affect the system yield. Inspection policy has received the most attention in the literature. Research on inspection policies can be divided into optimizing inspection parameters at a single station and the inspection station allocation problem. The former issue has been investigated extensively in the SQC literature [26]. Here, optimal SQC parameters such as control limits, sampling size, and frequency are sought for an optimal balance between the inspection cost and the cost of quality. The latter research looks for the optimal distributions of inspection stations along production lines [21].

Improving individual operation yield is another important way to increase the system yield. Many studies in this field try to stabilize the process either by finding root causes of variation and eliminating them or by making the process insensitive to external noise. The former topic has numerous qualitative research papers in the fields of Total Quality Management (TQM) [2] and Six Sigma [19]. Quantitative research is more oriented toward the latter topic. Robust engineering [20] is an area that has gained substantial attention.

It has been argued that inventory reduction is an effective means to improve system yield. Many lean manufacturing specialists have asserted that less inventory on the factory floor reveals problems in the manufacturing lines more quickly and helps quality improvement activities [1], [17].

There also have been investigations to explain the relationship between plant layout design and quality [7]. They argue that U-shaped lines are better than straight lines for producing higher quality products since there are more points of contact between operators. There is also less material movement, and there are other reasons.

There are many ways to improve system yield, but using only a single method will give limited gains. The effectiveness of each method is greatly dependent on the details of the factory. Thus, there is need to determine which method or which combination of methods is most effective in each case. The quantitative tools that will be developed from this research can help fulfill this need.

1.3 Outline

In Section 2 we introduce the structure of the modeling techniques used in this paper. We present modeling, solution techniques, and validation of the 2-machine-1-finite buffer case in Section 3. Discussions on the behavior of a production line based on numerical experiments are provided in Section 5. A future research plan is shown in Section 6. Parameters of many of the systems studied numerically here, and details of the analytical solution of the two-machine line, can be found in the appendices.

2 Mathematical Models

2.1 Single machine model

There are many possible ways to characterize a machine for the purpose of simultaneously studying quality and quantity issues. Here, we model a machine as a discrete state, continuous time Markov process. Material is assumed continuous, and μ_i is the speed at which Machine i processes material while it is operating and not constrained by the other machine or the buffer. It is a constant, in that μ_i does not depend on the repair state of the other machine or the buffer level.

Figure 2 shows the proposed state transitions of a single machine with persistent-type quality failures. In the model, the machine has three states:

- ⋆ State 1: The machine is operating and producing good parts.
- ⋆ State -1: The machine is operating and producing bad parts, but the operator does not know this yet.
- ⋆ State 0: The machine is not operating.

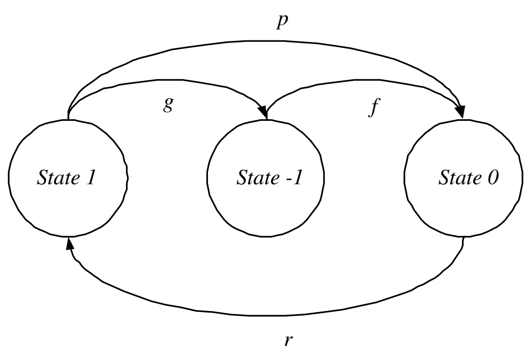

Fig. 2 States of a Machine

The machine therefore has two different failure modes (i.e. transition to failure states from state 1):

- ⋆ *Operational failure*: transition from state 1 to state 0. The machine stops producing parts due to failures like motor burnout.
- ⋆ *Quality failure*: transition from state 1 to state -1. The machine stops producing good parts (and starts producing bad parts) due to a failure like a sudden tool damage.

When a machine is in state 1, it can fail due to a non-quality related event. It goes to state 0 with transition probability rate p. After that an operator fixes it, and the machine goes back to state 1 with transition rate r. Sometimes, due to an assignable cause, the machine begins to produce bad parts, so there is a transition from state 1 to state -1 with a probability rate g. Here g is the reciprocal of the *Mean Time to Quality Failure (MTQF)*. A more stable operation leads to a larger MTQF and a smaller g.

The machine, when it is in state -1, can be stopped for two reasons: it may experience the same kind of operational failure as it does when it is in state 1; and the operator may stop it for repair when he learns that it is producing bad parts. The transition from state -1 to state 0 occurs at probability rate $f = p + h$ where h is the reciprocal of the *Mean Time To Detect (MTTD)*. A more reliable inspection leads to a shorter MTTD and a larger f. (The detection can take place elsewhere, for example at a remote inspection station.) Note that this implies that $f > p$. Here, for simplicity, we assume that whenever a machine is repaired, it goes back to state 1. All the indicated transitions are assumed to follow exponential distributions.

Single Machine Analysis To determine the production rate of a single machine, we first determine the steady-state probability distribution. This is calculated based on the probability balance principle: the probability of leaving a state is the same as the probability of entering that state. We have

$$(g+p)P(1) = rP(0) \tag{1}$$

$$fP(-1) = gP(1) \tag{2}$$

$$rP(0) = pP(1) + fP(-1) \tag{3}$$

The probabilities must also satisfy the normalization equation:

$$P(0) + P(1) + P(-1) = 1 \tag{4}$$

The solution of (1)–(4) is

$$P(1) = \frac{1}{1 + (p+g)/r + g/f} \tag{5}$$

$$P(0) = \frac{(p+g)/r}{1 + (p+g)/r + g/f} \tag{6}$$

$$P(-1) = \frac{g/f}{1 + (p+g)/r + g/f} \tag{7}$$

The *total production rate*, including good and bad parts, is

$$P_T = \mu(P(1) + P(-1)) = \mu \frac{1 + g/f}{1 + (p+g)/r + g/f} \tag{8}$$

The *effective production rate*, the production rate of good parts only, is

$$P_E = \mu P(1) = \mu \frac{1}{1 + (p+g)/r + g/f} \tag{9}$$

The *yield* is

$$\frac{P_E}{P_E + P_T} = \frac{P(1)}{P(1) + P(-1)} = \frac{f}{f+g} \tag{10}$$

2.2 2-Machine-1-Buffer continuous model

A flow (or transfer) line is a manufacturing system with a very special structure. It is a linear network of service stations or machines (M_1, M_2, ..., M_k) separated by buffer storages (B_1, B_2, ..., B_{k-1}). Material flows from outside the system to M_1, then to B_1, then to M_2, and so forth until it reaches M_k, after which it leaves. Figure 3 depicts a flow line. The rectangles represent machines and the circles represent buffers.

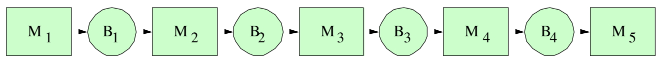

Fig. 3 Five-Machine Flow Line

2-machine-1-buffer (2M1B) models should be studied first. Then a decomposition technique, that divides a long transfer line into multiple 2-machine-1-buffer models, could be developed. (See [14].) Among the various modeling techniques for the 2M1B case, including deterministic, exponential, and continuous models, the continuous material line model is used for this research because it can handle deterministic but different operation times at each operation. This is an extension of the continuous material serial line modeling of [10] by adding another machine failure state. Figure 4 shows the 2M1B continuous model where the machines, buffer and discrete parts are represented as valves, a tank, and a continuous fluid.

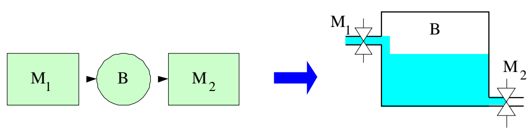

Fig. 4 Two-Machine-One-Buffer Continuous Model

We assume that an inexhaustible supply of workpieces is available upstream of the first machine in the line, and an unlimited storage area is present downstream of the last machine. Thus, the first machine is never starved, and the last machine is never blocked. Also, failures are assumed to be operation dependent (ODF).

Finally, we assume that each machine works on a different feature. For example, the two machines may be making two different holes. We do not consider cases where the both machines work on the same hole, in which the first machine does a roughing operation and the second does a finishing operation. This allows us to assume that the failures of the two machines are independent.

2.3 Infinite buffer case

An infinite buffer case is a special 2M1B line in which the size of the buffer (B) is infinite. This is an extreme case in which the first machine (M_1) never suffers from blockage. To derive expressions for the total production rate and the effective production rate, we observe that when there is infinite buffer capacity between two machines (M_1, M_2), the total production rate of the 2M1B system is a minimum of the total production rates of M_1 and M_2. The total production rate of machine i is given by (8), so the total production rate of the 2M1B system is

$$P_T^\infty = \min\left[\frac{\mu_1(1+g_1/f_1)}{1+(p_1+g_1)/r_1+g_1/f_1}, \frac{\mu_2(1+g_2/f_2)}{1+(p_2+g_2)/r_2+g_2/f_2}\right] \tag{11}$$

The probability that machine M_i does not add non-conformities is

$$Y_i = \frac{P_i(1)}{P_i(1)+P_i(-1)} = \frac{f_i}{f_i+g_i} \tag{12}$$

Since there is no scrap and rework in the system, the system yield becomes

$$\frac{f_1 f_2}{(f_1+g_1)(f_2+g_2)} \tag{13}$$

As a result, the effective production rate is

$$P_E^\infty = \frac{f_1 f_2}{(f_1+g_1)(f_2+g_2)} P_T^\infty \tag{14}$$

The effective production rate evaluated from (14) has been compared with a discrete-event, discrete-part simulation. Table 1 shows good agreement. The parameters for these cases are shown in Appendix B.

Table 1 Validation of Infinite Buffer Case

Case #	$P_E^\infty(Analytic)$	$P_E^\infty(Simulation)$	%Difference
1	0.762	0.761	0.17
2	0.708	0.708	0.00
3	0.657	0.657	0.00
4	0.577	0.580	-0.50
5	0.527	0.530	-0.42
6	0.745	0.745	0.01
7	0.762	0.760	0.30
8	1.524	1.522	0.14
9	0.762	0.762	0.00
10	1.524	1.526	-0.13

As indicated in Section 2.1, the detection of quality failures due to machine M_1 need not occur at that machine. For example, the inspection of the feature that M_1 works on could take place at an inspection station at M_2, and this inspection could trigger a repair of M_1. (We call this *quality information feedback*. See Section 4.) In that case, the MTTD of M_1 (and therefore f_1) will be a function of the amount of material in the buffer. We return to this important case in Section 4.

2.4 Zero buffer case

The zero buffer case is one in which there is no buffer space between the machines. This is the other extreme case where blockage and starvation take place most frequently.

In the zero-buffer case in which machines have different operation times, whenever one of the machines stops, the other one is also stopped. In addition, when both of them are working, the production rate is $\min[\mu_1, \mu_2]$. Consider a long time interval of length T during which M_1 fails m_1 times and M_2 fails m_2 times. If we assume that the average time to repair M_1 is $1/r_1$ and the average time to repair M_2 is $1/r_2$, then the total system down time will be close to $D = \frac{m_1}{r_1} + \frac{m_2}{r_2}$. Consequently, the total up time will be approximately

$$U = T - D = T - (\frac{m_1}{r_1} + \frac{m_2}{r_2}) \tag{15}$$

Since we assume operation-dependent failures, the rates of failure are reduced for the faster machine. Therefore,

$$p_i^b = p_i \frac{\min(\mu_1, \mu_2)}{\mu_i}, \; g_i^b = g_i \frac{\min(\mu_1, \mu_2)}{\mu_i}, \text{ and } f_i^b = f_i \frac{\min(\mu_1, \mu_2)}{\mu_i} \tag{16}$$

The reduction of p_i is explained in detail in [10]. The reductions of g_i and f_i are done for the same reasons.

Table 2 lists the possible working states α_1 and α_2 of M_1 and M_2. The third column is the probability of finding the system in the indicated state. The fourth and fifth columns indicate the expected number of transitions to down states during the time interval from each of the states in column 1.

Table 2 Zero-Buffer States, Probabilities, and Expected Numbers of Events

α_1	α_2	Probability $\pi(\alpha_1, \alpha_2)$	$Em_1(\alpha_1, \alpha_2)$	$Em_2(\alpha_1, \alpha_2)$
1	1	$\frac{f_1^b}{f_1^b+g_1^b} \frac{f_2^b}{f_2^b+g_2^b}$	$p_1^b U \pi(1,1)$	$p_2^b U \pi(1,1)$
1	-1	$\frac{f_1^b}{f_1^b+g_1^b} \frac{g_2^b}{f_2^b+g_2^b}$	$p_1^b U \pi(1,-1)$	$f_2^b U \pi(1,-1)$
-1	1	$\frac{g_1^b}{f_1^b+g_1^b} \frac{f_2^b}{f_2^b+g_2^b}$	$f_1^b U \pi(-1,1)$	$p_2^b U \pi(-1,1)$
-1	-1	$\frac{g_1^b}{f_1^b+g_1^b} \frac{g_2^b}{f_2^b+g_2^b}$	$f_1^b U \pi(-1,-1)$	$f_2^b U \pi(-1,-1)$

From Table 2, the expectations of m_1 and m_2 are

$$\begin{aligned} Em_1 &= \sum_{\alpha_1=-1}^{1} \sum_{\alpha_2=-1}^{1} Em_1(\alpha_1, \alpha_2) = \frac{Uf_1^b(p_1^b+g_1^b)}{f_1^b+g_1^b} \\ Em_2 &= \sum_{\alpha_1=-1}^{1} \sum_{\alpha_2=-1}^{1} Em_2(\alpha_1, \alpha_2) = \frac{Uf_2^b(p_2^b+g_2^b)}{f_2^b+g_2^b} \end{aligned} \quad (17)$$

By plugging them into equation (15), we find total production rate:

$$P_T^0 = \frac{\min[\mu_1, \mu_2]}{1 + \frac{f_1^b(p_1^b+g_1^b)}{r_1(f_1^b+g_1^b)} + \frac{f_2^b(p_2^b+g_2^b)}{r_2(f_2^b+g_2^b)}} \quad (18)$$

The effective production rate is

$$P_E^0 = \frac{f_1^b f_2^b}{(f_1^b + g_1^b)(f_2^b + g_2^b)} P_T^0 \quad (19)$$

The comparison with simulation is shown in in Table 3. The parameters of the cases are shown in Appendix B.

Table 3 Zero Buffer Case

Case #	$P_E^0(Analytic)$	$P_E^0(Simulation)$	%Difference
1	0.657	0.662	-0.73
2	0.620	0.627	-1.15
3	0.614	0.621	-1.03
4	0.529	0.534	-0.99
5	0.480	0.484	-0.77
6	0.647	0.651	-0.57
7	0.706	0.712	-0.91
8	1.377	1.406	-2.10
9	0.706	0.711	-0.77
10	1.377	1.380	-0.22

3 2-Machine-1-Finite-Buffer Line

The two-machine line is the simplest non-trivial case of a production line. In the existing literature on the performance evaluation of systems in which quality is not considered, two-machine lines are used in decomposition approximations of longer lines. (See [10].)

We define the model here and show the solution technique in Appendix A.

3.1 State definition

The state of the 2M1B line is defined as (x, α_1, α_2) where

- ⋆ x: the total amount of material in buffer B, $0 \le x \le N$,
- ⋆ α_1: the state of M_1. ($\alpha_1 = -1, 0$, or 1),
- ⋆ α_2: the state of M_2. ($\alpha_2 = -1, 0$, or 1)

The parameters of machine M_i are $\mu_i, r_i, p_i, f_i, g_i$ and the buffer size is N.

3.2 Model development

3.2.1 Internal transition equations When buffer B is neither empty nor full, its level can rise or fall depending on the states of adjacent machines. Since it can change only a small amount during a short time interval, it is natural to use differential equations to describe its behavior. The probability of finding both machines at state 1 with a total storage level between x and $x + \delta x$ at time $t + \delta t$ is given by $f(x, 1, 1)\delta t$, where

$$\begin{aligned} f(x,1,1,t+\delta t) &= \{1-(p_1+g_1+p_2+g_2)\delta t\} f(x+(\mu_2-\mu_1)\delta t,1,1) \\ &+ r_2 \delta t f(x-\mu_1\delta t,1,0) + r_1\delta t f(x+\mu_2\delta t,0,1) + o(\delta t) \end{aligned} \tag{20}$$

This is because if both machines are at state 1 at time t and the storage level is between x and $x + (\mu_2 - \mu_1)\delta t$, then there should be no failures before $t + \delta t$ to get $f(x, 1, 1)\delta t$. The probability of not having any failures between t and $t + \delta t$ is

$$\{1-(p_1+g_1)\delta t\}\{1-(p_2+g_2)\delta t\} \simeq \{1-(p_1+g_1+p_2+g_2)\delta t\} \tag{21}$$

Probabilities for transitions from the states $(x - \mu_1\delta t, 1, 0)$ and $(x + \mu_2\delta t, 0, 1)$ to $(x, 1, 1)$ can be found similarly. After linearizing and letting $\delta t \to 0$, this equation becomes

$$\begin{aligned} \frac{\partial f(x,1,1)}{\partial t} = &\ (\mu_2-\mu_1)\frac{\partial f(x,1,1)}{\partial x} - (p_1+g_1+p_2+g_2) f(x,1,1) + r_2 f(x,1,0) \\ &+ r_1 f(x,0,1) \end{aligned} \tag{22}$$

In steady state $\frac{\partial f}{\partial t} = 0$. Then, we have

$$(\mu_2-\mu_1)\frac{df(x,1,1)}{dx} - (p_1+g_1+p_2+g_2) f(x,1,1) + r_2 f(x,1,0) + r_1 f(x,0,1) = 0 \tag{23}$$

In the same way, the eight other internal transition equations for the probability density function are

$$p_2 f(x,1,1) - \mu_1 \frac{df(x,1,0)}{dx} - (p_1+g_1+r_2) f(x,1,0) + f_2 f(x,1,-1) + r_1 f(x,0,0) = 0 \tag{24}$$

$$g_2 f(x,1,1) + (\mu_2 - \mu_1)\frac{df(x,1,-1)}{dx} - (p_1 + g + f_2)f(x,1,-1) + r_1 f(x,0,-1) = 0 \quad (25)$$

$$p_1 f(x,1,1) + \mu_2 \frac{df(x,0,1)}{dx} - (r_1 + p_2 + g_2)f(x,0,1) + r_2 f(x,0,0) + f_1 f(x,-1,1) = 0 \quad (26)$$

$$p_1 f(x,1,0) + p_2 f(x,0,1) - (r_1 + r_2)f(x,0,0) + f_2 f(x,0,-1) + f_1 f(x,-1,0) = 0 \quad (27)$$

$$p_1 f(x,1,-1) + g_2 f(x,0,1) - (r_1 + f_2)f(x,0,-1) + \mu_2 \frac{df(x,0,-1)}{dx} + f_1 f(x,-1,-1) = 0 \quad (28)$$

$$g_1 f(x,1,1) - (p_2 + g_2 + f_1)f(x,-1,1) + (\mu_2 - \mu_1)\frac{df(x,-1,1)}{dx} + r_2 f(x,-1,0) = 0 \quad (29)$$

$$g_1 f(x,1,0) - \mu_1 \frac{df(x,-1,0)}{dx} - (r_2 + f_1)f(x,-1,0) + p_2 f(x,-1,1) + f_2 f(x,-1,-1) = 0 \quad (30)$$

$$g_1 f(x,1,-1) + g_2 f(x,-1,1) + (\mu_2 - \mu_1)\frac{df(x,-1,-1)}{dx} - (f_1 + f_2)f(x,-1,-1) = 0 \quad (31)$$

3.2.2 Boundary transition equations While the internal behavior of the system can be described by probability density functions, there is a nonzero probability of finding the system in certain boundary states. For example, if $\mu_1 < \mu_2$ and both machines are in state 1, the level of storage tends to decrease. If both machines remain operational for enough time, the storage will become empty ($x = 0$). Once the system reaches state $(0, 1, 1)$, it will remain there until a machine fails. There are 18 probability masses for boundary states ($P(N, \alpha_1, \alpha_2)$ and $P(0, \alpha_1, \alpha_2)$ where $\alpha_1 = -1$, 0 or 1, and $\alpha_2 = -1$, 0 or 1) and 22 boundary equations for the $\mu_1 = \mu_2$ case.

To arrive at state $(0, 1, 1)$ at time $t + \delta t$ when $\mu_1 = \mu_2$, the system may have been in one of two states at time t. It could have been in state $(0, 1, 1)$ without any of operational failures and quality failures for both of machines. It could have been in state $(0, 0, 1)$ with a repair of the first machine. (The second machine could not have failed since it was starved). If the second order terms are ignored,

$$P(0,1,1,t+\delta t) = \{1 - (p_1 + g_1 + p_2^b + g_2^b)\delta t\}P(0,1,1) + r_1 P(0,0,1) \quad (32)$$

After the usual analysis, (32) becomes

$$\frac{\partial P(0,1,1)}{\partial t} = (p_1 + g_1 + p_2^b + g_2^b)P(0,1,1) + r_1 P(0,0,1) \tag{33}$$

In steady state

$$-(p_1 + g_1 + p_2^b + g_2^b)P(0,1,1) + r_1 P(0,0,1) = 0 \tag{34}$$

There are 21 other boundary equations derived similarly for $\mu_1 = \mu_2$ [14]:

$$P(0,1,0) = 0 \tag{35}$$

$$g_2^b P(0,1,1) - (p_1 + g_1 + f_2^b)P(0,1,-1) + r_1 P(0,0,-1) = 0 \tag{36}$$

$$p_1 P(0,1,1) - r_1 P(0,0,1) + \mu_2 f(0,0,1) + f_1 P(0,-1,1) + r_2 P(0,0,0) = 0 \tag{37}$$

$$-(r_1 + r_2)P(0,0,0) = 0 \tag{38}$$

$$p_1 P(0,1,-1) - r_1 P(0,0,-1) + \mu_2 f(0,0,-1) + f_1 P(0,-1,-1) = 0 \tag{39}$$

$$g_1 P(0,1,1) - (f_1 + p_2^b + g_2^b)P(0,-1,1) = 0 \tag{40}$$

$$P(0,-1,0) = 0 \tag{41}$$

$$g_1 P(0,1,-1) + g_2^b P(0,-1,1) - (f_1 + f_2^b)P(0,-1,-1) = 0 \tag{42}$$

$$-(p_1^b + g_1^b + p_2 + g_2)P(N,1,1) + r_2 P(N,1,0) = 0 \tag{43}$$

$$p_2 P(N,1,1) - r_2 P(N,1,0) + \mu_1 f(N,1,0) + f_2 P(N,1,-1) + r_1 P(N,0,0) = 0 \tag{44}$$

$$g_2 P(N,1,1) - (p_1^b + g_1^b + f_2)P(N,1,-1) = 0 \tag{45}$$

$$P(N,0,1) = 0 \tag{46}$$

$$-(r_1 + r_2)P(N,0,0) = 0 \tag{47}$$

$$P(N,0,-1) = 0 \tag{48}$$

$$g_1^b P(N,1,1) - (f_1^b + g_2 + p_2)P(N,-1,1) + r_2 P(N,-1,0) = 0 \tag{49}$$

$$-r_2P(N,-1,0)+\mu_1f(N,-1,0)+f_2P(N,-1,-1)+p_2P(N,-1,1)=0 \quad (50)$$

$$g_1^bP(N,1,-1)+g_2P(N,-1,1)-(f_1^b+f_2)P(N,-1,-1)=0 \quad (51)$$

$$\mu_1f(0,1,0)=r_1P(0,0,0)+p_2^bP(0,1,1)+f_2^bP(0,1,-1) \quad (52)$$

$$\mu_1f(0,-1,0)=p_2^bP(0,-1,1)+f_2^bP(0,-1,-1) \quad (53)$$

$$\mu_2f(N,0,1)=r_2P(N,0,0)+p_1^bP(N,1,1)+f_1^bP(N,-1,1) \quad (54)$$

$$\mu_2f(N,0,-1)=p_1^bP(N,1,-1)+g_2P(N,0,1)+f_1^bP(N,-1,-1) \quad (55)$$

3.2.3 Normalization In addition to these, all the probability density functions and probability masses must satisfy the normalization equation:

$$\sum_{\alpha_1=-1,0,1}\sum_{\alpha_2=-1,0,1}\left[\int_0^N f(x,\alpha_1,\alpha_2)dx+P(0,\alpha_1,\alpha_2)+P(N,\alpha_1,\alpha_2)\right]=1 \quad (56)$$

3.2.4 Performance measures After finding all probability density functions and probability masses, we can calculate the average inventory in the buffer from

$$\overline{x}=\sum_{\alpha_1=-1,0,1}\sum_{\alpha_2=-1,0,1}\left[\int_0^N xf(x,\alpha_1,\alpha_2)dx+NP(N,\alpha_1,\alpha_2)\right] \quad (57)$$

The total production rate is

$$\begin{aligned}&P_T=P_T^1=\\&\sum_{\alpha_2=-1,0,1}\mu_1\left[\int_0^N\{f(x,-1,\alpha_2)+f(x,1,\alpha_2)\}dx+P(0,1,\alpha_2)+P(0,-1,\alpha_2)\right]\\&\quad+\mu_2\{P(N,1,-1)+P(N,1,1)+P(N,-1,-1)+P(N,-1,1\}\end{aligned} \quad (58)$$

The rate at which machine M_1 produces good parts is

$$P_E^1=\sum_{\alpha_2=-1,0,1}\mu_1[\int_0^N f(x,1,\alpha_2)dx+P(0,1,\alpha_2)]+\mu_2\{P(N,1,-1)+P(N,1,1)\} \quad (59)$$

The probability that the first machine produces a non-defective part is then $Y_1 = P_E^1/P_T$. Similarly, the probability that the second machine finishes its operation without adding a bad feature to a part is $Y_2 = P_E^2/P_T$, where

$$P_E^2 = \sum_{\alpha_1=-1,0,1} \mu_2[\int_0^N f(x,\alpha_1,1)dx + P(N,\alpha_1,1)] + \mu_1\{P(0,-1,1) + P(0,1,1)\} \tag{60}$$

Therefore, the effective production rate is

$$P_E = Y_1 Y_2 P_T \tag{61}$$

3.3 Validation

The 2M1B systems with the same machine speed ($\mu_1 = \mu_2$) are solved in Appendix A. As we have indicated, we represent discrete parts in this model as a continuous fluid and time as a continuous variable. We compare analytical and simulation results in this section. In the simulation, both material and time are discrete. Details are presented in [14].

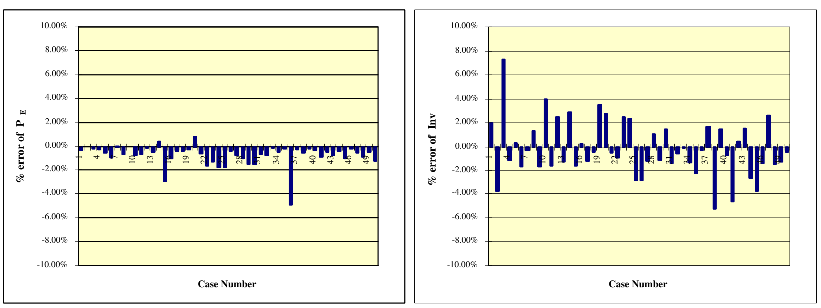

Fig. 5 Validation of the intermediate buffer size case

Figure 5 shows the comparison of the effective production rate and the average inventory from the analytic model and the simulation. 50 cases are generated by changing machines and buffer parameters and % errors are plotted in the vertical axis. The parameters for theses cases are given in Appendix B. The % error in the effective production rate is calculated from

$$P_E\%\text{error} = \frac{P_E(A) - P_E(S)}{P_E(S)} \times 100(\%) \tag{62}$$

where $P_E(A)$ and $P_E(S)$ are the effective production rates estimated from the analytical model and the simulation respectively. But the % error in the average inventory is calculated from

$$Inv_E\%\text{error} = \frac{Inv_E(A) - Inv_E(S)}{0.5 \times N} \times 100(\%) \tag{63}$$

where $Inv_E(A)$ and $Inv_E(S)$ are average inventory estimated from the analytical model and the simulation respectively and N is a buffer size[1].

The average absolute value of the % error in the effective production rate estimation is 0.76% and it is 1.89% for average inventory estimation.

4 Quality Information Feedback

Factory designers and managers know that it is ideal to have inspection after every operation. However, it is often costly to do this. As a result, factories are usually designed so that multiple inspections are performed at a small number of stations. In this case, inspection at downstream operations can detect bad features made by upstream machines. We call this *quality information feedback*. A simple example of the quality information feedback in 2M1B systems is when M_1 produces defective features but does not have inspection and M_2 has inspection and it can detect bad features made by M_1. In this situation, as we demonstrate below, the yield of a line is a function of the size of buffer. This is because when buffer gets larger, more material can accumulate between an operation (M_1) and the inspection of that operation (M_2). All such material will be defective if a persistent quality failure takes place. In other words, if buffer is larger, there tends to be more material in the buffer and consequently more material is defective. In addition it takes longer to have inspections after finishing operations. We can capture this phenomenon with the adjustment of a transition probability rate of M_1 from state -1 to state 0.

Let us define f_1^q as a transition rate of M_1 from state -1 to state 0 when there is a quality information feedback and f_1 as the transition rate without the quality information feedback. The adjustment can be done in a way that the yield of M_1 is the same as $\frac{Z_1^g}{Z_1^g+Z_1^b}$ where

- ⋆ Z_1^b: the expected number of bad parts generated by M_1 while it stays in state -1.
- ⋆ Z_1^g: the expected number of good parts produced by M_1 from the moment when M_1 leaves the -1 state to the next time it arrives at state -1.

[1] This is an unbiased way to calculate the error in average inventory. If it were calculated in the same way as the error in the effective production rate, the error would depend on the relative speeds of the machines. This is because there will be a lower error when the buffer is mostly full (ie, when M_1 is faster than M_2) and a higher error when the buffer is empty (when M_1 is faster than M_2).

From (10), the yield of M_1 is

$$\frac{P(1)}{P(1)+P(-1)} = \frac{f_1^q}{f_1^q + g_1} \tag{64}$$

Suppose that M_1 has been in state 1 for a very long time. Then all parts in the buffer B are non-defective. Suppose that M_1 goes to state -1. Defective parts will then begin to accumulate in the buffer. Until all the parts in the buffer are defective, the only way that M_1 can go to state 0 is due to its own inspection or its own operation failure. Therefore, the probability of a transition to 0 before M_1 finishes a part is

$$\frac{f_1}{\mu_1} \equiv \chi_{11}$$

Eventually all the parts in the buffer are bad so that defective parts reach M_2. Then, there is another way that M_1 can move to state 0 from state -1: quality information feedback. The probability that the inspection at M_2 detects a nonconformity made by M_1 is

$$\chi_{21} \equiv \frac{h_{21}}{\mu_2}$$

where $1/h_{21}$ is the mean time until the inspection at M_2 detects a bad part made by M_1 after M_2 receives the bad part.

The expected value of the number of bad parts produced by M_1 before it is stopped by either operational failures or quality information feedback is

$$Z_1^b = [\chi_{11} + 2\chi_{11}(1-\chi_{11}) + 3\chi_{11}(1-\chi_{11})^2 + \ldots + w\chi_{11}(1-\chi_{11})^{w-1}]$$
$$+[(w+1)(1-\chi_{11})^w\chi_{21} + (w+2)(1-\chi_{11})^{w+1}\chi_{21}(1-\chi_{21}) + \ldots] \tag{65}$$

where w is average inventory in the buffer B. This is an approximate formula since we simply use the average inventory rather than averaging the expected number of bad parts produced by M_1 depending on different inventory levels w_i. After some mathematical manipulation,

$$Z_1^b = \frac{1-(1-\chi_{11})^w}{\chi_{11}} - w(1-\chi_{11})^w$$
$$+\frac{(1-\chi_{11})^w\chi_{21}[(w+1) - w(1-\chi_{11})(1-\chi_{21})]}{[1-(1-\chi_{11})(1-\chi_{21})]^2} \tag{66}$$

On the other hand, Z_1^g is given as

$$Z_1^g = \frac{\mu_1}{p_1+g_1} + \frac{p_1}{p_1+g_1}\frac{\mu_1}{p_1+g_1} + (\frac{p_1}{p_1+g_1})^2(\frac{\mu_1}{p_1+g_1})\ldots = \frac{\mu_1}{g_1} \tag{67}$$

By setting $\frac{f_1^q}{f_1^q+g_1} = \frac{Z_1^g}{Z_1^g+Z_1^b}$ we have

$$f_1^q = \frac{\mu_1}{\frac{1-(1+w\chi_{11})(1-\chi_{11})^w}{\chi_{11}} + \frac{(1-\chi_{11})^w \chi_{21}[1+w(\chi_{21}+\chi_{11}-\chi_{21}\chi_{11})]}{[1-(1-\chi_{11})(1-\chi_{21})]^2}} \tag{68}$$

Since the average inventory is a function of f_1^q and f_1^q is dependent on the average inventory, an iterative method is required to determine these values.

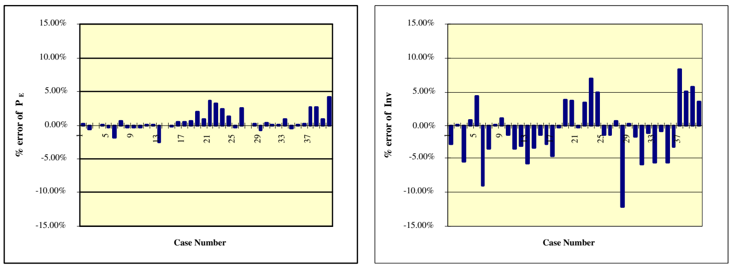

Fig. 6 Validation of the quality information feedback formula

Figure 6 shows the comparison of the effective production rate and the average inventory from the analytic model and the simulation. 50 cases are generated by selecting different machine and buffer parameters and % errors are plotted in the y-axis. The parameters for theses cases are given in Appendix B. % errors in the effective production rate and average inventory are calculated using equations (62) and (63) respectively. The average absolute value of the % error in P_E and $\overline{x}$ estimations are 1.01% and 3.67% respectively.

5 Insights From Numerical Experimentation

In this section, we perform a set of numerical experiments to provide intuitive insight into the behavior of production lines with inspection. The parameters of all the cases are presented in Appendix B.

5.1 Beneficial buffer case

5.1.1 Production rates Having quality information feedback means having more inspection than otherwise. Therefore, machines tend to stop more frequently. As a result, the total production rate of the line decreases. However, the effective production rate can increase since added inspections prevent the making of defective parts. This phenomenon is shown in Figure 7. Note that the total production rate P_T without quality information feedback is consistently higher than P_T with quality information feedback regardless of

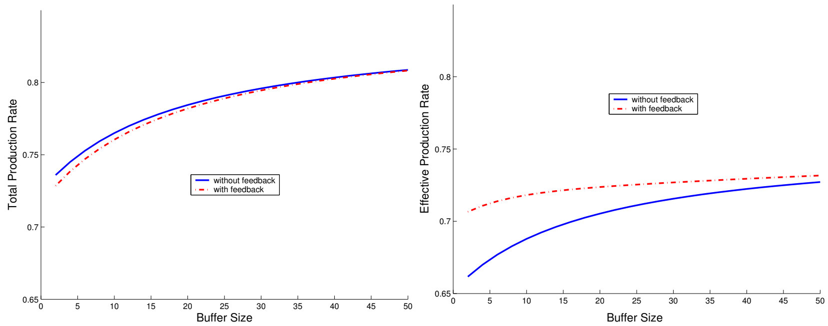

Fig. 7 Production Rates with/without Quality Information Feedback

buffer size and the opposite is true for the effective production rate P_E. Also it should be noted that in this case, both the total production rate and the effective production rate increase with buffer size, with or without quality information feedback.

5.1.2 System yield and buffer size Even though a larger buffer increases both total and effective production rates in this case, it decreases yield. As explained in Section 4, the system yield is a function of the buffer size if there is quality information feedback. Figure 8 shows system yield decreasing as buffer size increases when there is quality information feedback. This happens because when the buffer gets larger, more material accumulates between an operation and the inspection of that operation. All such material will be defective when the first machine is at state -1 but the inspection at the first machine does not find it. This is a case in which *a smaller buffer improves quality,* which is widely believed to be generally true. If there is no quality information feedback, then the system yield is independent of the buffer size (and is substantially less).

5.2 Harmful buffer case

5.2.1 Production rates Typically, increasing the buffer size leads to higher effective production rate. This is the case in Figure 7. But under certain conditions, the effective production rate can actually decrease as buffer size increases. This can happen when

- ⋆ The first machine produces bad parts frequently: this means g_1 is large.
- ⋆ The inspection at the first machine is poor or non-existent and inspection at the second machine is reliable: this means $h_1 \ll h_2$ or $f_1 - p_1 \ll f_2 - p_2$.
- ⋆ There is quality information feedback.
- ⋆ The isolated production rate of the first machine is higher than that of the second machine:

$$\mu_1 \frac{1 + g_1/f_1}{1 + (p_1 + g_1)/r_1 + g_1/f_1} > \mu_2 \frac{1 + g_2/f_2}{1 + (p_2 + g_2)/r_2 + g_2/f_2}$$

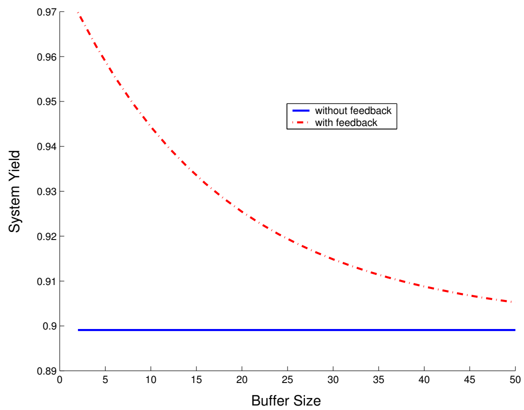

Fig. 8 System Yield as a Function of Buffer Size

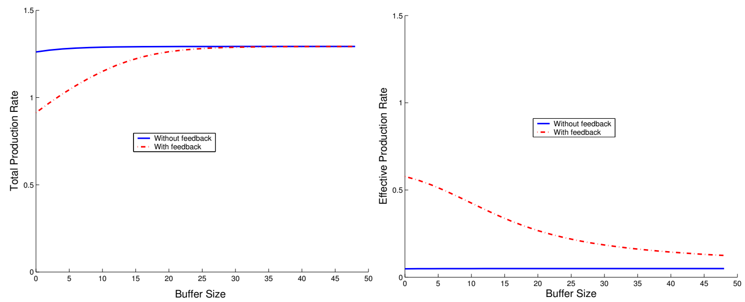

Fig. 9 Total Production Rate and Effective Production Rate

Figure 9 shows a case in which a buffer size increase leads to a lower effective production rate. Note that even in this case, total production rate monotonically increases as buffer size increases.

5.2.2 System Yield The system yield for this case is shown in Figure 10. Note that the yield decreases dramatically as the buffer size increases. In this case, the decrease of the system yield is more than the increase of the total production rate so that the effective production rate monotonically decreases as buffer size gets bigger.

5.3 How to improve quality in a line with persistent quality failures

There are two major ways to improve quality. One is to increase the yield of individual operations and the other is to perform more rigorous inspection. Having extensive preventive maintenance on manufacturing equipment and using robust engineering techniques to stabilize operations have been suggested as tools to increase yield of individual operations. Both approaches

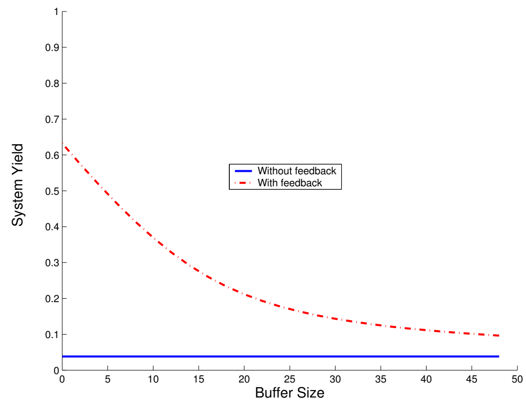

Fig. 10 System Yield as a Function of Buffer Size

increase the Mean Time to Quality Failure (MTQF) (i.e. decrease g). On the other hand, the inspection policy aims to detect bad parts as soon as possible and prevent their flow toward downstream operations. More rigorous inspection decreases the mean time to detect (MTTD) (i.e. increases h and therefore increases f). It is natural to believe that using only one kind of method to achieve a target quality level would not give the most cost efficient quality assurance policy. Figure 11 indicates that the impact of individual operation stabilization on the system yield decreases as the operation becomes more stable. It also shows that effect of improving inspection (MTTD) on the system yield decreases. Therefore, it is optimal to use a combination of both methods to improve quality.

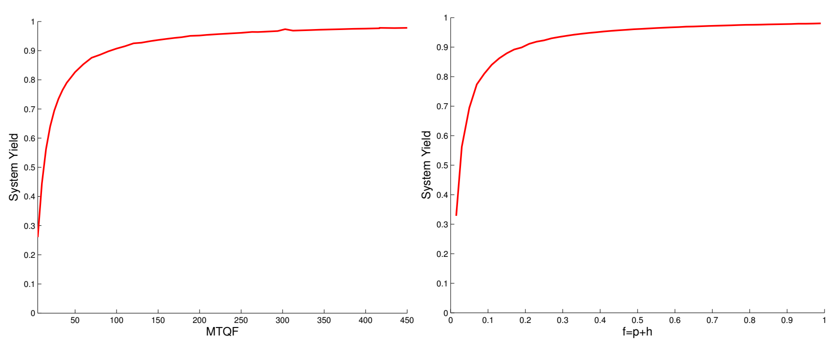

Fig. 11 Quality Improvement

5.4 How to increase productivity

Improving the stand-alone throughput of each operation and increasing the buffer space are typical ways to increase the production rate of manufacturing systems. If operations are apt to have quality failures, however, there may be other ways to increase the effective production rate: increasing the yield of each operation and conducting more extensive inspections. Stabilizing operations, thus improving the yield of individual operations, will increase effective throughput of a manufacturing system regardless of the type of quality failure. On the other hand, reducing the mean time to detect (MTTD) will *increase* the effective production rate only if the quality failure is persistent but it will *decrease* the effective production rate if the quality failure is Bernoulli. This is because the quality of each part is independent of the others when the quality failure is Bernoulli. Therefore, stopping the line does not reduce the number of bad parts in the future.

In a situation in which machines produce defective parts frequently and inspection is poor, increasing inspection reliability is more effective than increasing buffer size to boost the effective production rate. Figure 12 shows this. Also, in other situations in which machines produce defective parts frequently and inspection is reliable, increasing machine stability is more effective than increasing buffer size to enhance effective production rate. Figure 13 shows this phenomenon.

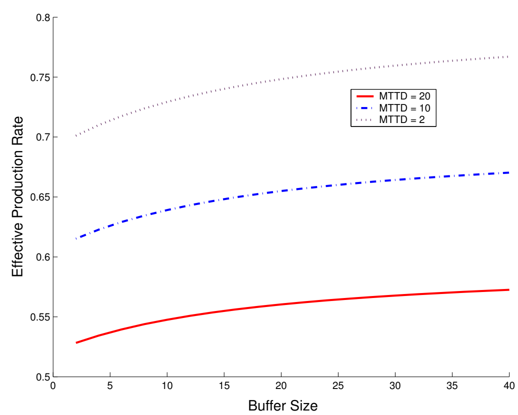

Fig. 12 Mean Time to Detect and Effective Production Rate

6 Future Research

The 2-Machine-1-Buffer (2M1B) model with $\mu_1 \neq \mu_2$ is analyzed in [14]. This case is more challenging because the number of roots of the internal transition equations depends on parameters of machine. A more general

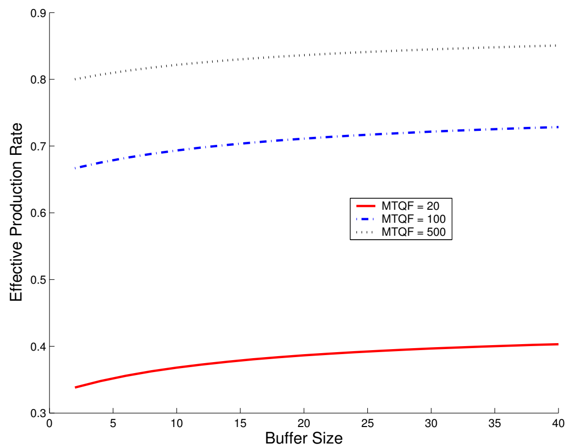

Fig. 13 Quality Failure Frequency and Effective Production Rate

2M1B model with multiple-yield quality failures (a mixture of Bernoulli- and persistent-type quality failures) should also be studied. A long line analysis using decomposition is under the development. Refer to Kim [14] for more detailed information.

Acknowledgments

The authors would like to thank the Singapore-MIT Alliance and PSA Peugeot-Citroën for their support of this research.

References

1. Alles M, Amershi A, Datar S, and Sarkar R (2000) Information and incentive effects of inventory in JIT production. Management science 46 (12): 1528–1544.
2. Besterfield D H, Besterfield-Michna C, Besterfield G, and Besterfield-Sacre M, (2003) Total quality management. Prentice Hall.
3. Black J T (1991) The design of the factory with a future. McGraw-Hill.
4. Bonvik A M, Couch C E, and Gershwin S B (1997) A comparison of production line control mechanisms. International journal of production research 35 (3): 789–804.
5. Burman M, Gershwin S B, and Suyematsu C (1998) Hewlett-Packard uses operations research to improve the design of a printer production line. Interfaces 28 (1): 24–26.
6. Buzacott J A and Shantikumar J G (1993) Stochastic models of manufacturing systems. Prentice-Hall.
7. Cheng C H, Miltenburg J, and Motwani J (2000) The effect of straight and U shaped lines on quality. IEEE Transactions on Engineering Management 47 (3): 321-334.

8. Dallery Y. and Gershwin S B (1992) Manufacturing flow line systems: a review of models and analytical results. Queuing Systems Theory and Applications 12: 3–94.
9. Fujimoto T (1999) The evolution of a manufacturing systems at Toyota. Oxford University Press.
10. Gershwin S B (1994) Manufacturing systems engineering. Prentice Hall.
11. Gershwin S B (2000) Design and Operation of Manufacturing Systems — The Control-Point Policy. IIE Transactions 32 (2): 891-906.
12. Gershwin S B and Schor J E (2000) Efficient algorithms for buffer space allocation. Annals of Operations Research 93: 117-144.
13. Inman R R, Blumenfeld D E, Huang N, and Li J (2003) Designing production systems for quality: research opportunities from an automotive industry perspective. International journal of production research 41 (9): 1953–1971.
14. Kim J (2004) Integrated Quality and Quantity Modeling of a Production Line. Massachusetts Institute of Technology Ph.D. thesis in preparation.
15. Law A M, Kelton D W, Kelton W D, and Kelton D M(1999) Simulation modeling and analysis. McGraw-Hill.
16. Ledolter J and Burrill C W (1999) Statistical quality control. John Wiley & Sons.
17. Monden Y (1998) Toyota production system — An integrated approach to Just-In-Time. EMP Books, 1998.
18. Montgomery D C (2001) Introduction to statistical quality control 4th edition. John Wiley & Sons.
19. Pande P and Holpp L (2002) What is six sigma? McGraw-Hill.
20. Phadke M (1989) Quality engineering using robust design. Prentice Hall.
21. Raz T (1986) A survey of models for allocating inspection effort in multistage production systems. Journal of quality technology 18 (4): 239–246.
22. Shin W S, Mart S M, and Lee H F (1995) Strategic allocation of inspection stations for a flow assembly line: a hybrid procedure. IIE Transactions 27: 707–715.
23. Shingo S, (1989) A study of the Toyota production system from an industrial engineering viewpoint. Productivity Press.
24. Toyota Motor Corporation (1996) The Toyota production system.
25. Wein L, (1988) Scheduling semiconductor wafer fabrication. IEEE Transactions on semiconductor manufacturing 1(3): 115–130.
26. Wooddall W H and Montgomery D C (1999) Research issues and ideas in statistical process control. Journal of Quality Technology 31 (4): 376–386.

Appendix

A Solution technique

It is natural to assume an exponential form for the solution to the steady state density functions since equations (23)–(31) are coupled ordinary linear differential equations. A solution of the form $e^{\lambda x} K_1^{\alpha_1} K_2^{\alpha_2}$ worked successfully in the continuous material two-machine line with perfect quality [10]. Therefore, a solution of a form

$$f(x, \alpha_1, \alpha_2) = e^{\lambda x} G_1(\alpha_1) G_2(\alpha_2) \tag{69}$$

is assumed here. This form satisfies the transition equations if all of the following equations are met. Equations (23)–(31) become, after substituting (69),

$$\{(\mu_2-\mu_1)\lambda-(p_1+g_1+p_2+g_2)G_1(1)G_2(1)\}+r_2G_1(1)G_2(0)+r_1G_1(0)G_2(1)=0 \tag{70}$$

$$-\{\mu_1\lambda+(p_1+g_1+r_2)\}G_1(1)G_2(0)+p_2G_1(1)G_2(1)+f_2G_1(1)G_2(-1) \\ +r_1G_1(0)G_2(0)=0 \tag{71}$$

$$\{(\mu_2-\mu_1)\lambda-(p_1+g_1+f_2)\}G_1(1)G_2(-1)+g_2G_1(1)G_2(1)+r_1G_1(0)G_2(-1)=0 \tag{72}$$

$$\{\mu_2\lambda-(r_1+p_2+g_2)\}G_1(0)G_2(1)+p_1G_1(1)G_2(1)+r_2G_1(0)G_2(0) \\ +f_1G_1(-1)G_2(1)=0 \tag{73}$$

$$p_1G_1(1)G_2(0)+p_2G_1(0)G_2(1)-(r_1+r_2)G_1(0)G_2(0)+f_2G_1(0)G_2(-1) \\ +f_1G_1(-1)G_2(0)=0 \tag{74}$$

$$\{\mu_2\lambda-(r_1+f_2)\}G_1(0)G_2(-1)+p_1G_1(1)G_2(-1)+g_2G_1(0)G_2(1) \\ +f_1G_1(-1)G_2(-1)=0 \tag{75}$$

$$\{(\mu_2-\mu_1)\lambda-(p_2+g_2+f_1)\}G_1(-1)G_2(1)+g_1G_1(1)G_2(1)+r_2G_1(-1)G_2(0)=0 \tag{76}$$

$$-\{\mu_1\lambda+(r_2+f_1)\}G_1(-1)G_2(0)+g_1G_1(1)G_2(0)+p_2G_1(-1)G_2(1) \\ +f_2G_1(-1)G_2(-1)=0 \tag{77}$$

$$\{(\mu_2-\mu_1)\lambda-(f_1+f_2)\}G_1(-1)G_2(-1)+g_1G_1(1)G_2(-1)+g_2G_1(-1)G_2(1)=0 \tag{78}$$

These are nine equations in seven unknowns ($\lambda, G_1(1)$, $G_2(0)$, $G_1(-1)$, $G_2(1)$, $G_2(0)$, and $G_2(-1)$). Thus, there must be seven independent equations and two dependent ones.

If we divide equations (70) – (78) by $G_1(0)G_2(0)$ and define new parameters

$$\Gamma_i = p_i\frac{G_i(1)}{G_i(0)} - r_i + f_i\frac{G_i(-1)}{G_i(0)} \tag{79}$$

$$\Psi_i = -p_i - g_i + r_i\frac{G_i(0)}{G_i(1)} \tag{80}$$

$$\Theta_i = -f_i + g_i \frac{G_i(1)}{G_i(-1)} \tag{81}$$

then equations (70)–(78) can be rewritten as

$$\Gamma_1 + \Gamma_2 = 0 \tag{82}$$

$$-\mu_2\lambda = \Gamma_1 + \Psi_2 \tag{83}$$

$$\mu_1\lambda = \Gamma_2 + \Psi_1 \tag{84}$$

$$(\mu_1 - \mu_2)\lambda = \Psi_1 + \Psi_2 \tag{85}$$

$$(\mu_1 - \mu_2)\lambda = \Theta_1 + \Theta_2 \tag{86}$$

$$\mu_1\lambda = \Gamma_2 + \Theta_1 \tag{87}$$

$$-\mu_2\lambda = \Gamma_1 + \Theta_2 \tag{88}$$

$$(\mu_1 - \mu_2)\lambda = \Psi_2 + \Theta_1 \tag{89}$$

$$(\mu_1 - \mu_2)\lambda = \Psi_1 + \Theta_2 \tag{90}$$

From equations (82)–(90), it is clear that only seven equations are independent. After much mathematical manipulation [14], these equations become

$$\begin{aligned} 0 = {} & \frac{\{(M+r_1)(\mu_1 N-1)-f_1\}^2}{(f_1-p_1)(\mu_1 N-1)} \\ & - \frac{\{(p_1+g_1-f_1)+r_1(\mu_1 N-1)\}\{(M+r_1)(\mu_1 N-1)-f_1\}}{(f_1-p_1)(\mu_1 N-1)} - r_1 \end{aligned} \tag{91}$$

$$\begin{aligned} 0 = {} & \frac{\{(-M+r_2)(\mu_2 N-1)-f_2\}^2}{(f_2-p_2)(\mu_2 N-1)} \\ & - \frac{\{(p_2+g_2-f_2)+r_2(\mu_2 N-1)\}\{(-M+r_2)(\mu_2 N-1)-f_2\}}{(f_2-p_2)(\mu_2 N-1)} - r_2 = 0 \end{aligned} \tag{92}$$

where

$$p_1 \frac{G_1(1)}{G_1(0)} - r_1 + f_1 \frac{G_1(-1)}{G_1(0)} = -\left(p_2 \frac{G_2(1)}{G_2(0)} - r_2 + f_2 \frac{G_2(-1)}{G_2(0)} \right) = M \tag{93}$$

$$\begin{aligned} & \frac{1}{\mu_1}\left(1 + \frac{1}{G_1(1)/G_1(0) + G_1(-1)/G_1(0)}\right) = \\ & \frac{1}{\mu_2}\left(1 + \frac{1}{G_2(1)/G_2(0) + G_2(-1)/G_2(0)}\right) = N \end{aligned} \tag{94}$$

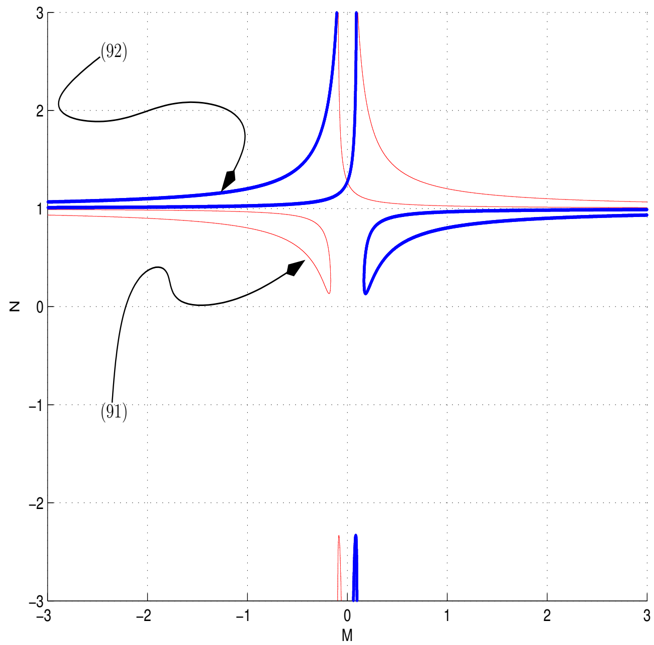

Fig. 14 Plot of Equations (91) and (92)

Now all the equations and unknowns are simplified into two unknowns and two equations. By solving equations (91) and (92) simultaneously we can calculate M and N. An example of these equations is plotted in Figure 14. Equation (91) is represented with lighter lines and equation (92) is shown as darker lines. The intersections of the two sets of lines are the solutions of the equations.

These are high order polynomial equations for which no general analytical solution exists. A numerical approach is required to find the roots of the equations. A special algorithm to find the solutions has been developed [14] based on the characteristics of the equations. Once we find roots of equations (91) and (92), we can get ratios $\frac{G_i(1)}{G_i(0)}$ and $\frac{G_i(-1)}{G_i(0)}$ $(i = 1, 2)$ from equation (94). By setting $G_1(0) = G_2(0) = 1$, we can calculate $G_1(1), G_1(-1), G_2(1)$, and $G_2(-1)$. After some mathematical manipulation, we find that λ can be expressed as

$$\lambda = \frac{-p_1 - g_1 + {}^{r_1}\!/_{G_1(1)} - p_1 G_1(1) + r_1 - f_1 G_1(-1)}{\mu_1} \tag{95}$$

Therefore, we can get a probability density function $f(x, \alpha_1, \alpha_2)$ corresponding to an (M, N) pair. The number of roots in equations (91) and (92) depends on machine parameters. There are only 3 roots when $\mu_1 = \mu_2$

regardless of other parameters. Therefore, a general expression of the probability density function in this case is

$$f(x,\alpha_1,\alpha_2) = c_1 f_1(x,\alpha_1,\alpha_2) + c_2 f_2(x,\alpha_1,\alpha_2) + c_3 f_3(x,\alpha_1,\alpha_2) \qquad (96)$$

where $f_1(x,\alpha_1,\alpha_2)$, $f_2(x,\alpha_1,\alpha_2)$, $f_3(x,\alpha_1,\alpha_2)$ are the roots of the equations (91) and (92).

The remaining unknowns, including c_1, c_2, c_3 and probability masses at the boundaries, can be calculated by solving boundary equations ((34)–(55)) and the normalization equation (56) with $f_i(x,\alpha_1,\alpha_2)$ given by equation (96).

B Machine Parameters for Numerical and Simulation Experiments

Table 4 Machine Parameters for Infinite Buffer Case and Zero Buffer Case

Case #	μ_1	μ_2	r_1	r_2	p_1	p_2	g_1	g_2	f_1	f_2
1	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2
2	1.0	1.0	0.3	0.3	0.005	0.005	0.05	0.05	0.5	0.5
3	1.0	1.0	0.2	0.05	0.01	0.01	0.01	0.01	0.2	0.2
4	1.0	1.0	0.1	0.1	0.05	0.005	0.01	0.01	0.2	0.2
5	1.0	1.0	0.1	0.1	0.01	0.01	0.05	0.005	0.2	0.2
6	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.5	0.1
7	2.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.5	0.1
8	3.0	2.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2
9	1.0	2.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2
10	2.0	3.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2

Table 5 Machine Parameters for Figures 7 and 8

μ_1	μ_2	r_1	r_2	p_1	p_2	g_1	g_2	f_1	f_2
1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.1	0.9

Table 6 Machine Parameters for Figures 9 and 10

μ_1	μ_2	r_1	r_2	p_1	p_2	g_1	g_2	f_1	f_2
2.0	2.0	0.5	0.1	0.005	0.05	0.5	0.005	0.02	0.9

Table 7 Machine Parameters for Figure 11

μ_1	μ_2	r_1	r_2	p_1	p_2	g_1	g_2	f_1	f_2
1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2

Table 8 Machine Parameters for Figure 12

μ_1	μ_2	r_1	r_2	p_1	p_2	g_1	g_2
1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01

Table 9 Machine Parameters for Figure 13

μ_1	μ_2	r_1	r_2	p_1	p_2	f_1	f_2
1.0	1.0	0.1	0.1	0.01	0.01	0.2	0.2

Table 10 Machine Parameters for Intermediate Buffer Case Validation

Case #	μ_1	μ_2	r_1	r_2	p_1	p_2	g_1	g_2	f_1	f_2	N
1	1.0	1.0	0.1	0.1	0.01	0.01	0.02	0.01	0.1	0.2	30
2	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	5
3	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	10
4	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	15
5	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	20
6	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	25
7	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	35
8	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	40
9	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	45
10	0.5	0.5	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	30
11	1.5	1.5	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	30
12	2.0	2.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	30
13	2.5	2.5	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	30
14	3.0	3.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.2	30
15	1.0	1.0	0.01	0.01	0.01	0.01	0.01	0.01	0.2	0.2	30
16	1.0	1.0	0.05	0.05	0.01	0.01	0.01	0.01	0.2	0.2	30
17	1.0	1.0	0.2	0.2	0.01	0.01	0.01	0.01	0.2	0.2	30
18	1.0	1.0	0.5	0.5	0.01	0.01	0.01	0.01	0.2	0.2	30
19	1.0	1.0	0.8	0.8	0.01	0.01	0.01	0.01	0.2	0.2	30
20	1.0	1.0	0.1	0.1	0.001	0.001	0.01	0.01	0.2	0.2	30
21	1.0	1.0	0.1	0.1	0.005	0.005	0.01	0.01	0.2	0.2	30
22	1.0	1.0	0.1	0.1	0.02	0.02	0.01	0.01	0.2	0.2	30
23	1.0	1.0	0.1	0.1	0.05	0.05	0.01	0.01	0.2	0.2	30
24	1.0	1.0	0.1	0.1	0.1	0.1	0.01	0.01	0.2	0.2	30
25	1.0	1.0	0.1	0.1	0.01	0.01	0.001	0.001	0.2	0.2	30
26	1.0	1.0	0.1	0.1	0.01	0.01	0.005	0.005	0.2	0.2	30
27	1.0	1.0	0.1	0.1	0.01	0.01	0.02	0.02	0.2	0.2	30
28	1.0	1.0	0.1	0.1	0.01	0.01	0.05	0.05	0.2	0.2	30
29	1.0	1.0	0.1	0.1	0.01	0.01	0.10	0.10	0.2	0.2	30
30	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.02	0.02	30
31	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.05	0.05	30
32	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.1	0.1	30
33	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.5	0.5	30
34	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.95	0.95	30
35	1.0	1.0	0.5	0.1	0.01	0.01	0.01	0.01	0.2	0.2	30
36	1.0	1.0	0.01	0.1	0.01	0.01	0.01	0.01	0.2	0.2	30
37	1.0	1.0	0.1	0.5	0.01	0.01	0.01	0.01	0.2	0.2	30
38	1.0	1.0	0.1	0.01	0.01	0.01	0.01	0.01	0.2	0.2	30
39	1.0	1.0	0.1	0.1	0.1	0.01	0.01	0.01	0.2	0.2	30
40	1.0	1.0	0.1	0.1	0.001	0.01	0.01	0.01	0.2	0.2	30
41	1.0	1.0	0.1	0.1	0.01	0.1	0.01	0.01	0.2	0.2	30
42	1.0	1.0	0.1	0.1	0.01	0.001	0.01	0.01	0.2	0.2	30
43	1.0	1.0	0.1	0.1	0.01	0.01	0.1	0.01	0.2	0.2	30
44	1.0	1.0	0.1	0.1	0.01	0.01	0.001	0.01	0.2	0.2	30
45	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.1	0.2	0.2	30
46	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.001	0.2	0.2	30
47	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.9	0.2	30
48	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.05	0.2	30
49	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.9	30
50	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	0.05	30

Table 11 Machine Parameters for Quality Information Feedback Validation

Case #	μ_1	μ_2	r_1	r_2	p_1	p_2	g_1	g_2	f_1	f_2	N
1	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	10
2	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	0
3	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	5
4	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	20
5	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	30
6	1.0	1.0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	1.0	10
7	1.0	1.0	0.05	0.05	0.01	0.01	0.01	0.01	0.01	1.0	10
8	1.0	1.0	0.4	0.4	0.01	0.01	0.01	0.01	0.01	1.0	10
9	1.0	1.0	0.8	0.8	0.01	0.01	0.01	0.01	0.01	1.0	10
10	1.0	1.0	0.1	0.1	0.001	0.001	0.01	0.001	0.01	1.0	10
11	1.0	1.0	0.1	0.1	0.005	0.005	0.01	0.005	0.01	1.0	10
12	1.0	1.0	0.1	0.1	0.02	0.02	0.01	0.01	0.02	1.0	10
13	1.0	1.0	0.1	0.1	0.1	0.1	0.01	0.01	0.1	1.0	10
14	1.0	1.0	0.1	0.1	0.01	0.01	0.001	0.001	0.01	1.0	10
15	1.0	1.0	0.1	0.1	0.01	0.01	0.005	0.005	0.01	1.0	10
16	1.0	1.0	0.1	0.1	0.01	0.01	0.02	0.02	0.01	1.0	10
17	1.0	1.0	0.1	0.1	0.01	0.01	0.05	0.05	0.01	1.0	10
18	0.5	0.5	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	10
19	1.5	1.5	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	10
20	2.0	2.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	1.0	10
21	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.05	1.0	10
22	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	1.0	10
23	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.5	1.0	10
24	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.8	1.0	10
25	1.0	1.0	0.5	0.1	0.01	0.01	0.01	0.01	0.01	1.0	10
26	1.0	1.0	0.01	0.1	0.01	0.01	0.01	0.01	0.01	1.0	10
27	1.0	1.0	0.1	0.5	0.01	0.01	0.01	0.01	0.01	1.0	10
28	1.0	1.0	0.1	0.01	0.01	0.01	0.01	0.01	0.01	1.0	10
29	1.0	1.0	0.1	0.1	0.1	0.01	0.01	0.01	0.1	1.0	10
30	1.0	1.0	0.1	0.1	0.001	0.01	0.01	0.01	0.001	1.0	10
31	1.0	1.0	0.1	0.1	0.01	0.1	0.01	0.01	0.01	1.0	10
32	1.0	1.0	0.1	0.1	0.01	0.001	0.01	0.01	0.01	1.0	10
33	1.0	1.0	0.1	0.1	0.01	0.01	0.05	0.01	0.01	1.0	10
34	1.0	1.0	0.1	0.1	0.01	0.01	0.001	0.01	0.01	1.0	10
35	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.05	0.01	1.0	10
36	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.001	0.01	1.0	10
37	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.5	1.0	10
38	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.2	1.0	10
39	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	0.8	10
40	1.0	1.0	0.1	0.1	0.01	0.01	0.01	0.01	0.01	0.2	10