

AN ANALYTICAL FORMULA FOR THROUGHPUT OF A PRODUCTION LINE WITH IDENTICAL STATIONS AND RANDOM FAILURES

DENNIS E. BLUMENFELD AND JINGSHAN LI

Received 22 November 2004

We derive a simple formula for the throughput (jobs produced per unit time) of a serial production line with workstations that are subject to random failures. The derivation is based on equations developed for a line flow model that takes into account the impact of finite buffers between workstations. The formula applies in the special case of a line with identical workstations and buffers of equal size. It is a closed-form expression that shows the mathematical relationships between the system parameters, and that can be used to gain basic insight into system behavior at the initial design stage.

1. Introduction

An important measure of performance for a production line is the system throughput (i.e., the average number of jobs produced per hour). Various analytical models have been developed to analyze throughput and identify bottlenecks for a production line composed of a series of workstations separated by buffers (Gershwin [11, 12]; Buzacott and Shan-thikumar [6]; Jacobs and Meerkov [15]; Chiang et al. [7]; Alden [1]). These models take into account that the stations are subject to random failures. The throughput of a line depends on each station's speed (processing rate) and reliability, and the sizes of the buffers.

The objective of this paper is to obtain a simple formula for throughput from general equations given in the model by Alden [1]. The model considers a two-station line and provides a building block for modeling longer lines. It analyzes the flow of jobs through a line of stations and derives analytical equations for line performance.

Alden's model is developed for the general case in which stations can have different speeds and reliabilities, and buffers can have different sizes. Since the model yields analytical equations, it can be used to compute throughput for such general serial lines very efficiently, and allows quick "what-if" comparisons. As a result of this generality, however, the equations are very involved. They are suited for conveniently computing numerical results rather than providing insight from their functional form.

This paper derives a formula for throughput in the special case of a serial line with identical stations and buffers of equal sizes. From the general equations of the model, the paper utilizes simplifications for the special case. The purpose is to have a simple

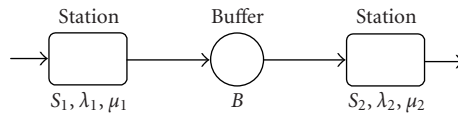


Figure 2.1. Two-station production line.

analytical result that shows the mathematical relationships between the key system parameters. This is useful in the initial design stage, when basic insight into system behavior is needed before detailed numerical analyses are performed.

A formula with a simple structure is helpful in several ways. It helps provide an intuitive understanding of the underlying model, so that analysts and decision makers can use the model with confidence. It highlights tradeoffs between the different parameters in the model, and allows general conclusions to be drawn about system behavior. A formula also allows quick “back of the envelope” calculations. It does not depend on a particular type of operating system, programming language, or user interface. It can easily be incorporated into a spreadsheet tool as part of a wider analysis for strategic planning.

The paper starts with the basic model for a general two-station line developed by Alden [1], and uses the model to derive a throughput formula in the special case of identical stations. The paper then extends the formula to apply to a line of any length. The extended formula is compared with numerical results obtained from simulation.

2. Two-station line

2.1. Alden’s model. The model developed by Alden [1] analyzes a production line consisting of two stations in series, separated by a buffer. Jobs flow along the line to be processed at each station. The average number of jobs per unit time that can flow along the line is the line’s throughput. The two stations are each subject to failures, which affect the throughput. Each station is characterized by its fixed speed (processing rate) and reliability parameters (failure rate and repair rate).

Figure 2.1 depicts the two-station line. The buffer holds jobs processed at the first station and waiting to be processed at the second station. If the buffer is full, the first station is “blocked” and cannot release a job or process new jobs. If the buffer is empty, the second station is “starved” and has no new jobs to process. Alden’s model captures the impact of blocking and starving on the number of jobs in the buffer, and uses results on the buffer content to obtain the line’s throughput. The equations are developed in the model for a two-station line in general and for a line where the two stations have equal speeds.

In this paper, the parameters for the general two-station model are first defined, and the assumptions are stated. The basic equations developed in the model for stations with equal speeds are then presented. These equations are used to derive a formula for throughput in the special case of identical stations (i.e., stations with equal speeds, and also equal failure rates and repair rates).

For the general two-station line shown in Figure 2.1, the parameters are

- (i) S_i , the speed of station i ($i = 1, 2$) (jobs per hour),
- (ii) λ_i , the failure rate of station i ($i = 1, 2$) (failures per hour),

- (iii) μ_i , the repair rate of station i ($i = 1, 2$) (repairs per hour),
- (iv) B , the buffer size ($0 \leq B < \infty$) (number of jobs).

The speed S_i is the number of jobs station i processes per hour when not blocked, starved, or failed. The failure rate λ_i is the number of failures per hour of operating time, that is, $1/\lambda_i$ is the mean operating time between failures (MTBF _{i}) ($i = 1, 2$).

The repair rate μ_i is the number of repairs per hour, that is, $1/\mu_i$ is the mean time to repair (MTTR _{i}) or mean downtime ($i = 1, 2$).

The buffer size B is the number of jobs the buffer can hold.

For given reliability parameters λ_i and μ_i , the fraction of time that station i is available for processing jobs if never blocked or starved is

$$\frac{\mu_i}{\lambda_i + \mu_i} \quad \text{or} \quad \frac{\text{MTBF}_i}{\text{MTBF}_i + \text{MTTR}_i}. \quad (2.1)$$

This fraction is known as the station's *stand-alone availability* or *efficiency*. The effective speed \hat{S}_i of station i , accounting for its stand-alone availability, is therefore

$$\hat{S}_i = S_i \left(\frac{\mu_i}{\lambda_i + \mu_i} \right) \quad (i = 1, 2). \quad (2.2)$$

The speed \hat{S}_i in (2.2) is often referred to as the station's *stand-alone throughput*.

Alden's model treats the movement of jobs through the line as a fluid flow, and develops equations for the system in steady state in terms of the above parameters. The model is based on the following assumptions about the system.

- (a) The buffer does not fail, and jobs flow through it with zero transit time.
- (b) A station does not fail if it is blocked or starved (i.e., it is subject to failure only when operating).
- (c) Operating times between failures at a station are exponentially distributed (with mean $1/\lambda_i$, $i = 1, 2$).
- (d) Repair times at a station are exponentially distributed (with mean $1/\mu_i$, $i = 1, 2$).
- (e) The first station is never starved and the second (i.e., last) station is never blocked, so that there are no external impediments to the line's operation. This assumption ensures that the analysis determines the maximum number of jobs that can flow through the line.

The equations derived for the two-station model also require the following assumption.

- (f) While one station is down, the other station does not fail, but its speed is reduced to its normal speed multiplied by its stand-alone availability, that is, its speed is reduced to \hat{S}_i as given by (2.2). This assumption is an approximation to account for the possibility of both stations being down simultaneously.

The above assumptions are discussed fully in [1].

The model analyzes the buffer content dynamics by considering the possible system states. For the case where stations have equal speeds, the states are

- (i) U (up): both stations are processing,
- (ii) F (filling): station 1 is processing while station 2 has failed,

- (iii) *E* (emptying): station 2 is processing while station 1 has failed,
- (iv) *FB* (fail-blocked): station 1 is blocked because station 2 has failed,
- (v) *FS* (fail-starved): station 2 is starved because station 1 has failed.

In general, there are two additional states: *SB* (speed-blocked), where station 1 is blocked because station 2 has a slower speed, and *SS* (speed-starved), where station 2 is starved because station 1 has a slower speed. However, for the case of stations with equal speeds, speed blocking and speed starving do not occur and therefore these two states (*SB* and *SS*) need not be considered here. The general model is described in [1] and the main results are summarized in [16].

The approach in Alden’s model is to apply renewal theory to obtain the distribution of buffer content at a given renewal epoch (Alden [1]), and derive the expected times over a renewal period that the system is in each of the above states. From the basic relationships for the case of equal speeds (Alden [1, (8.28)-(8.29), page 84]), the expected renewal period $E(T^C)$ is the sum of the expected times spent in each of the possible states, that is,

$$E(T^C) = E(T^U) + E(T^F) + E(T^E) + E(T^{FB}) + E(T^{FS}), \tag{2.3}$$

where $E(T^U)$, $E(T^F)$, $E(T^E)$, $E(T^{FB})$, and $E(T^{FS})$ denote the expected times spent in states *U*, *F*, *E*, *FB*, and *FS*, respectively, and the fractions of time $P(U)$ and $P(E)$ that the system is in states *U* and *E*, respectively, are given by

$$P(U) = \frac{E(T^U)}{E(T^C)}, \tag{2.4}$$

$$P(E) = \frac{E(T^E)}{E(T^C)}. \tag{2.5}$$

The system throughput ρ for this case is given by (Alden [1, (8.30), page 84])

$$\rho = S_2P(U) + \hat{S}_2P(E), \tag{2.6}$$

where $\hat{S}_2 = S_2\mu_2/(\lambda_2 + \mu_2)$ from (2.2).

In developing equations for each of these quantities, Alden introduces the following intermediate variables in terms of the station parameters (Alden [1, pages 18, 21, 23, and 107]):

$$\theta_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad \theta_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \quad r_1 = \frac{\mu_1}{\hat{S}_2}, \quad r_2 = \frac{\mu_2}{\hat{S}_1}. \tag{2.7}$$

Equations for the expected times in each state are expressed in terms of these variables. For the general model, different sets of equations are developed according to whether the station speeds are equal or different, and on whether a quantity denoted in [1] by α_2 is zero or nonzero. In the case of equal speeds, $\alpha_2 = r_1\theta_2 - r_2\theta_1$ (Alden [1, (8.11)]).

For the special case of identical stations, the station speeds are equal (i.e., $S_1 = S_2$), and also $r_1\theta_2 - r_2\theta_1 = 0$ (i.e., $\alpha_2 = 0$). The equations that apply in this case are (Alden [1, (8.21)–(8.27), pages 111–112])

$$\begin{aligned} E(T^U) &= \frac{1}{\lambda_1 + \lambda_2}, \\ E(T^F) &= \frac{B\theta_2}{\hat{S}_1} P_0, \\ E(T^E) &= \frac{B\theta_2}{\hat{S}_2} P_0, \\ E(T^{FB}) &= \frac{\theta_2}{\theta_1\mu_2} P_0, \\ E(T^{FS}) &= \frac{1}{\mu_2} P_0, \end{aligned} \tag{2.8}$$

where P_0 is the probability that the buffer is empty at the renewal epoch (taken as the moment a repair is completed) and is given by (Alden [1, (8.20), page 81])

$$P_0 = \frac{\theta_1}{1 + r_2\theta_1 B}. \tag{2.9}$$

2.2. Formula for throughput. Given the above equations, we now derive a formula for throughput in the special case of identical stations. For this case, the station parameters become

$$\begin{aligned} S_1 &= S_2 = S, \\ \lambda_1 &= \lambda_2 = \lambda, \\ \mu_1 &= \mu_2 = \mu, \end{aligned} \tag{2.10}$$

where S , λ , and μ denote station speed, failure rate, and repair rate, respectively. The above variables then reduce to

$$\begin{aligned} \theta_1 &= \theta_2 = \frac{1}{2}, \\ r_1 &= r_2 = \frac{\lambda + \mu}{S}, \end{aligned} \tag{2.11}$$

and the expected times spent in each state simplify to give

$$\begin{aligned} E(T^U) &= \frac{1}{2\lambda}, \\ E(T^F) &= E(T^E) = \frac{B(\lambda + \mu)}{2\mu S} P_0, \\ E(T^{FB}) &= E(T^{FS}) = \frac{1}{\mu} P_0, \end{aligned} \tag{2.12}$$

where P_0 is given by

$$P_0 = \frac{1}{2 + (\lambda + \mu)B/S}. \quad (2.13)$$

Substituting in (2.3), (2.4), and (2.5), the expected renewal period $E(T^C)$ becomes

$$E(T^C) = \frac{1}{2\lambda} + \frac{1}{\mu}, \quad (2.14)$$

and $P(U)$ and $P(E)$ become

$$\begin{aligned} P(U) &= \frac{\mu}{2\lambda + \mu}, \\ P(E) &= \frac{\mu}{2\lambda + \mu} \left(\frac{1}{1 + 2S/\{(\lambda + \mu)B\}} \right). \end{aligned} \quad (2.15)$$

Hence, the throughput ρ in (2.6) is given by

$$\rho = S \left(\frac{1}{1 + 2(\lambda/\mu)} \right) \left(1 + \frac{(\lambda/\mu)(B\mu/S)}{2 + (1 + \lambda/\mu)(B\mu/S)} \right). \quad (2.16)$$

With some algebraic manipulation, (2.16) for the throughput ρ can be rewritten as

$$\rho = \frac{S}{1 + \lambda/\mu + (\lambda/\mu)/\{1 + (1/2)(1 + 2\lambda/\mu)(B\mu/S)\}}. \quad (2.17)$$

Similar closed-form results for a two-station line can be obtained from the models by Gershwin [11, 12], Buzacott and Shanthikumar [6], Jacobs and Meerkov [15], Chiang et al. [7], and Li and Meerkov [17]. Numerical comparisons in [16] show close agreement between (2.17) and the other closed-form results.

Although the system is defined by four parameters (S , λ , μ , and B), (2.16) and (2.17) show that throughput ρ depends on just three basic quantities: S , λ/μ , and $B\mu/S$. Each of these three quantities has a clear physical meaning. Station speed S defines the rate of production if there were no failures, the ratio λ/μ defines the station's stand-alone availability or efficiency given there are failures, and the ratio $B\mu/S$ is a measure of buffer effectiveness during failures, indicating how many jobs a buffer can accommodate during $B\mu/S$ average downtimes.

One of the benefits of a simple analytical formula is that these key relationships between the parameters become apparent. This is important in gaining insight into the factors affecting throughput performance. In particular, the formula shows that it is not buffer size B in isolation that is critical in understanding the impact of a buffer on throughput, but rather buffer size in relation to station speed and downtime. What matters for throughput is not simply the number of jobs that a buffer can hold, but the amount of time it can keep the line operating in the event of a failure. The time to empty a full buffer of size B (or fill an empty buffer) is B/S . This is the amount of extra time the buffer can buy for production when a failure occurs. The ratio of this time B/S to the mean downtime $1/\mu$ gives the number of downtimes that the line can continue operating.

This ratio (i.e., $B\mu/S$) is therefore the basic measure of a buffer's effectiveness in maintaining throughput, and is the quantity that should be considered in buffer sizing decisions. A similar conclusion is introduced in [9, 10], where the level of buffering is also defined by a station's average downtime.

The throughput formula given in (2.16) and (2.17) can be checked against standard results in the two extreme cases of no buffer ($B = 0$) and infinite buffer ($B = \infty$). When $B = 0$, the formula reduces to

$$\rho = \frac{S}{1 + 2\lambda/\mu}, \quad (2.18)$$

which is the result for a two-station transfer line (Buzacott [4]). When $B = \infty$, there is no interaction between the two stations, and the formula reduces to a station's stand-alone throughput:

$$\rho = \frac{S}{1 + \lambda/\mu}, \quad (2.19)$$

as given in (2.2).

The formula given in (2.16) and (2.17) is derived directly from Alden's model for a two-station line, in the special case of identical stations. The following section uses this formula to develop an approximate formula for throughput of a line with any number of identical stations.

3. Line with M stations

Alden's model for a two-station line (Alden [1]) provides the basis for modeling longer lines. The blocking and starving interactions in a longer line become too complex to allow an exact analysis, so we develop here an approximation based on the result for two stations.

The purpose of this section is to extend the two-station throughput formula derived above, so that it applies to a line with any number of stations. The extension holds under the same assumptions, (a)–(e), as for the two-station model. We consider again the special case of identical stations, and assume also that the buffers between stations all have the same size.

Figure 3.1 depicts the line of general length. The parameters for this case are

- (i) S , the station speed (jobs per hour),
- (ii) λ , the failure rate (failures per hour),
- (iii) μ , the repair rate (repairs per hour),
- (iv) B , the buffer size (number of jobs),
- (v) M , the line length (number of stations).

To obtain a formula for throughput ρ that includes the additional parameter M , we make the conjecture that two-station formula in (2.17) be extended to have the following functional form:

$$\rho \cong \frac{S}{1 + \lambda/\mu + \{f_1(M)\lambda/\mu\}/\{1 + f_2(M)(1 + 2\lambda/\mu)(B\mu/S)\}}, \quad (3.1)$$

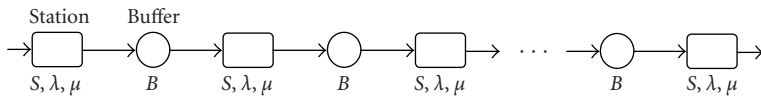


Figure 3.1. Serial line of general length with identical stations and buffers of equal size.

where $f_1(M)$ and $f_2(M)$ are functions to be determined. The function $f_1(M)$ provides the M -station generalization for the impact of station failures, independent of the buffers in the line, while $f_2(M)$ accounts for the impact of buffers specifically.

Consider first the standard result for an M -station line with no buffers (i.e., for a transfer line). The throughput in this case is given by (Buzacott [4])

$$\rho = \frac{S}{1 + M\lambda/\mu} \tag{3.2}$$

Putting $B = 0$ in (3.1) gives

$$\rho = \frac{S}{1 + \lambda/\mu + f_1(M)\lambda/\mu} \tag{3.3}$$

Matching (3.3) with the Buzacott result (3.2), we obtain the following form for the function $f_1(M)$:

$$f_1(M) = M - 1. \tag{3.4}$$

The simplest form for $f_2(M)$ is a linear function, $f_2(M) = kM$, where k is a constant. With this function, the impact of buffers is proportional to M . For the special case of $M = 2$, comparing (2.17) and (3.1) gives $f_2(2) = 1/2$. Thus, in terms of k , we have $2k = 1/2$ or $k = 1/4$, and hence $f_2(M)$ is given by

$$f_2(M) = \frac{M}{4}. \tag{3.5}$$

Putting (3.4) and (3.5) for $f_1(M)$ and $f_2(M)$ in (3.1), we obtain the following approximate expression for throughput ρ of an M -station line:

$$\rho \cong \frac{S}{1 + \lambda/\mu + \{(M - 1)\lambda/\mu\}/\{1 + (M/4)(1 + 2\lambda/\mu)(B\mu/S)\}} \tag{3.6}$$

Alternative functional forms for an extension of the two-station formula to an M -station line were also considered, but these did not perform as well as the functional form given by (3.6). Comparisons of (3.6) with simulation results for throughput are given in the next section.

Note that the extension given by (3.6) preserves the basic relationships between parameters, as seen in the two-station formula (2.17). These include the ratio λ/μ that determines a station's efficiency, and the ratio $B\mu/S$ that provides a measure of buffer effectiveness during failures.

Table 4.1. Station parameter values used in throughput comparisons.

	S (jobs/h)	λ (failures/h)	μ (repairs/h)	MTBF (min)	MTTR (min)	λ/μ
Short frequent failures	60	3	30	20	2	0.1
Long infrequent failures	60	0.3	3	200	20	0.1

4. Numerical comparisons

This section compares the throughput formula given by (3.6) with numerical throughput estimates using the simulation software “Simul8” (Hauge and Paige [13]). The comparisons are made for given values of the station parameters S , λ , and μ , and a range of different line lengths M and buffer sizes B .

Table 4.1 shows the parameter values used for each station. Station speed S is taken as 60 jobs per hour (i.e., cycle time for a job is 1 minute). Two sets of values for the station reliability parameters λ and μ are considered: *short frequent failures* and *long infrequent failures*. These cover two representative scenarios of reliability. In both scenarios, the ratio λ/μ is 0.1, and resulting stand-alone availability $\mu/(\lambda + \mu)$ is 91%. The table also gives the corresponding values for the mean operating time between failures MTBF ($= 1/\lambda$) and mean time to repair MTTR ($= 1/\mu$), expressed in minutes.

In all the simulation experiments, the runs were performed with zero initial occupancy of all buffers and 10 000 time units for the warm-up period. The next 100 000 time units of stationary regime results were used to statistically evaluate the throughput, with 20 replications for each case. The resulting throughput estimates have 95% confidence intervals that are consistently within ± 0.07 jobs per hour for short frequent failures, and within ± 0.2 jobs per hour for long infrequent failures.

Figures 4.1 and 4.2 show the fit of the M -station formula (3.6) to numerical simulation data, for short frequent failures and long infrequent failures, respectively. The continuous curves show throughput from the formula and the data points show the numerical simulation results. Both are plotted against line length M from 2–20 stations, for different buffer sizes B . The comparisons show close agreement between the formula and simulation results. For all the results in Figures 4.1 and 4.2, the formula is within 6% of the simulation results.

Differences between the plots in these two figures show that increasing buffer size has a much bigger improvement in throughput for short frequent failures (Figure 4.1) than for long infrequent failures (Figure 4.2). As pointed out earlier, it is not buffer size B in isolation that is critical in determining how buffers improve throughput, but rather buffer size in relation to station speed S and downtime ($1/\mu$). The formula shows that it is the ratio $B\mu/S$ that is the measure of buffer effectiveness for maintaining throughput. Thus, for fixed station speed S and for the repair rates used here ($\mu = 3$ and 0.3), the throughput curves in Figures 4.1 and 4.2 show that a buffer size $B = 1$ in the case of short frequent failures is equivalent to a buffer size $B = 10$ in the case of long infrequent failures.

Figure 4.3 displays the comparison between the formula and simulation data, with the throughput plotted against buffer size B , for a fixed line length of $M = 20$ stations and for the two reliability scenarios, short frequent failures and long infrequent failures. As

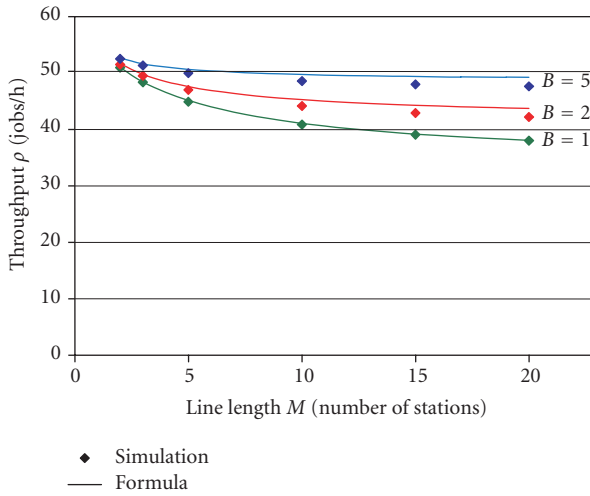


Figure 4.1. Fit of formula (3.6) to simulation results for short frequent failures.

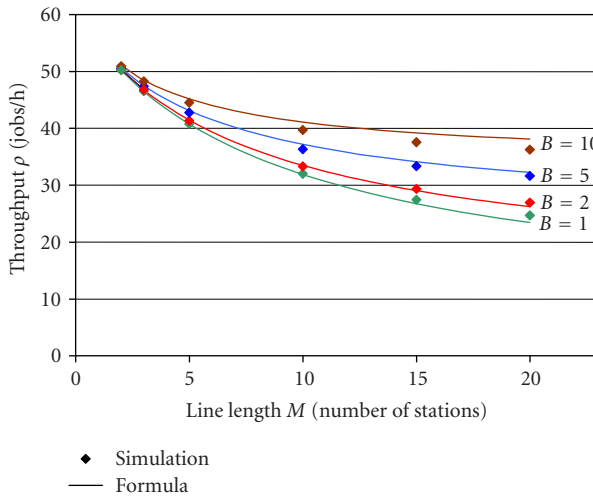


Figure 4.2. Fit of formula (3.6) to simulation results for long infrequent failures.

in the earlier figures, the continuous curves show the formula results and the data points the simulation values. The plots show again that the impact of increasing buffer size on improving throughput depends on $B\mu/S$ and is therefore much greater for short frequent failures than for long infrequent failures.

Tables 4.2 and 4.3 list the complete set of numerical results for the comparisons plotted in the above figures, for the cases of short frequent failures and long infrequent failures, respectively. The tables display the throughput results from the formula and from

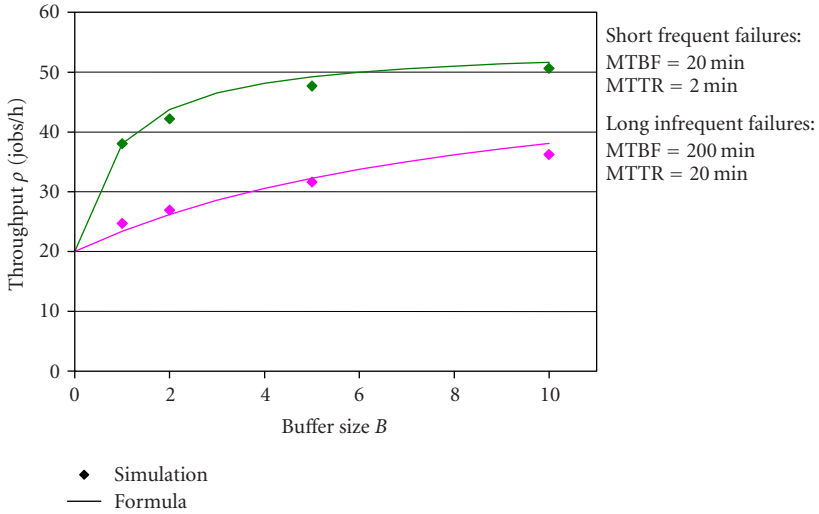


Figure 4.3. Fit of formula (3.6) to simulation results for line length $M = 20$ stations.

the simulation using Simul8 software, and the relative difference between the two sets of results.

The above comparisons indicate that the formula given by (3.6) provides a sufficiently accurate estimate of throughput for a wide range of parameter values. The agreement holds for lines with different buffer sizes. Note that, for a line with no buffers ($B = 0$), the formula reduces to the exact result by Buzacott [4], as given in (3.2). The line in this case is a synchronous system (i.e., all jobs move together in unison as a transfer line), while for cases of nonzero buffers, the line is modeled as an asynchronous system (i.e., jobs move independently of one another).

5. Discussion of throughput formula

Equation (3.6) can be used to determine the buffer size needed for a given system performance target. Rewriting (3.6), we obtain the necessary buffer size B to achieve a desired throughput ρ , that is,

$$B = \frac{4S\{((M - 1)\rho\lambda/\mu)/(S - \rho(1 + \lambda/\mu)) - 1\}}{M\mu(1 + 2\lambda/\mu)}. \tag{5.1}$$

In terms of buffer effectiveness ratio $B\mu/S$, we obtain

$$\frac{B\mu}{S} = \frac{4\{((M - 1)\rho\lambda/\mu)/(S - \rho(1 + \lambda/\mu)) - 1\}}{M(1 + 2\lambda/\mu)}. \tag{5.2}$$

This result shows that the effective buffer level $B\mu/S$ needed to reach a desired throughput ρ depends only on S , ρ , M , and λ/μ . Similar results are also obtained in [9] and a closed-form formula is presented to evaluate the smallest level of buffering (i.e., lean) necessary to achieve desired line efficiency.

Table 4.2. Comparison of formula and simulation throughput results for short frequent failures ($S = 60$ jobs/h, MTBF = 20 minutes, MTTR = 2 minutes).

Buffer size B (no. of jobs)	Line length M (no. of stations)	Simulation ρ_s (jobs/h)	Formula ρ_f (jobs/h)	Relative difference $100 \times (\rho_f - \rho_s) / \rho_s$ (%)
1	2	50.96	50.98	0.04
1	3	48.38	48.47	0.19
1	5	44.93	45.16	0.51
1	10	40.86	41.10	0.59
1	15	39.07	39.20	0.33
1	20	38.03	38.10	0.18
2	2	51.54	51.61	0.14
2	3	49.54	49.78	0.48
2	5	47.02	47.62	1.28
2	10	44.16	45.28	2.54
2	15	42.92	44.30	3.22
2	20	42.21	43.75	3.65
5	2	52.53	52.63	0.19
5	3	51.36	51.66	0.58
5	5	49.99	50.67	1.36
5	10	48.59	49.76	2.41
5	15	48.01	49.41	2.92
5	20	47.66	49.23	3.29
10	2	53.24	53.33	0.17
10	3	52.57	52.80	0.44
10	5	51.82	52.31	0.95
10	10	51.11	51.89	1.53
10	15	50.83	51.74	1.79
10	20	50.66	51.67	1.99
20	2	53.76	53.85	0.17
20	3	53.41	53.46	0.09
20	5	53.03	53.33	0.57
20	10	52.71	53.14	0.82
20	15	52.59	53.08	0.93
20	20	52.53	53.04	0.97

For the parameter values shown in Table 4.1, and a line length $M = 10$ stations, the buffer size B (rounded to the next integer) needed to achieve a target throughput of 40 jobs per hour is $B = 1$ in the case of short frequent failures and $B = 9$ in the case of long infrequent failures. In either case, the effective buffer level $B\mu/S$ is 0.42, that is, the buffer must be sufficient to accommodate the number of jobs produced during 0.42 average downtimes.

Table 4.3. Comparison of formula and simulation throughput results for long infrequent failures ($S = 60$ jobs/h, MTBF = 200 minutes, MTTR = 20 minutes).

Buffer size B (no. of jobs)	Line length M (no. of stations)	Simulation ρ_s (jobs/h)	Formula ρ_f (jobs/h)	Relative difference $100 \times (\rho_f - \rho_s) / \rho_s$ (%)
1	2	50.25	50.12	-0.26
1	3	46.51	46.46	-0.11
1	5	40.77	40.76	-0.02
1	10	31.99	31.87	-0.38
1	15	27.45	26.75	-2.55
1	20	24.70	23.42	-5.18
2	2	50.25	50.24	-0.02
2	3	46.76	46.75	-0.02
2	5	41.33	41.44	0.27
2	10	33.31	33.48	0.51
2	15	29.33	29.05	-0.95
2	20	26.93	26.23	-2.60
5	2	50.52	50.55	0.06
5	3	47.39	47.50	0.23
5	5	42.76	43.14	0.89
5	10	36.33	37.17	2.31
5	15	33.34	34.11	2.31
5	20	31.62	32.26	2.02
10	2	50.91	50.98	0.14
10	3	48.25	48.47	0.46
10	5	44.51	45.16	1.46
10	10	39.71	41.10	3.50
10	15	37.53	39.20	4.45
10	20	36.24	38.10	5.13
20	2	51.52	51.61	0.17
20	3	49.47	49.78	0.63
20	5	46.79	47.62	1.77
20	10	43.54	45.28	4.00
20	15	42.21	44.30	4.95
20	20	41.39	43.75	5.70

The curves in Figures 4.1 and 4.2 show that, for a given buffer size B , throughput decreases as line length M increases. The more stations in the line, the more system interactions occur due to blocking and starving. However, the curves flatten out as M increases, indicating that the additional impact of these interactions diminishes with each additional station. In the limit as M tends to infinity, the throughput approaches a minimum value that is greater than zero, provided the buffer size B is one or greater. Thus, for a line

where $B \geq 1$, throughput ρ satisfies the property

$$\lim_{M \rightarrow \infty} \rho > 0. \quad (5.3)$$

The result given in (5.3) can be deduced by means of contradiction. Consider a line of any length M , with nonzero buffers. If the throughput ρ is zero, then the last station (station M) does not process any jobs. Since the last station is never blocked (assumption (e) in Section 2), this implies that last station is always starved, that is, the buffer before the last station is always empty. It follows that next-to-last station (station $M-1$) cannot be blocked and is always starved, that is, the buffer before station $M-1$ is always empty. Repeating this argument up the line to the second station (station 2), we obtain that station 2 is always starved and the buffer before station 2 (i.e., the first buffer) is always empty. Therefore, the first station cannot be blocked. Since the first station is never starved (assumption (e)), it is able to process jobs to fill the buffer before station 2. This contradicts that this buffer is always empty. For any line length M , therefore, throughput ρ cannot be zero. Thus, we obtain $\rho > 0$ when $M \rightarrow \infty$.

In the special case of no buffers ($B = 0$), the above argument does not hold, since the system reduces to a transfer line in which a failure at any station stops the entire line. Thus, for $B = 0$, the probability that the line is down approaches 1 as the number of stations M tends to infinity, and the throughput therefore approaches zero.

Using (3.6) and letting line length $M \rightarrow \infty$, system throughput ρ approaches the following result:

$$\rho \rightarrow \frac{S}{1 + \lambda/\mu + (\lambda/\mu)/\{(1/4)(1 + 2\lambda/\mu)(B\mu/S)\}}. \quad (5.4)$$

As discussed above, this asymptotic result is nonzero for any buffer size $B \geq 1$, and is zero only in the special case $B = 0$.

Note that the formula derived in this paper applies to a line in which stations have fixed processing times ($1/S$) and random failures. This type of line is typical for automated operations. For lines with manual operations, stations typically have processing times that vary randomly from job to job, but are not subject to machine failures. Analytical results for throughput of a line with variable processing times have been presented by several authors (Haydon [14]; Buzacott [5]; Blumenfeld [3]; De Kok [8]; Askin and Standridge [2]; Martin [19]; Liu et al. [18]).

6. Conclusion

This paper presents an analytical formula for the throughput of a serial production line, in which workstations are subject to random failures. The formula is derived from the equations developed in [1] for a two-station line and extended to give a close approximation for a line of any length. It gives an estimate of throughput in the special case of identical stations and equally sized buffers. This special case result allows quick and convenient “back of the envelope” calculations with minimal data. Such calculations are valuable in initial design stage of a new system, when detailed data are not available. The formula has a simple analytical form, so that the mathematical relationships between

the system parameters are readily apparent. This is important in gaining insight into system behavior and in assessing the benefits of system improvements.

With the recent development of web-based technology, the formula has an additional benefit. It is a closed-form function that does not involve any iterative procedures. The formula is therefore ideally suited to being programmed in a Java applet for interactive analyses on a graphical display, so that the impact on throughput of any parameter changes can be seen virtually instantaneously.

In summary, the formula provides a handy way of assessing throughput performance as a first step in production system design. It allows quick comparisons between design alternatives before more detailed analyses are conducted. Initial approximate estimates of throughput are useful in decision-making, where it is necessary to consider tradeoffs between throughput and other performance criteria, such as operating costs and investment planning. Since the formula is not limited to a particular programming language or operating system, it can be used on its own for throughput calculations or be included as one component of a comprehensive decision-making tool for planning purposes.

7. Acknowledgment

Our special thanks go to Dr. Jeff Alden, General Motors R&D, for many valuable discussions on his analytical model for throughput analysis.

References

- [1] J. M. Alden, *Estimating performance of two workstations in series with downtime and unequal speeds*, Research Publication R&D-9434, General Motors R&D Center, Warren, Michigan, 2002.
- [2] R. G. Askin and C. R. Standridge, *Modeling and Analysis of Manufacturing Systems*, Wiley, New York, 1993.
- [3] D. E. Blumenfeld, *A simple formula for estimating throughput of serial production lines with variable processing times and limited buffer capacity*, *International Journal of Production Systems* **28** (1990), 1163–1182.
- [4] J. A. Buzacott, *Prediction of the efficiency of production systems without internal storage*, *International Journal of Production Systems* **6** (1968), no. 3, 173–188.
- [5] ———, *Abandoning the moving assembly line: models of human operators and job sequencing*, *International Journal of Production Systems* **28** (1990), no. 5, 821–839.
- [6] J. A. Buzacott and J. G. Shanthikumar, *Stochastic Models of Manufacturing Systems*, Prentice-Hall, New Jersey, 1993.
- [7] S. Y. Chiang, C. T. Kuo, and S. M. Meerkov, *DT-bottlenecks in serial production lines: theory and application*, *IEEE Transactions on Robotics and Automation* **16** (2000), no. 5, 567–580.
- [8] A. G. De Kok, *Computationally efficient approximations for balanced flowlines with finite intermediate buffers*, *International Journal of Production Systems* **28** (1990), no. 2, 401–419.
- [9] E. Enginarlar, J. Li, and S. M. Meerkov, *How lean can lean buffers be?* *IIE Transactions* **37**, 333–342.
- [10] E. Enginarlar, J. Li, S. M. Meerkov, and R. Q. Zhang, *Buffer capacity for accommodating machine downtime in serial production lines*, *International Journal of Production Research* **40** (2002), no. 3, 601–624.
- [11] S. B. Gershwin, *An Efficient decomposition algorithm for the approximate evaluation of tandem queues with finite storage space and blocking*, *Operations Research* **35** (1987), no. 2, 291–305.

- [12] ———, *Manufacturing Systems Engineering*, PTR Prentice-Hall, New Jersey, 1994.
- [13] J. W. Hauge and K. N. Paige, *Learning simul8, The Complete Guide*, Plain Vu Publishers, Bellingham, Washington, 2001.
- [14] B. J. Haydon, *The Behaviour of Systems of Finite Queues*, Ph.D. thesis, University of New South Wales, Kensington, 1975.
- [15] D. Jacobs and S. M. Meerkov, *Mathematical theory of improvability of production systems*, *Mathematical Problems in Engineering*, vol. 1, 1995, pp. 95–137.
- [16] J. Li, D. E. Blumenfeld, and J. M. Alden, *Throughput analysis of serial production lines with two machines: summary and comparisons.*, Research Publication R&D-9532, General Motors R&D Center, Warren, Michigan, 2003.
- [17] J. Li and S. M. Meerkov, *Due-time performance in production systems with Markovian machines*, *Analysis and Modeling of Manufacturing Systems*, Kluwer Academic Publishers, Massachusetts, 2003, pp. 221–253.
- [18] C. M. Liu, S. F. Su, and C. L. Lin, *Predictive models for performance of serial production lines*, *International Journal of Production Research* **34** (1996), no. 3, 1279–1291.
- [19] G. E. Martin, *Predictive formulae for unpaced line efficiency*, *International Journal of Production Research* **31** (1993), no. 8, 1981–1990.

Dennis E. Blumenfeld: Manufacturing Systems Research Lab, R&D Center, General Motors, Mail Code 480-106-359, 30500 Mound Road, Warren, MI 48090-9055, USA

E-mail address: dennis.blumenfeld@gm.com

Jingshan Li: Manufacturing Systems Research Lab, R&D Center, General Motors, Mail Code 480-106-359, 30500 Mound Road, Warren, MI 48090-9055, USA

E-mail address: jingshan.li@gm.com