34

Modeling and Analysis of Manufacturing Systems

E. Lefeber
Eindhoven University of Technology

J.E. Rooda
Eindhoven University of Technology

34.1 Introduction

The dynamics of manufacturing systems has been a subject of study for several decades (Forrester, 1961; Hopp and Spearman, 2000). Over the last years, manufacturing systems have become more and more complex and therefore a good understanding of their dynamics has become even more important.

The goal of this chapter is to introduce a large variety of models for manufacturing systems. By means of examples it is illustrated how certain modeling techniques can be used to derive models that can be used for analysis or control. In addition to references that can be used as a starting point for further inquiry, recent developments in the modeling, analysis, and control of manufacturing systems are presented.

Since no familiarity with manufacturing systems is assumed, in Section 34.2 some terminology and basic properties of manufacturing systems are introduced. Section 34.3 provides some analytical modeling techniques and methods for analyzing steady-state behavior. Section 34.4 is concerned with deriving discrete-event models, which yield a more detailed insight in the dynamics of a manufacturing system. To reduce the complexity of discrete-event models, the concept of effective process times (EPTs) is introduced in Section 34.5, which results in modeling a manufacturing system as a large queueing network. This way of modeling a manufacturing system is a first step in a larger control framework, which is introduced in Section 34.6. This control framework makes it possible to study problems of controlling the dynamics of manufacturing systems by means of the available inputs. An important role in this control framework is played by approximation models. The most commonly used approximation models are presented in Section 34.7. Recently, a new class of approximation models has been proposed, which is presented in Section 34.8. Section 34.9 concludes this chapter.

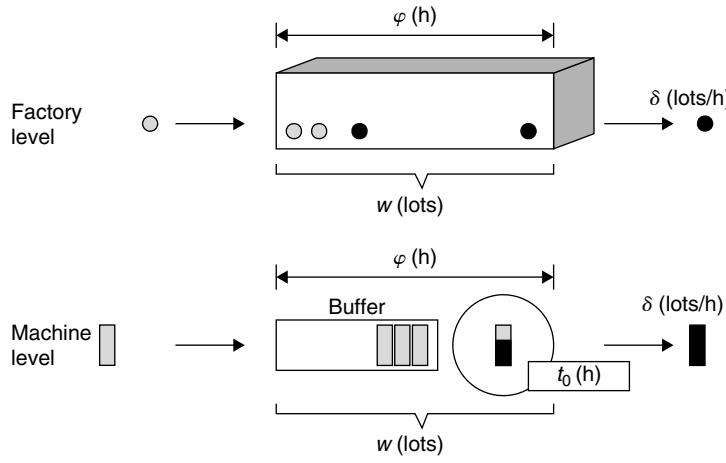


FIGURE 34.1 Basic quantities for manufacturing systems.

34.2 Preliminaries

First a few basic quantities are introduced as well as the main principles for manufacturing system analysis. The items produced by a manufacturing system are called *lots*. Also the words product and job are commonly used. Other important notions are throughput, flow time, wip, and utilization. These notions are illustrated in Figure 34.1 at factory and machine level.

Raw process time t_0 of a lot denotes the net time a machine needs to process the lot. This process time excludes additions such as setup time, breakdown, or other sources that may increase the time a lot spends in the machine. The raw process time is typically measured in hours or minutes.

Throughput δ denotes the number of lots per unit time that leaves the manufacturing system. At a machine level, this denotes the number of lots that leave a machine per unit time. At a factory level it denotes the number of lots that leave the factory per unit time. The unit of throughput is typically lots/hour.

Flow time φ denotes the time a lot is in the manufacturing system. At a factory level this is the time from the release of the lot into the factory until the finished lot leaves the factory. At a machine level this is the time from entering the machine (or the buffer in front of the machine) until leaving the machine. Flow time is typically measured in days, hours, or minutes. Instead of flow time the words cycle time and throughput time are also commonly used.

Work in process (wip) w denotes the total number of lots in the manufacturing system, i.e., in the factory or in the machine. Wip is measured in lots.

Utilization u denotes the fraction a machine is not idle. A machine is considered idle if it could start processing a new lot. Thus process time as well as downtime, setup time, and preventive maintenance time all contribute to the utilization. Utilization has no dimension and can never exceed 1.0.

Ideally, a manufacturing system should both have a high throughput and a low flow time or low wip. Unfortunately, these goals are conflicting (cf. Figure 34.2) and both cannot be met simultaneously. If a high throughput is required, machines should always be busy. As from time to time disturbances like machine failures happen, buffers between two consecutive machines are required to make sure that the second machine can still continue if the first machine fails (or vice versa). Therefore, for a high throughput many lots are needed in the manufacturing system, i.e., wip needs to be high. As a result, if a new lot starts in the system it has a large flow time, since all lots that are currently in the system need to be completed first.

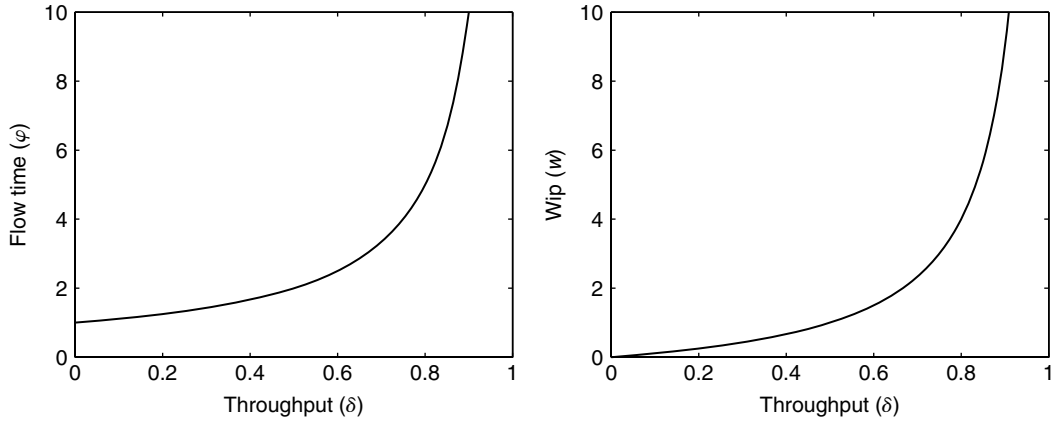


FIGURE 34.2 Basic relations between basic quantities for manufacturing systems.

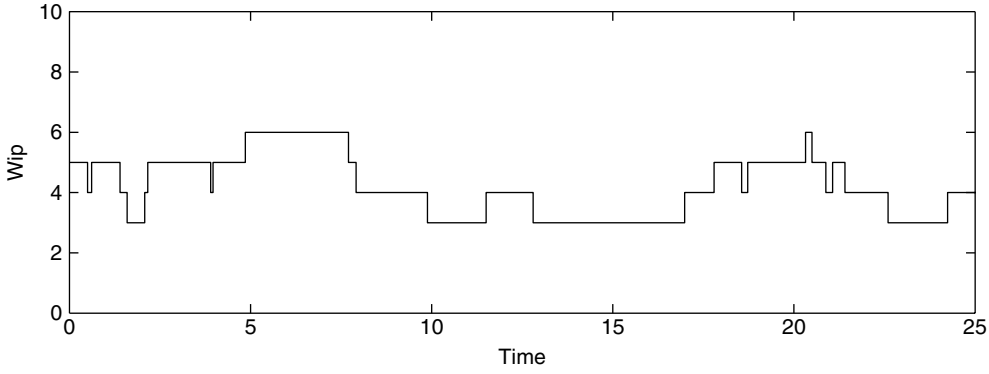


FIGURE 34.3 A characteristic time-behavior of wip at a workstation.

Conversely, the least possible flow time can be achieved if a lot arrives at a completely empty system and never has to wait before processing takes place. As a result, the wip level is small. However, for most of the time machines are not processing, yielding a small throughput.

When trying to control manufacturing systems, a trade-off needs to be made between throughput and flow time, so the nonlinear (steady-state) relations depicted in Figure 34.2 need to be incorporated in any reasonable model of manufacturing systems.

A final observation of relevance for modeling manufacturing systems is the nature of the system signals. In Figure 34.3 a characteristic graph of the wip at a workstation as a function of time is shown. Wip always takes integer values with arbitrary (nonnegative real) duration. One could consider a manufacturing system to be a system that takes values from a finite set of states and jumps from one state to the other as time evolves. This jump from one state to the other is called an *event*. As we have a countable (discrete) number of states, it is clear that discrete-event models are often used in modeling manufacturing systems. Discrete-event models for manufacturing systems are considered in Section 34.4. But first some analytical models for analyzing steady-state behavior of manufacturing systems are presented in the next section.

34.3 Analytical Models for Steady-State Analysis

To get some insights in the steady-state performance of a given manufacturing system simple relations can be used. In this section, we deal with mass conservation for determining the mean utilization of

workstations and the number of machines required for meeting a required throughput. Furthermore, relations from queueing theory are used to obtain estimates for the mean wip and mean flow time.

34.3.1 Mass Conservation (Throughput)

Using mass conservation the mean utilization of workstations can easily be determined.

Example 1

Consider the manufacturing system with rework and bypassing in Figure 34.4. The manufacturing system consists of three buffers and four machines. Lots are released at a rate of λ lots/h. The numbers near the arrows indicate the fraction of the lots that follow that route. For instance, of the lots leaving buffer B_1 90% goes to machine M_1 and 10% goes to buffer B_3. The process time of each machine is listed in the table in Figure 34.4.

Let δ_{M_i} and δ_{B_i} denote the throughput of machine M_i $(i = 1, 2, 3, 4)$ and buffer B_i $(i = 1, 2, 3)$, respectively. Using mass conservation we obtain

$$\begin{aligned} \delta_{M_1} &= 0.9\delta_{B_1} & \delta_{B_1} &= \lambda \\ \delta_{M_2} &= 0.2\delta_{B_2} & \delta_{B_2} &= \delta_{M_1} + \delta_{M_2} \\ \delta_{M_3} &= 0.8\delta_{B_2} & \delta_{B_3} &= \delta_{M_3} + 0.1\delta_{B_1} \\ \delta_{M_4} &= \delta_{B_3} & \delta &= \delta_{M_4} \end{aligned}$$

Solving these linear relations results in:

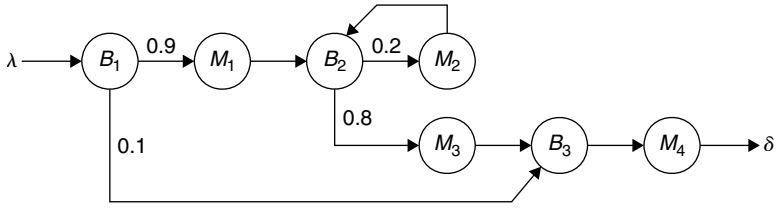
$$\begin{aligned} \delta_{M_1} &= 0.9\lambda & \delta_{B_1} &= \lambda \\ \delta_{M_2} &= 0.225\lambda & \delta_{B_2} &= 1.125\lambda \\ \delta_{M_3} &= 0.9\lambda & \delta_{B_3} &= \lambda \\ \delta_{M_4} &= \lambda & \delta &= \lambda \end{aligned}$$

Using the process times of the table in Figure 34.4, we obtain for the utilizations:

$$\begin{aligned} u_{M_1} &= 0.9\lambda \cdot 2.0/1 = 1.8\lambda & u_{M_3} &= 0.9\lambda \cdot 1.8/1 = 1.62\lambda \\ u_{M_2} &= 0.225\lambda \cdot 6.0/1 = 1.35\lambda & u_{M_4} &= \lambda \cdot 1.6/1 = 1.6\lambda \end{aligned}$$

Clearly, machine M_1 is the bottleneck and the maximal throughput for this line is $\lambda = 1/1.8 = 0.56$ jobs/h.

Using mass conservation, utilizations of workstations can be determined straightforwardly. This also provides a way for determining the number of machines required for meeting a given throughput. By modifying the given percentages the effect of rework or a change in product mix can also be studied.



Machine	t_0 (h)
M_1	2.0
M_2	6.0
M_3	1.8
M_4	1.6

FIGURE 34.4 Manufacturing system with rework and bypassing.

34.3.2 Queueing Relations (Wip, Flow Time)

For determining a rough estimate of the corresponding mean flow time and mean wip, basic relations from queueing theory can be used.

Consider a single machine workstation that consists of infinite buffer B_∞ and machine M (see Figure 34.5). Lots arrive at the buffer with a stochastic interarrival time. The interarrival time distribution has mean t_a and a standard deviation σ_a, which we characterize by the coefficient of variation $c_a = \sigma_a/\mu_a$. The machine has stochastic process times, with mean process time t_0 and coefficient of variation c_0. Finished lots leave the machine with a stochastic interdeparture time, with mean t_d and coefficient of variation c_d. Assuming independent interarrival times and independent process times, the mean waiting time φ_B in buffer B can be approximated for a stable system by means of Kingman's equation (Kingman, 1961):

$$\varphi_B = \frac{c_a^2 + c_0^2}{2} \frac{u}{1-u} t_0 \tag{34.1}$$

with the utilization u defined by: $u = t_0/t_a$. Eq. (34.1) is exact for an $M/G/1$ system, i.e., a single machine workstation with exponentially distributed interarrival times and any distribution for the process time. For other single machine workstations it is an approximation.

For a stable system, we have $t_d = t_a$. We can approximate the coefficient of variation c_d by Kuehn's linking equation (Kuehn, 1979):

$$c_d^2 = (1-u^2)c_a^2 + u^2 c_0^2. \tag{34.2}$$

This result is exact for an $M/M/1$ system. For other single machine workstations it is an approximation. Having characterized the departure process of a workstation, the arrival process at the next workstation has been characterized as well. As a result, a line of workstations can also be described.

Example 2 (Three workstations in line)

Consider the three workstation flow line in Figure 34.6. For the interarrival time at workstation 0 we have $t_a = 4.0$ h and $c_a^2 = 1$. The three workstations are identical with respect to the process times: $t_{0,i} = 3.0$ h for $i = 0, 1, 2$ and $c_{0,i}^2 = 0.5$ for $i = 0, 1, 2$. We are interested to determine the mean total flow time per lot.

Since $t_a > t_{0,i}$ for $i = 0, 1, 2$, we have a stable system and $t_{a,i} = t_{d,i} = 4.0$ h for $i = 0, 1, 2$. Subsequently, the utilization for each workstation is $u_i = 3.0/4.0 = 0.75$ for $i = 0, 1, 2$.

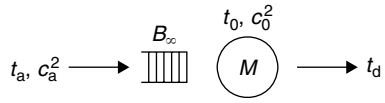


FIGURE 34.5 Single-machine workstation.

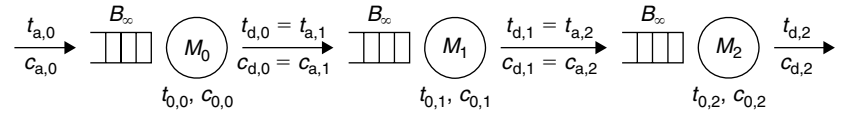


FIGURE 34.6 Three workstation flow line.

Using Eq. (34.1) we calculate the mean flow time for workstation 0

$$\varphi_0 = \varphi_B + t_0 = \frac{c_a^2 + c_0^2}{2}\frac{u}{1-u}t_0 + t_0 = \frac{1+0.5}{2}\frac{0.75}{1-0.75}3.0 + 3.0 = 9.75\,\text{h}$$

Using Eq. (34.2), we determine the coefficient of variation on the interarrival time $c_{a,1}$ for workstation W_1

$$c_{a,1}^2 = c_{d,0}^2 = (1-u^2)c_a^2 + u^2 c_0^2 = (1-0.75^2)1 + 0.75^2 \cdot 0.5 = 0.719$$

and the mean flow time for workstation 1

$$\varphi_1 = \frac{0.719+0.5}{2}\frac{0.75}{1-0.75}3.0 + 3.0 = 8.49\,\text{h}$$

In a similar way, we determine that $c_{a,2}^2 = 0.596$, $\varphi_2 = 7.93$ h. We then calculate the mean total flow time to be

$$\varphi_{\text{tot}} = \varphi_0 + \varphi_1 + \varphi_2 = 26.2\,\text{h}$$

Note that the minimal flow time without variability ($c_a^2 = c_{0,i}^2 = 0$) equals 9.0 h.

Eq. (34.1) and Eq. (34.2) are particular instances of a workstation consisting of a single machine. For workstations consisting of m identical machines, in parallel the following approximations can be used:

$$\varphi_B = \frac{c_a^2 + c_0^2}{2}\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \cdot t_0 \tag{34.3}$$

$$c_d^2 = (1-u^2)c_a^2 + u^2\frac{c_0^2 + \sqrt{m} - 1}{\sqrt{m}} \tag{34.4}$$

Note that in case $m = 1$ these equations reduce to Eq. (34.1) and Eq. (34.2).

Once the mean flow time has been determined, a third basic relation from queueing theory, Little's law (Little, 1961), can be used for determining the mean wip level. Little's law states that the mean wip level (number of lots in a manufacturing system) w is equal to the product of the mean throughput δ and the mean flow time φ, provided the system is in steady state

$$w = \delta\varphi \tag{34.5}$$

An example illustrates how Kingman's equation and Little's law can be used.

Example 3

Consider the system of Example 2 as depicted in Figure 34.6. From Example 34.3.2 we know that the flow times for the three workstations are, respectively,

$$\varphi_0 = 9.75\,\text{h}, \qquad \varphi_1 = 8.49\,\text{h}, \qquad \varphi_2 = 7.93\,\text{h}$$

Since the steady-state throughput was assumed to be $\delta = 1/t_a = 1/4.0 = 0.25$ lots/h, we obtain via Little's law

$$w_0 = 0.25 \cdot 9.75 = 2.44\ \text{lots}$$
$$w_1 = 0.25 \cdot 8.49 = 2.12\ \text{lots}$$
$$w_2 = 0.25 \cdot 7.93 = 1.98\ \text{lots}$$

The above-mentioned relations are simple approximations that can be used for getting a rough idea about the possible performance of a manufacturing system. These approximations are fairly accurate for high

degrees of utilization but less accurate for lower degrees of utilization. A basic assumption when using these approximations is the independence of the interarrival times, which in general is not the case, e.g., for merging streams of jobs. Furthermore, using these equations only steady-state behavior can be analyzed. For studying things like ramp-up behavior or for incorporating more details like operator-behavior, more sophisticated models are needed, as described in the next section.

34.4 Discrete-Event Models

In the previous section simple methods have been introduced for analyzing steady-state behavior of manufacturing systems. For analyzing the dynamics of manufacturing systems, more sophisticated models are required. Using Figure 34.3 in Section 34.2 it was illustrated that typical models of manufacturing systems are the so-called discrete-event models. In this section, we present examples how to build a discrete-event model of a manufacturing system using the specification language χ, explained in more detail in Chapter 19 of this handbook.

The way to build a discrete-event model is to consider the manufacturing system as a network of concurrent processes through which jobs and other types of information flows. For example, a basic machine can be modeled as a process which repeatedly tries to receive a lot, waits for the period of time (the process time), and tries to send a lot. Using χ, we can write

$$\texttt{proc}\ M\,(\texttt{chan}\ a?b!:\text{lot},\ \texttt{var}\ t_e:\texttt{real}) = |[\texttt{var}\ x:\text{lot} :: (a?x;\ \Delta t_e;\ b!x)]|$$

The machine is able to receive lots via external channel a, is able to send lots via external channel b, and the process time of the machine is given by parameter t_e. Repeatedly, the machine tries to receive a lot over external channel a and store this lot in discrete variable x. Next, the machine waits for t_e, after which the machine tries to send x via external channel b.

A buffer can be modeled as a process that may receive new lots if it is not full and may send lots if it is not empty. Using χ, a finite first in, first out (FIFO) buffer with a maximal buffer size n can be modeled as

$$\begin{array}{l}\texttt{proc}\ B\ (\texttt{chan}\ a?b!:\text{lot},\ \texttt{var}\ n:\texttt{nat}) =\\ |[\ \texttt{disc}\ xs:[\text{lot}]=[\,],\ x:\text{lot}\\ \ ::(\text{len}(xs)<n \rightarrow a?x;\ xs:=xs{+}{+}[x]\\ \ \ []\ \text{len}(xs)>0 \rightarrow b!\text{hd}(xs);\ xs:=\text{tl}(xs)\\ \ \ )\\ ]|\end{array}$$

This process can receive lots via external channel a, send lots via external channel b, and has its maximal buffer size n as a parameter. Repeatedly, two alternatives can be executed:

- Trying to receive a lot via channel a (only if the length of the list xs is less than n) into discrete variable x and consecutively adding it to list xs of lots (using a concatenation of lists).
- Trying to send the head of the list (its first element) via channel b (only if the list is not empty) and consecutively reducing list xs to its tail (everything but the first element).

These two processes can be used to model a workstation that consists of a 3-place buffer and a machine with process time t_e by simply executing the two previously specified processes in parallel:

$$\texttt{proc}\ W(\texttt{chan}\ a?b!:\text{lot},\ \texttt{var}\ t_e:\texttt{real}) = |[\texttt{chan}\ c:\text{lot} :: B(a,c,3)\|M(c,b,t_e)]|$$

We assume that lots arrive at this workstation according to a Poisson arrival process with mean arrival rate of λ jobs per unit time. This can be modeled by means of the generator process

$$\begin{array}{l}\texttt{proc}\ G\ (\texttt{chan}\ a!:\text{lot},\ \texttt{var}\ \lambda:\texttt{real}) =\\ |[\texttt{disc}\ u:\rightarrow\ \texttt{real}=\text{exponential}(1/\lambda) :: (a!\tau;\ \Delta\sigma u;\ b!x)]|\end{array}$$

Here the type lot is a real number which contains the time this lot entered the system. Generator G is able to send lots via external channel a and has a mean departure rate, which is given by parameter λ. The

discrete variable u contains an exponential distribution with mean $1/\lambda$. Repeatedly, the generator tries to send a lot over external channel a, where at departure it gets assigned the current time τ. Next, the generator waits for a period which is given by a sample from the distribution u.

Once a lot has been served by workstation W it leaves to the exit process E:

$$\texttt{proc}\ E(\texttt{chan}\ a? : \text{lot}) = |[\texttt{var}\ x : \text{lot} :: (a?x)]|$$

This process repeatedly tries to receive a lot via external channel a.

For an arrival rate of $\lambda = 0.5$ and a process time of $t_e = 1.5$, the specification of the discrete-event model can be completed by

$$\texttt{model}\ GWE() = |[\texttt{chan}\ a,b : \text{lot} :: G(a,0.5) \parallel W(a,b,1.5) \parallel E(\text{b})]|$$

In this way a manufacturing system can be modeled as a network of concurrent processes through which jobs and other types of information flows. The presented model is rather simple, but clearly many more ingredients can be added. For example, to include an operator for the processing of the machine we can modify the process M into

$$\texttt{proc}\ \bar{M}\ (\texttt{chan}\ a?b! : \text{lot}, c?, d! : \text{operator}, \texttt{var}\ t_e : \texttt{real}) =$$
$$|[\texttt{var}\ x : \text{lot}, y : \text{operator} :: (c?y; a?x; \Delta t_e; b!x; d!y)]|$$

Highly detailed models of manufacturing systems can be made in this way, even before the system has been build. The influence of parameters can be analyzed by running several experiments with the discrete-event model using different parameter settings. This is common practice when designing a several billion wafer fab. However, since in practice manufacturing systems are changing continuously, it is very hard to keep these detailed discrete-event models up-to-date.

Fortunately, for a manufacturing system in operation it is possible to arrive at more simple/less detailed discrete-event models by using the concept of EPTs as introduced in the next section.

34.5 Effective Process Times

As mentioned in the previous section, for the processing of a lot at a machine, many steps may be required. It could be that an operator needs to get the lot from a storage device, set up a specific tool that is required for processing the lot, put the lot on an available machine, start a specific program for processing the lot, wait until this processing has finished (meanwhile doing something else), inspect the lot to determine if all went well, possibly perform some additional processing (e.g., rework), remove the lot from the machine and put it on another storage device, and transport it to the next machine. At all of these steps something might go wrong: the operator might not be available, after setting up the machine the operator finds out that the required recipe cannot be run on this machine, the machine might fail during processing, no storage device is available anymore so the machine cannot be unloaded and is blocked, etc.

It is impossible to measure all sources of variability that might occur in a manufacturing system. While some of the sources of variability could be incorporated into a discrete-event model (tool failures and repairs, maintenance schedules), not all sources of variability can be included. This is clearly illustrated in Figure 34.7, obtained from Jacobs et al. (2003).

The left graph contains actual realizations of flow times of lots leaving a real manufacturing system, whereas the right graph contains the results of a detailed deterministic simulation model and the graph in the middle contains the results of a similar model including stochasticity. It turns out that in reality flow times are much higher and much more irregular than simulation predicts. So, even if one tries hard to capture all variability present in a manufacturing system, still the outcome predicted by the model is far from reality.

The term EPT has been introduced by Hopp and Spearman (2000) as the time seen by lots from a logistical point of view. To determine the EPT they assume that the contribution of the individual sources of variability is known.

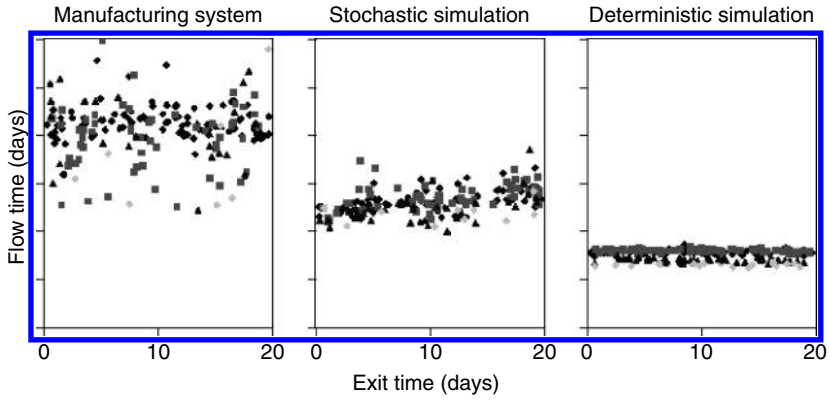


FIGURE 34.7 A comparison.

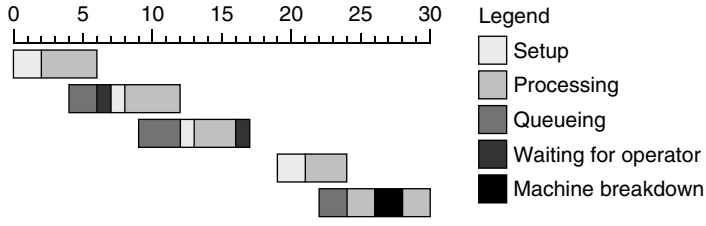


FIGURE 34.8 Gantt chart of five lots at a single machine workstation.

A similar description is given in Sattler (1996) where the effective process time has been defined as all flow time except waiting for another lot. It includes waiting owing to machine down time and operator availability and a variety of other activities. In Sattler (1996) it was also noticed that this definition of effective process time is difficult to measure.

Instead of taking the bottom-up view of Hopp and Spearman, a top-down approach can also be taken, as shown by Jacobs et al. (2003), where algorithms have been introduced that enable determination of effective process time realizations from a list of events.

We consider a single machine workstation and assume that the Gantt chart of Figure 34.8 describes a given time period.

- At $t = 0$ the first lot arrives at the workstation. After a setup, the processing of the lot starts at $t = 2$ and is completed at $t = 6$.
- At $t = 4$ the second lot arrives at the workstation. At $t = 6$ this lot could have been started, but apparently there was no operator available, so only at $t = 7$ the setup for this lot starts. Eventually, at $t = 8$ the processing of the lot starts and is completed at $t = 12$.
- The fifth lot arrives at the workstation at $t = 22$, processing starts at $t = 24$, but at $t = 26$ the machine breaks down. It takes until $t = 28$ before the machine has been repaired and the processing of the fifth lot continues. The processing of the fifth lot is completed at $t = 30$.

From a lot's point of view we observe:

- The first lot arrives at an empty system at $t = 0$ and departs from this system at $t = 6$. Its processing took 6 units of time.
- The second lot arrives at a nonempty system at $t = 4$ and needs to wait. At $t = 6$, the system becomes available again and hence from $t = 6$ on there is no need for the second lot to wait. At $t = 12$ the

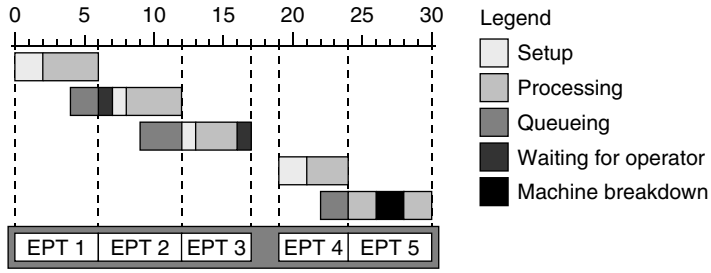


FIGURE 34.9 EPT realizations of five lots at a workstation.

second lot leaves the system, so from the point of view of this lot, its processing took from $t = 6$ till $t = 12$; the lot does not know whether waiting for an operator and a setup is part of its processing.

- The third lot sees no need for waiting after $t = 12$ and leaves the system at $t = 17$, so it assumes to have been processed from $t = 12$ till $t = 17$.

Following this reasoning, the resulting effective process times for lots are as depicted in Figure 34.9. Note that only arrival and departure events of lots to a workstation are needed for determining the effective process times. Furthermore, none of the contributing disturbances needs to be measured. In highly automated manufacturing systems, arrival and departure events of lots are being registered, so for these manufacturing systems, effective process time realizations can be determined rather easily. These EPT realizations can be used in a relatively simple discrete-event model of the manufacturing system. Such a discrete-event model only contains the architecture of the manufacturing system, buffers, and machines. The process times of these machines are samples from their EPT distribution as measured from real manufacturing data. There is no need for incorporating machine failures, operators, etc., as this is all included in the EPT-distributions. Furthermore, the EPTs are utilization independent. That is, EPTs determined collected at a certain throughput rate are also valid for different throughput rates. Also, machines with the same EPT-distribution can be added to a workstation. This makes it possible to study how the manufacturing system responds in case a new machine is added, or all kinds of other what-if-scenarios. Finally, since EPT-realizations characterize operational time variability, they can be used for performance measuring. For more on this issue, the interested reader is referred to Jacobs et al. (2003, 2006). What is most important is that EPTs can be determined from real manufacturing data and yield relatively simple discrete-event models of the manufacturing system under consideration. These relatively simple discrete-event models can serve as a starting point for controlling manufacturing systems dynamically.

34.6 Control of Manufacturing Systems: A Framework

In the previous section, the concept of effective process times has been introduced as a means to arrive at relatively simple discrete-event models for manufacturing systems, using measurements from the real manufacturing system under consideration. The resulting discrete-event models are large queueing networks which capture the dynamics reasonably well. These relatively simple discrete-event models are not only a starting point for analyzing the dynamics of a manufacturing system, but can also be used as a starting point for controller design. If one is able to control the dynamics of the discrete-event model of the manufacturing system, the resulting controller can also be used for controlling the real manufacturing system.

Even though control theory exists for controlling discrete-event systems, unfortunately none of it is appropriate for controlling discrete-event models of real-life manufacturing systems. This is mainly due to the large number of states of a manufacturing system. Therefore, a different approach is needed.

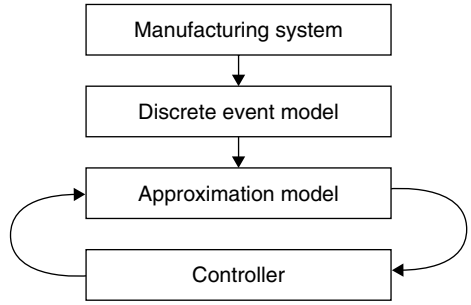


FIGURE 34.10 Control framework (I).

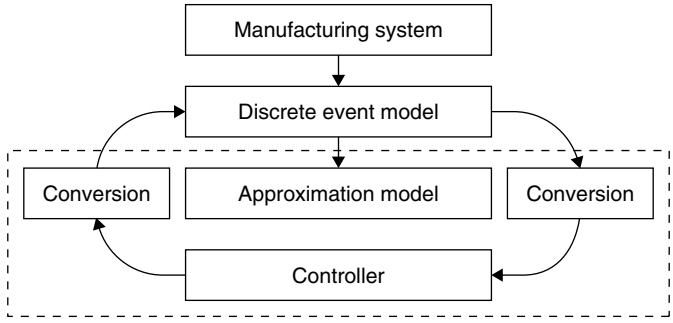


FIGURE 34.11 Control framework (II).

If we concentrate on mass production, the distinction between lots is not really necessary and lots can be viewed in a more continuous way. Therefore, instead of the discrete-event model we consider an approximation model.

Using the approximation model, we can use standard control theory to derive a controller for the approximation model (cf. Figure 34.10).

When the closed-loop system of the approximation model and the controller behaves as desired, the controller can be connected to the discrete-event model. However, since the derived controller is not a discrete-event controller its control actions still need to be transformed into events. It might very well be that the optimal control action would be to produce 2.75 lots during the next shift. One still needs to decide how many jobs to really start (2 or 3), and also when to start them. This is the left conversion block in Figure 34.11. Similarly, a conversion is needed from discrete-event model to controller: a simple conversion would be to sample the discrete-event model once every shift, but other sampling strategies might also be followed. For example, if at the beginning of a shift a machine breaks down it might not be such a good idea to wait until the end of the shift before setting new production targets.

Once the two conversion blocks have been properly designed a suitable discrete-event controller for the discrete-event model is obtained, as illustrated in Figure 34.11 (dashed).

Eventually, as a final step, the designed controller can be disconnected from the discrete-event model, and attached to the real manufacturing system.

In the presented control framework two crucial steps can be distinguished. First, the discrete-event model should be a good enough approximation of the real manufacturing system, i.e., the model needs to be validated and if found unsatisfactory it needs to be improved. Second, the approximation model should be a good enough approximation of the discrete-event model, or actually, of the discrete-event model and conversion block(s), since that is the system that needs to be controlled by the continuous controller. Depending on the variables of interest, a valid approximation model needs to be used. An overview of

common used approximation models, assuming mass production, is provided in the next two sections. In Section 34.7 approximation models are presented that mainly focus on throughput. In Section 34.8 approximation models are presented that incorporate both throughput and flow time, taking into account the nonlinear relations as depicted in Figure 34.2.

34.7 Standard Fluid Model and Extensions

The analytical approximations models of Section 34.3 are only concerned with steady state, no dynamics is included. This disadvantage is overcome by discrete-event models as discussed in Section 34.4. However, since they model each job separately and stochastically long simulation times are required for obtaining satisfactory results. Using an approximation model where jobs are viewed in a continuous way we can overcome these long simulation times.

34.7.1 A Common Fluid Model

The current standard way of deriving fluid models is most easily explained by means of an example. Consider a simple manufacturing system consisting of two machines in series, as displayed in Figure 34.12. Let $u_0(t)$ denote the rate at which jobs arrive at the system at time t, $u_i(t)$ the rate at which machine M_i produces lots at time t, $y_i(t)$ the number of lots in buffer B_i at time t ($i \in \{1, 2\}$), and $y_3(t)$ the number of lots produced by the manufacturing system at time t. Assume that machines M_1 and M_2 have a maximum capacity of μ_1 and μ_2 lots per time unit, respectively.

The rate of change of the buffer contents is given by the difference between the rates at which lots enter and leave the buffer. Under the assumption that the number of lots can be considered continuously, we get

$$\begin{aligned} \dot{y}_1(t) &= u_0(t) - u_1(t), \\ \dot{y}_2(t) &= u_1(t) - u_2(t), \\ \dot{y}_3(t) &= u_2(t) \end{aligned} \tag{34.6}$$

which can be rewritten as

$$\dot{x}(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} u(t) \tag{34.7a}$$

$$y(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} u(t) \tag{34.7b}$$

where $u = [u_0, u_1, u_2]^\top$ and $y = [y_1, y_2, y_3]^\top$. We also have capacity constraints on the input as well as the constraint that the buffer contents should remain positive, i.e.,

$$0 \le u_1(t) \le \mu_1, 0 \le u_2(t) \le \mu_2 \quad \text{and} \quad y_1(t) \ge 0, y_2(t) \ge 0, y_3(t) \ge 0 \tag{34.8}$$

This system is a controllable linear system of the form $\dot{x} = Ax + Bu, y = Cx + Du$, extensively studied in control theory.

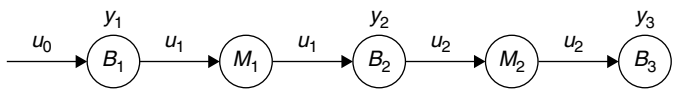


FIGURE 34.12 A simple manufacturing system.

Note that instead of a description in continuous time, a description in discrete time can also be used:

$$\dot{x}(k+1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} u(k) \tag{34.9a}$$

$$y(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} u(k) \tag{34.9b}$$

Also, the description of Eq. (34.7) is not the only possible input/output/state model that yields the input/output behavior Eq. (34.6). To illustrate this, consider the change of coordinates

$$x(t) = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \bar{x}(t) \tag{34.10}$$

which results in the following input/output/state model:

$$\bar{x}(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bar{x}(t) + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u(t) \tag{34.11a}$$

$$y(t) = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \bar{x}(t) + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} u(t) \tag{34.11b}$$

Note that in this description, the state $\bar{x}$ denotes the cumulative production at each workstation.

We would like to study the response of the output of the system Eq. (34.7), or equivalently Eq. (34.11). Assume we initially start with an empty production line (i.e., $x(0) = 0$), that both machines have a capacity of 1 lot per unit time (i.e., $\mu_1 = \mu_2 = 1$) and that we feed the line at a rate of 1 lot per time unit (i.e., $u_0 = 1$). Furthermore, assume that machines produce at full capacity, but only in case something is in the buffer in front of it, i.e.,

$$u_i(t) = \begin{cases} \mu_i & \text{if } y_i(t) > 0 \\ & \qquad\qquad i \in \{1, 2\} \\ 0 & \text{otherwise} \end{cases} \tag{34.12}$$

Under these assumptions, the resulting contents of buffer B_3 are as displayed in Figure 34.13. Note that immediately lots start coming out of the system. Clearly, this is not what happens in practice. Since both machines M_1 and M_2 need to process the first lot, it should take the system at least $(1/\mu_1) + (1/\mu_2)$ time

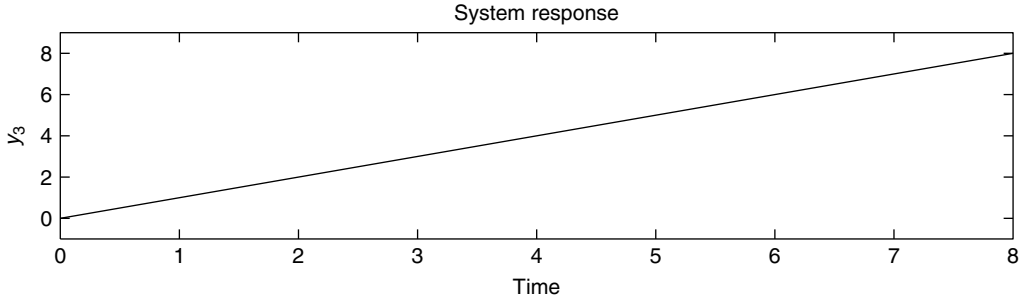


FIGURE 34.13 Output of the manufacturing system using model Eq. (34.6).

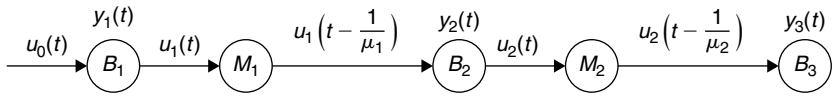


FIGURE 34.14 A simple manufacturing system revisited.

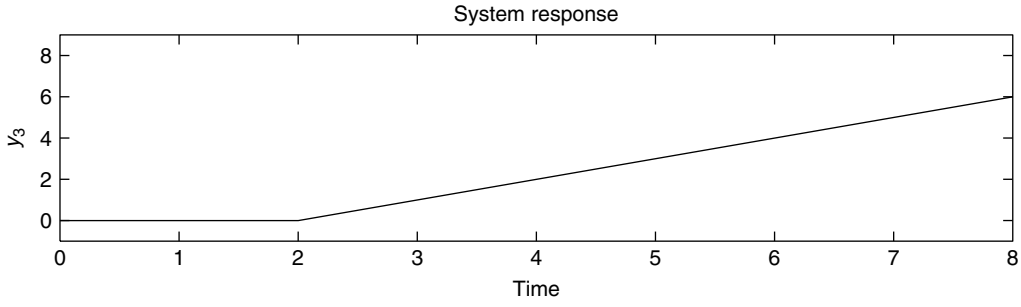


FIGURE 34.15 Output of the manufacturing system using model Eq. (34.13).

units before lots can come out. This illustrates fluid models as given by Eq. (34.7) or Eq. (34.11) do not incorporate flow times.

34.7.2 An Extended Fluid Model

In the previous subsection, we noticed that in the standard fluid model lots immediately come out of the system, once we start producing. A way to overcome this problem is to explicitly take into account the required delay. Whenever we decide to change the production rate of machine M_1, buffer B_2 notices this $1/\mu_1$ time units later. As a result, the rate at which lots arrive at buffer B_2 at time t is equal to the rate at which machine M_1 was processing at time $t - 1/\mu_1$. This observation results in the following model (see also Figure 34.14):

$$\begin{aligned}
\dot{y}_1(t) &= u_0(t) - u_1(t) \\
\dot{y}_2(t) &= u_1\left(t - \frac{1}{\mu_1}\right) - u_2(t) \\
\dot{y}_3(t) &= u_2\left(t - \frac{1}{\mu_2}\right)
\end{aligned} \tag{34.13}$$

Clearly, the constraints of Eq. (34.8) also apply to the model given by Eq. (34.13).

Figure 34.15 shows the response of the system given by Eq. (34.12) to the ramp-up experiment that lead to Figure 34.13. Comparing these two figures we see that no products enter buffer B_3 during the first 2.0 time units for the extended fluid model. Clearly, the extended fluid model produces more realistic results than the standard fluid model.

34.7.3 An Approximation to the Extended Fluid Model

In the previous subsection, we proposed an extended version of the standard fluid model. Although the model of Eq. (34.13) still is a linear model, standard linear control theory is not able to deal with this model, due to the time delay. Instead we have to rely on control theory of infinite-dimensional linear systems (see, e.g., Curtain and Zwart, 1995).

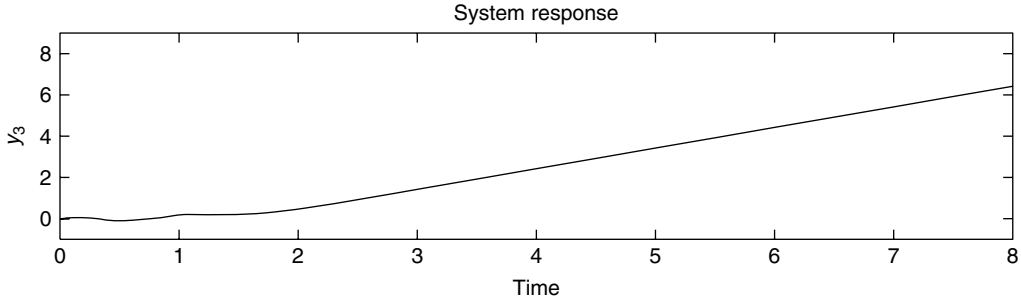


FIGURE 34.16 Output of the manufacturing system using model Eq. (34.14).

Another possibility would be to approximate the time delays by means of a Padé approximation (Baker, 1965). Using a second-order Padé approximation, the model of Eq. (34.13) can be approximated as:

$$\dot{x} = \left[\begin{array}{ccccccc} \mathbf{0} & \mathbf{0} & 0 & 0 & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & 4 & 6 & -3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 6 & -3 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 \end{array}\right] x + \left[\begin{array}{ccc} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] u \tag{34.14a}$$

$$y = \left[\begin{array}{ccccccc} \mathbf{1} & \mathbf{-1} & 0 & 0 & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{1} & -3 & 0 & \mathbf{-1} & 0 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 0 & \mathbf{1} & -3 & 0 \end{array}\right] x + \left[\begin{array}{ccc} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right] u \tag{34.14b}$$

Note the structure in Eq. (34.14). In bold face we can easily recognize the dynamics of Eq. (34.11). The additional dynamics results from the Padé approximation.

If we initiate the system of Eq. (34.14) from $x(0) = 0$ and feed it at a rate $u_0 = 1$ while using Eq. (34.12), we obtain the system response as depicted in Figure 34.16. It is clear that we do not get the same response as in Figure 34.15, but the result is rather acceptable from a practical point of view. At least it is closer to reality than the response displayed in Figure 34.13.

34.7.4 A Hybrid Model

In the previous subsections, we provided extensions to the standard fluid model by taking into account the time delay lots encounter owing to the processing of machines. We also mentioned the constraints of Eq. (34.8) that have to be obeyed. These are constraints that we have to take into account when designing a controller for our manufacturing system. The way we dealt with these constraints in the previous subsections was by requiring the machines to produce only in case the buffer contents in front of that machine were positive (cf. Eq. 34.12).

A way to extend the standard fluid model Eq. (34.7) is to think of these constraints in a different way. As illustrated in Subsection 34.7.1, when we turn on both machines, immediately lots start coming out of the system. This is an undesirable feature that we would like to avoid. In practice, the second machine can only start producing when the first machine has finished a lot. Keeping this in mind, why do we allow machine M_2 to start producing as soon as the buffer contents of the buffer in front of it are positive? Actually, machine M_2 should only start producing as soon as a whole product has been finished by the machine M_1.

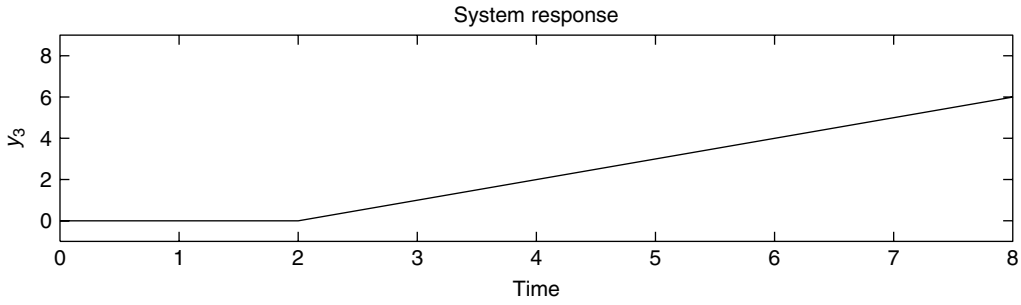


FIGURE 34.17 Output of the manufacturing system.

In words, machine M_2 should only start producing as soon as the buffer contents of the buffer in front of it becomes 1. Therefore, we should not allow for a positive u_2 as soon as $y_2 > 0$, but only in case $y_2 \geq 1$.

When we consider the initially empty system Eq. (34.7), i.e., $x(0) = 0$, and assume

$$u_i(t) = \begin{cases} \mu_i & \text{if } y_i(t) \geq 1 \\ 0 & \text{otherwise} \end{cases} \qquad i \in \{1, 2\} \tag{34.15}$$

the resulting system response to an input of $u_0 = 1$ is shown in Figure 34.17. Note that we obtain exactly the same response as in Figure 34.15.

While this hybrid model has an acceptable behavior when we ramp up our manufacturing system, it stops producing at buffer level $y_i = 1$ when we ramp down. This is not what we would like to have. Therefore, in case $u_1 = 0$, machine M_1 should be allowed to produce until $y_1 = 0$.

Under these conditions, we could also think of our model operating in different modes. For the manufacturing system under consideration, we can distinguish the following modes:

mode 1:	$0 \leq y_1 \leq 1$,	$0 \leq y_2 \leq 1$,	$u_0 = 0$,	$u_1 \geq 0$,	$u_2 = 0$.
mode 2:	$0 \leq y_1 \leq 1$,	$0 \leq y_2 \leq 1$,	$u_0 \geq 0$,	$u_1 = 0$,	$u_2 \geq 0$.
mode 3:	$1 \leq y_1$,	$0 \leq y_2 \leq 1$,		$u_1 = 0$,	$u_2 \geq 0$.
mode 4:	$1 \leq y_1$,	$0 \leq y_2 \leq 1$,		$u_1 \geq 0$,	$u_2 = 0$.
mode 5:	$0 \leq y_1 \leq 1$,	$1 \leq y_2$,	$u_0 = 0$,	$u_1 \geq 0$.	
mode 6:	$0 \leq y_1 \leq 1$,	$1 \leq y_2$,	$u_0 \geq 0$,	$u_1 = 0$.	
mode 7:	$1 \leq y_1$,	$1 \leq y_2$.			

In all of these modes, the system dynamics is described by Eq. (34.7).

The hybrid model just presented is also known as a piecewise affine (PWA) system (Sontag, 1981). Other well-known descriptions are linear complementarity (LC) systems (Heemels et al., 2000; Schaft and Schumacher, 1998) and mixed logical dynamical (MLD) systems (Bemporad and Morari, 1999). In Bemporad et al. (2000b) and Heemels et al. (2001) it was shown that (under certain assumptions like well-posedness) these three descriptions are equivalent. This knowledge is useful, as each modeling class has its own advantages. Stability criteria for PWA systems were proposed in DeCarlo et al. (2000) and Johansson and Rantzer (1998), and control and state-estimation techniques for MLD hybrid models have been presented in Bemporad et al. (2000a, 1999) and Bemporad and Morari (1999). These results can now be applied for controlling the hybrid systems model of our manufacturing system.

34.8 Flow Models

The fluid models presented in the previous section are (still) not satisfactory. While they do not suffer from the problem that lots come out of the system as soon as we start producing, flow times are not truly present in these models. It is not possible to determine the time it takes lots to leave once they have entered

the system. Furthermore, according to these models any throughput can be achieved by means of zero inventory, whereas in Section 34.2 we already noticed that the nonlinear (steady-state) relations depicted in Figure 34.2 should be incorporated in any reasonable model of manufacturing systems.

In this section, we present approximation models that do incorporate both throughput and flow time. These dynamic models are inspired by the continuum theory of highway traffic. Therefore, before presenting this dynamic model we first present some results from traffic theory.

34.8.1 Introduction to Traffic Flow Theory: The LWR Model

In the mid-1950s Lighthill and Whitham (1955) and Richards (1956) proposed a first-order fluid approximation of traffic flow dynamics. This model nowadays is known in traffic flow theory as the LWR model.

Traffic behavior for a single one-way road can be described using three variables that vary in time t and space x: flow $u(x,t)$, density $\rho(x,t)$, and speed $v(x,t)$. Flow is the product of speed and density:

$$u(x,t) = \rho(x,t)v(x,t) \quad \forall x,t \tag{34.16}$$

For a highway without entrances or exits, the number of cars between any two locations x_1 and x_2 ($x_1 < x_2$) needs to be conserved at any time t, i.e., the change in the number of cars between x_1 and x_2 is equal to the flow entering via x_1 minus the flow leaving via x_2:

$$\frac{\partial}{\partial t}\int_{x_1}^{x_2} \rho(x,t)\mathrm{d}x = u(x_1,t) - u(x_2,t) \tag{34.17a}$$

or in differential form:

$$\frac{\partial \rho}{\partial t}(x,t) + \frac{\partial u}{\partial x}(x,t) = 0 \tag{34.17b}$$

The two relations Eq. (34.16) and Eq. (34.17) are basic relations that any model must satisfy. As we have three variables of interest, a third relation is needed. For this third relation, several choices can be made. The LWR model assumes in addition to the relations Eq. (34.16) and Eq. (34.17) that the relation between flow and density observed under steady-state conditions also holds when flow and density vary with x and/or t; i.e., for a homogeneous highway:

$$u(x,t) = S(\rho(x,t)) \tag{34.18}$$

The model given by Eqs. (34.16)–(34.18) can predict some traffic phenomena rather well. To overcome some of the deficiencies of the LWR model, in the early 1970s higher order theories have been proposed where Eq. (34.18) has been replaced by another partial differential equation, containing diffusion or viscosity terms. Unfortunately, these extended models experience some undesirable properties, as made clear in (Daganzo, 1995). The most annoying of these properties is the fact that in these second-order models cars can travel backward. Second-order models that do not suffer from this deficiency have been presented in Jiang et al. (2002) and Zhang (2002).

34.8.2 A Traffic Flow Model for Manufacturing Flow

In the previous subsection, we introduced the LWR model from traffic flow theory. This model describes the dynamic behavior of cars along the highway at a macroscopic level and contains information both about the number of cars passing a certain point and about the time it takes cars to go from one point to the other. The LWR model can not only be used for describing the flow of cars along the highway, but also for describing the flow of products through a manufacturing line.

Consider, instead of a homogeneous highway, a homogeneous manufacturing line, i.e., a manufacturing line that consists of a lot of identical machines. Let t denote the time and x the position in the manufacturing line. The behavior of lots flowing through the manufacturing line can also be described by three variables

that vary with time and position: flow $u(x,t)$ measured in unit lots per unit time, density $\rho(x,t)$ measured in unit lots per unit machine, and speed $v(x,t)$ measured in unit machines per unit time. Now we can relate these three variables by means of Eqs. (34.16)–(34.18), where in Eq. (34.18) the function S describes the relation between flow and density observed under steady-state conditions.

To make this last statement more explicit, consider a manufacturing line consisting of m machines with exponentially distributed process times and an average capacity of μ lots per unit time. Furthermore, consider a Poisson arrival process where lots arrive at the first machine with a rate of λ lots per unit time ($\lambda < \mu$), and assume that buffers have infinite capacity. Then we know from queueing theory (Kleinrock, 1975) that the average number of lots N in each workstation (consisting of a buffer and a machine) in steady-state is given by

$$N = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda} \tag{34.19}$$

In words, in steady-state we have $\rho(x,t)$ is constant and

$$\frac{1}{m}\rho(x,t) = \frac{u(x,t)}{\mu - u(x,t)} \tag{34.20}$$

from which we can conclude that in steady-state

$$u(x,t) = \frac{\mu\rho(x,t)}{m + \rho(x,t)} \tag{34.21}$$

For this example, this is the mentioned function $S(\rho)$.

With this information, we can conclude that the dynamics of this manufacturing line might be described by means of the partial differential equation

$$\frac{\partial\rho}{\partial t} + \mu\frac{\partial}{\partial x}\left(\frac{\rho}{m+\rho}\right) = 0 \tag{34.22a}$$

Together with the relations

$$u = \frac{\mu\rho}{m+\rho} \quad \text{and} \quad v = \frac{u}{\rho} \quad \text{or} \quad v = \frac{\mu}{m+\rho} \tag{34.22b}$$

this completes our model.

Note that contrary to the fluid models presented in the previous section, the dynamic model of Eq. (34.22) is able to incorporate the stochasticity as experienced in manufacturing lines. If the manufacturing line would be in steady-state, the throughput and flow time as predicted by the model of Eq. (34.22) is exactly the same as those predicted by queueing theory. However, contrary to queueing theory, the model of Eq. (34.22) is not a steady-state model, but also incorporates dynamics. Therefore, the model Eq. (34.22) is a dynamic model that incorporates both throughput and flow time. Furthermore, given the experience in the field of fluid dynamics, the model is computationally feasible as well. For more on these flow models, the interested reader is referred to Armbruster et al. (2004, 2005) and Armbruster and Ringhofer (2005).

34.9 Conclusions

In this chapter, we presented some of the models used in the modeling, analysis, and control of manufacturing systems. In Section 34.3 some analytical modeling techniques and methods for analyzing steady-state behavior of manufacturing systems have been introduced. To get a more detailed insight in the dynamics of a manufacturing system discrete-event models, as introduced in Section 34.4 can be used. A disadvantage

of discrete-event models is their complexity. To reduce the complexity of discrete-event models, EPTs have been introduced in Section 34.5. This enables the modeling of a manufacturing system as a large queueing network.

Once the dynamics of manufacturing systems can be well described by a relatively simple discrete-event model, the problem of controlling the dynamics of manufacturing systems becomes of interest. In Section 34.6 a control framework has been presented. A crucial role in this framework is played by approximation models of manufacturing systems. In Section 34.7 the most common approximation models, fluid models, have been introduced, together with some extensions of these models. These fluid models mainly focus on throughput and do not contain information on flow times. Finally, in Section 34.8, flow models have been presented that do incorporate both throughput and flow time information.

References

Armbruster, D., P. Degond, and C. Ringhofer (2005). Continuum models for interacting machines. In D. Armbruster, K. Kaneko, and A. Mikhailov (Eds.), *Networks of Interacting Machines: Production Organization in Complex Industrial Systems and Biological Cells*. Singapore: World Scientific Publishing.

Armbruster, D., D. Marthaler, and C. Ringhofer (2004). Kinetic and fluid model hierarchies for supply chains. *SIAM Journal on Multiscale Modeling and Simulation 2*(1), 43–61.

Armbruster, D. and C. Ringhofer (2005). Thermalized kinetic and fluid models for re-entrant supply chains. *SIAM Journal on Multiscale Modeling and Simulation 3*(4), 782–800.

Baker Jr., G. A. (1965).The theory and application of the Pade approximant method. In K. A. Brueckner (Ed.), *Advances in Theoretical Physics*, Volume 1, pp. 1–58. New York: Academic Press.

Bemporad, A., F. Borrelli, and M. Morari (2000a, June). Piecewise linear optimal controllers for hybrid systems. In *Proceedings of the 2000 American Control Conference*, Chicago, IL, pp. 1190–1194.

Bemporad, A., G. Ferrari-Trecate, and M. Morari (2000b, October). Observability and controllability of piecewise affine and hybrid systems. *IEEE Transactions on Automatic Control 45*(10), 1864–1876.

Bemporad, A., D. Mignone, and M. Morari (1999, June). Moving horizon estimation for hybrid systems and fault detection. In *Proceedings of the 1999 American Control Conference*, San Diego, CA, pp. 2471–2475.

Bemporad, A. and M. Morari (1999). Control of systems integrating logic, dynamics, and constraints. *Automatica 35*, 407–427.

Curtain, R. F. and H. Zwart (1995). *An Introduction to Infinite-Dimensional Linear Systems Theory*. Berlin, Germany: Springer.

Daganzo, C. F. (1995). Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B 29*(4), 277–286.

DeCarlo, R. A., M. Branicky, S. Petterson, and B. Lennartson (2000). Perspectives and results on the stability and stabilizability of hybrid systems. *Proceedings of the IEEE 88*(7), 1069–1082.

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Heemels, W. P. M. H., J. M. Schumacher, and S. Weiland (2000). Linear complementarity systems. *SIAM Journal on Applied Mathematics 60*(4), 1234–1269.

Heemels, W. P. M. H., B. d. Schutter, and A. Bemporad (2001). Equivalence of hybrid dynamical models. *Automatica 37*(7), 1085–1091.

Hopp, W. J. and M. L. Spearman (2000). *Factory Physics*, second ed. New York: Irwin/McGraw-Hill.

Jacobs, J. H., P. P. v. Bakel, L. F. P. Etman, and J. E. Rooda (2006). Quantifying variability of batching equipment using effective process times. *IEEE Transactions on Semiconductor Manufacturing 19*(2), 269–275.

Jacobs, J. H., L. F. P. Etman, E. J. J. v. Campen, and J. E. Rooda (2003). Characterization of the operational time variability using effective processing times. *IEEE Transactions on Semiconductor Manufacturing 16*(3), 511–520.

Jiang, R., Q. S. Wu, and Z. J. Zhu (2002). A new continuum model for traffic flow and numerical tests. *Transportation Research. Part B, Methodological 36*, 405–419.

Johansson, M. and A. Rantzer (1998). Computation of piece-wise quadratic Lyapunov functions for hybrid systems. *IEEE Transactions on Automatic Control 43*(4), 555–559.

Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Proceedings of the Cambridge Philosophical Society 57*, 902–904.

Kleinrock, L. (1975). *Queueing Systems, Volume I: Theory*. New York: Wiley.

Kuehn, P. J. (1979). Approximate analysis of general queueing networks by decomposition. *IEEE Transactions on Communication 27*, 113–126.

Lighthill, M. J. and J. B. Whitham (1955). On kinematic waves. I: Flow movement in long rivers. II: A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society A 229*, 281–345.

Little, J. D. C. (1961). A proof of the queueing formula $l = \lambda w$. *Operations Research 9*, 383–387.

Richards, P. I. (1956). Shockwaves on the highway. *Operations Research 4*, 42–51.

Sattler, L. (1996, November). Using queueing curve approximations in a fab to determine productivity improvements. In *Proceedings of the 1996 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop*. Cambridge, MA, pp. 140–145.

Schaft, A. J. v. d. and J. M. Schumacher (1998). Complementarity modelling of hybrid systems. *IEEE Transactions on Automatic Control 43*, 483–490.

Sontag, E. D. (1981). Nonlinear regulation: the piecewise linear approach. *IEEE Transactions on Automatic Control 26*(2), 346–358.

Zhang, H. M. (2002). A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research. Part B, Methodological 36*, 275–290.