

DETERMINISTIC AND STOCHASTIC DYNAMIC MODELING OF CONTINUOUS MANUFACTURING SYSTEMS USING ANALOGIES TO ELECTRICAL SYSTEMS

Bashar H. Sader
Carl D. Sorensen

Mechanical Engineering Department
Brigham Young University
Provo, UT 84602, U.S.A.

ABSTRACT

A dynamic system model of continuous manufacturing systems has been developed based on analogies with electrical systems. This model has the capability to model both deterministic and stochastic systems. The model provides physically meaningful governing equations to describe both the steady state and transient responses of continuous manufacturing systems. For stochastic solutions, the model is not limited to any specific probabilistic distribution. The model is demonstrated by application to a representative continuous manufacturing line for both deterministic and stochastic cases. The results of the stochastic case are compared to those from a discrete event simulation tool using a paired t-test at the 95% confidence level. For some results, the difference is statistically insignificant. For others, there is a statistically significant difference. However, in both cases the percentage difference is within a reasonable range.

1 INTRODUCTION

Manufacturing systems are commonly modeled using discrete event simulation and queuing theory. While the discrete event simulation approach provides a powerful tool to understand and improve manufacturing systems, it does not provide closed form solutions to physically describe the dynamics of the system. The common modeling approach, queuing theory, is limited to special cases and cannot be used in many situations (de Souza et al. 1996). A third alternative is to use the system dynamics approach. Although bond-graphs and state equations have been used to model manufacturing systems (Ferneu 2000, Besombes and Marcon 1993), this approach has not been very common (Baines and Harrison 1999).

In this paper we present a dynamic model for single product continuous manufacturing systems. The basic idea of the model is to employ analogies between continuous manufacturing systems and electrical systems. These

analogies are intended to simplify the modeling process and produce a visually understandable graphical representation of the manufacturing system being modeled. The graphical model can then be used to write the governing equations of the system. However, the analogy to electrical systems will be carried no further than the graphical representation and basic variables and components. The governing equations will be developed for the manufacturing system variables without referring to the equations of the analogous electrical system.

2 THE MODEL

The main variables of interest in a manufacturing system are the material flow (throughput) measured in units of material per unit time, and the amount of work in process (WIP) measured in units of material. The material flow through any place in the manufacturing system is analogous to the current in an electrical system. WIP existing at any place in the manufacturing system is analogous to the charge on a capacitor in an electrical system. For this model, all capacitors will have a capacitance of 1, so that WIP is also equal to the voltage on the capacitor. A third important variable is the amount of time a material unit spends in the system (cycle time). This last variable will not be included in the model. But at steady state, from the WIP and throughput it is possible to calculate the cycle time using Little's law (Hopp and Spearman 2001).

In order to process materials at a manufacturing station and thus produce material flow, some conditions have to be satisfied. First, there must be some raw materials to work on. Second, there must be some capacity available at the station. And third there must be a signal to allow the station to work on that material. This signal can take the form of an operator manually feeding a machine, an electrical signal in an automated line or any other form. It is this signal that is commonly used to control a station.

A model of a manufacturing station needs to represent these three conditions. An ideal transistor in parallel with a

capacitor provides these requirements, (Figure 1). The transistor passes a current (i_{pass}) proportional to the control current (i_{cont}) as long as the voltage across the capacitor is at least equal to the threshold voltage (V_{th}). If the voltage is less than the threshold voltage, the transistor current is reduced, see Figure 2. If there is an enforced input or output material flow rate, a current source in the proper direction can be used to represent it. A voltage follower is used in the circuit to ensure that no current will flow from the capacitor without going through the transistor and to keep the voltage across the transistor (V_{tran}) equal to that across the capacitor, i.e., $V_{tran}=V_{cap}$. The control current is usually much smaller than the current passing through the capacitor (i_{cap}), thus $i_{pass}=i_{cap}+i_{cont}\cong i_{cap}$.

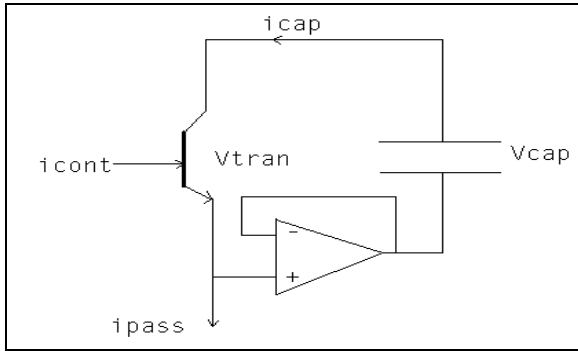


Figure 1: Ideal Transistor in Parallel with a Capacitor

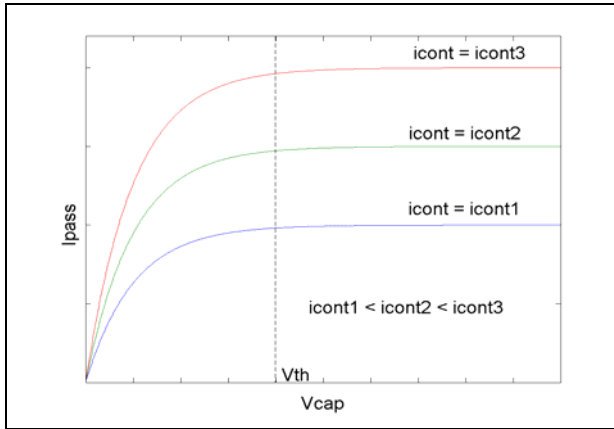


Figure 2: Ideal Transistor Voltage-Current Relation

Now we can establish the analogy between a manufacturing station and the ideal transistor in parallel with a capacitor. There should be some material in the station (analogous to a voltage across the capacitor) to have a material flow rate out of it (analogous to some current through the transistor). The value of the possible output rate is determined by the control signal (analogous to the control current) but cannot exceed some maximum value which is a property of the station (analogous to the rated transistor current). It should be noted here that a station can have either a single machine or a number of machines working in

parallel. It is the total capacity rate of the station that needs to be modeled. Table 1 summarizes the manufacturing-electrical analogies developed in this section.

Table 1: Manufacturing-Electrical Analogies

Manufacturing systems	Electrical Systems
Throughput	Current
WIP	Voltage
Buffer	Capacitor
Machine	Transistor
Enforced input or output rate	Current source

Using the above analogy, we model a single station manufacturing system. The modeling will be based on three assumptions:

1. Single product line.
2. Infinite storing capacity at the station.
3. A push policy is applied (any time material is present, the machine is authorized to work on it, which is analogous to the control current being set to its maximum value).

Such a system can be modeled as shown in Figure 3. Although the remainder of this paper deals with push systems as outlined in assumption 3, the model technique is not limited to this case. Other papers in preparation will show the application to a variety of pull systems, including CONWIP, Kanban and MAXWIP (Sader and Sorensen, in preparation). In fact, the ability to model a variety of control schemes is one of the strong points of this model.

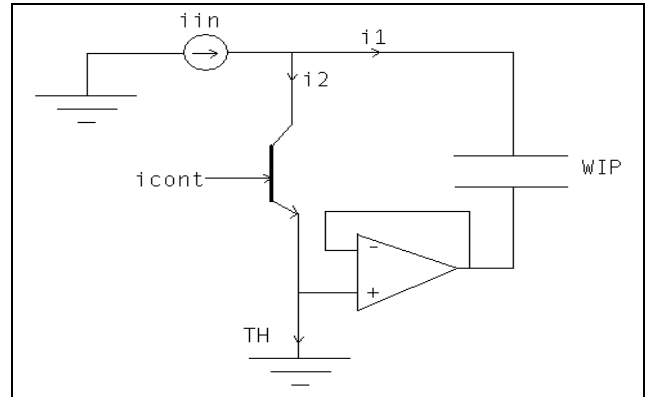


Figure 3: Single Station Continuous Manufacturing System

In the above model, i_{in} represents the arrival rate, which is the rate at which materials are released to the system. i_{in} is modeled as a current source, and it cannot be a negative value ($i_{in} \geq 0$). i_1 represents the rate at which materials are flowing in or out of the WIP storage in the station. i_2 represents the material flow through the machine(s) which is equal to the throughput (TH) of the station, and is always a nonnegative value ($i_2 \geq 0$). Note that i_{in} , i_1 , i_2 and WIP are all functions of time. The performance of this sys-

tem is independent of any amounts of WIP (analogous to voltage) outside it, thus it is grounded on both ends. Let's assume that the effective time needed to process a single unit of material is t_e (minutes), and that the machine (or parallel machines) can simultaneously process a maximum of k (units). This yields a maximum possible capacity rate of (k/t_e) units per minute. So, we can define i_2 as:

$$i_2 = TH = \begin{cases} WIP/t_e & : WIP < k \\ k/t_e & : WIP \geq k \end{cases} \quad (1)$$

where WIP is the amount of material units available in the station. The WIP includes both the units being processed and the units waiting to be processed. Equation (1) is a direct result of the fact that the station cannot process at maximum capacity if there is not enough material in the station.

Applying Kirchoff's first law gives:

$$i_{in} = i_1 + TH. \quad (2)$$

Since i_1 is the rate at which material is flowing in or out of the capacitor, i_1 can be defined as:

$$i_1 = \frac{dWIP}{dt}. \quad (3)$$

Assume we started with an empty system (i.e., $WIP = 0$), and consider the system after some time t has elapsed. If WIP is less than k , then Equations (1), (2) and (3) give:

$$\frac{dWIP}{dt} + \frac{WIP}{t_e} = i_{in} \quad (WIP < k). \quad (4)$$

But if WIP is greater than or equal to k , then:

$$\frac{dWIP}{dt} = i_{in} - \frac{k}{t_e} \quad (WIP \geq k). \quad (5)$$

Equations (4) and (5) describe the performance of a single station continuous manufacturing system subject to the assumptions previously mentioned.

Using Equation (1), we can define the utilization of the station. The utilization, u , is defined as the arrival rate to the station divided by the maximum possible processing rate, thus:

$$u = \frac{i_{in}}{(k/t_e)}. \quad (6)$$

3 THE DETERMINISTIC CASE

In this section, we add two more assumptions to the ones we made in section 2:

4. The arrival rate (i_{in}) is a deterministic constant value.
5. The effective processing time (t_e) is a deterministic constant value.

Using these assumptions, and applying the initial condition $WIP(0) = 0$, Equation (4) can be solved to give:

$$WIP(t) = i_{in} t_e (1 - e^{-t/t_e}) = uk (1 - e^{-t/t_e}) \quad (t < t_s). \quad (7)$$

Equation (7) is valid up to the time t_s at which saturation is reached (i.e., $WIP(t_s) = k$). This time, t_s , can be determined by setting $WIP(t_s)$ equal to k , thus getting:

$$t_s = -t_e \ln(1 - k/i_{in} t_e) = -t_e \ln(1 - 1/u). \quad (8)$$

Equation (5) can also be solved for the condition $WIP(t_s) = k$, which yields:

$$WIP(t) = (i_{in} - k/t_e)[t + t_e \ln(1 - k/i_{in} t_e)] + k = (k/t_e)(u - 1)(t - t_s) + k \quad (t \geq t_s). \quad (9)$$

Equation (9) is valid after the saturation time, t_s .

Note from Equation (8) that if $(k/i_{in} t_e)$ is larger than or equal to 1 (utilization is less than or equal to 100%), then t_s does not exist. This means that WIP will never reach k and Equation (7) will be valid for all time. If we evaluate Equation (7) at time $t \rightarrow \infty$, we get $WIP(t \rightarrow \infty) = i_{in} t_e$. Note that the quantity $i_{in} t_e$ is a constant, so, $dWIP/dt = 0$ as $t \rightarrow \infty$ and Equations (2) and (3) will give the value of TH as:

$$TH = i_{in}.$$

In that case Equation (7) can be written as:

$$\lim_{t \rightarrow \infty} WIP(t) = TH t_e \quad (u \leq 1). \quad (10)$$

For the steady state case (i.e., $t \rightarrow \infty$), Equation (7) reduces to Equation (10), which is readily recognizable as Little's law. This suggests that Equation (7) is a general form of Little's law that covers both the transient and steady state responses, given the assumptions here.

The previous analysis can be carried out for a series line of n stations, where the output rate of machine j is the input rate of machine $j+1$. MATLAB code was written to numerically solve the resulting linked set of differential equations and plot the results. Consider a simple continuous manufacturing system that consists of three stations in series. All the assumptions stated in section 2 in addition to assumption 5 from this section are still employed. Assumption 4 is still made for arrivals to the line, all of which go

to station 1. For this example: $i_{in1} = i_{in}$, $i_{out1} = i_{in2}$, $i_{out2} = i_{in3}$, $i_{out3} = TH$. This system is shown in Figure 4.

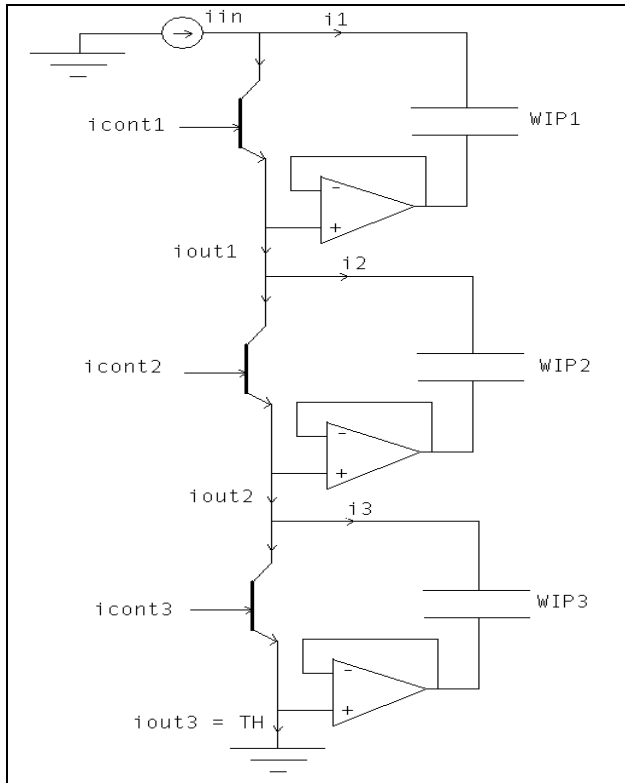


Figure 4: Three Station Continuous Manufacturing System

We solve the equations for the case of $i_{in} = 10$ units/minute, $k_1 = k_2 = k_3 = 10$ units and $u_1 = u_2 = u_3 = 100\%$. We will generate a graphical solution for the first hour starting with an empty system. A plot showing the WIP levels and i_{out} at each station as a function of time is shown in Figure 5. Note the time delays in both WIP and flow rate between the three stations and note the transient response. Now let's decrease the available capacity rate at station 2 (by increasing t_{e2}) so that its utilization is 110%. This case is shown in Figure 6. Note that WIP_2 increases linearly with time and that machine 2 is the bottleneck (it determines the throughput of the system). Theoretically, WIP_2 will be infinite as $t \rightarrow \infty$. Nevertheless, at any given finite time, a finite value of WIP_2 can be determined. Thus, this system can be effectively modeled when there are finite time periods where release rate exceeds capacity. This is an advantage of this modeling technique over other models in which only steady state behavior can be modeled. Now let's relieve the bottleneck by increasing k_2 (the maximum number of material units that can simultaneously be processed at station 2) to 11. This case is shown in Figure 7. Note that the station reaches steady state.

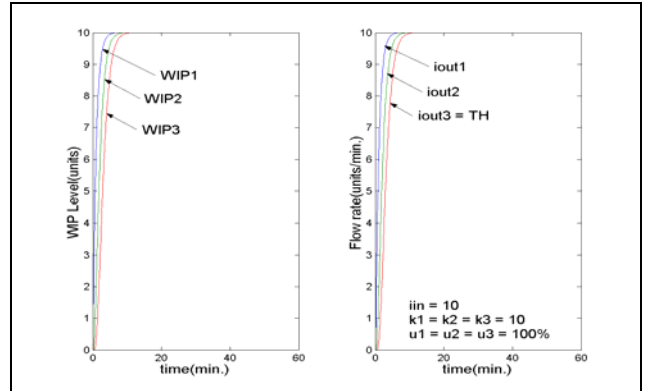


Figure 5: Deterministic Three Station System without a Bottleneck

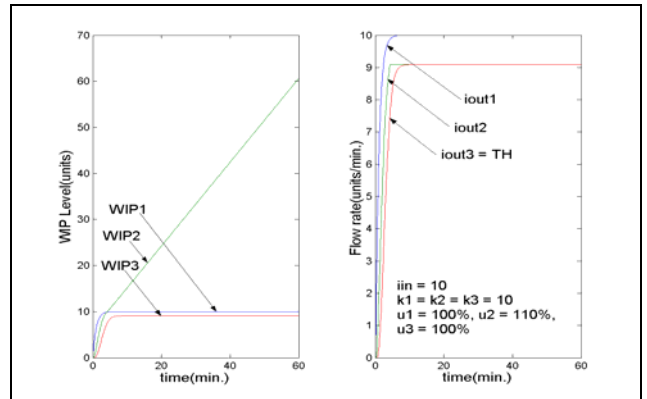


Figure 6: Deterministic Three Station System with a Bottleneck

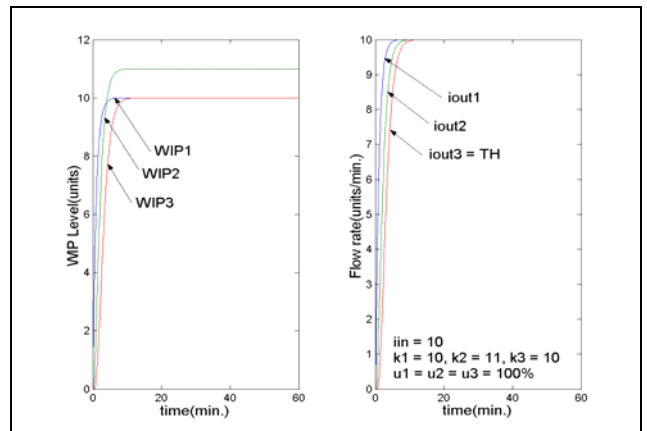


Figure 7: Deterministic Three Station System with the Bottleneck Relieved

4 THE STOCHASTIC CASE

To model stochastic processes we allow effective processing times of the stations to be random numbers that come from distributions with known means and standard deviations. In other words, assumption 5 from the previous sec-

tion will be relaxed while assumptions 1, 2, 3 and 4 will still be employed.

Equations (4) and (5) remain the governing equations, however, t_e varies stochastically with time. Rather than searching for a closed-form solution for a specific distribution, we choose to solve the ordinary differential equations by numerical integration with random values for t_e .

The utilization definition, Equation (6), can be rewritten for any station j with stochastic t_{ej} and deterministic i_{inj} as:

$$u_j = \frac{i_{inj}}{E(k_j/t_{ej})} = \frac{(i_{inj}/k_j)}{E(1/t_{ej})}$$

where E represents the expectation operator. It should be noted that $E(1/t_{ej})$ is not equal to $1/E(t_{ej})$. Also note that $\text{Var}(1/t_{ej})$ is not equal to $1/\text{Var}(t_{ej})$, where Var is the variance operator. For a uniformly distributed random number x on the interval $[a,b]$:

$$E(x) = (a+b) / 2,$$

$$\text{Var}(x) = (b-a)^2 / 12,$$

$$\text{Coefficient of variation } (x) = \frac{\sqrt{\text{Var}(x)}}{E(x)}.$$

On the other hand,

$$E(1/x) = \frac{\ln(b) - \ln(a)}{b - a},$$

$$\text{Var}(1/x) = \frac{1}{ab} \frac{(\ln(b) - \ln(a))^2}{(b-a)^2},$$

$$\text{Coefficient of variation } (1/x) = \frac{\sqrt{\text{Var}(1/x)}}{E(1/x)}.$$

This means that to achieve a mean available capacity rate, $E(k/t_e)$, of say 10 parts per minute with a coefficient of variation (cv) of 0.1 at a machine with $k = 10$ units, we need to choose the interval $[a,b]$ for a uniform distribution in such a way that makes $E(1/t_e) = 1$ and the cv of $(1/t_e) = 0.1$. It is important to note that the interval $[a,b]$ that meets these requirements will not necessarily give a mean value of t_e equal to 1 or a cv value of t_e equal to 0.1. The parameters of any probabilistic distribution for t_e should be chosen such that $(1/t_e)$ has the desired expected value and variance.

The MATLAB code mentioned in section 3 was modified to use random numbers generated from known distributions instead of deterministic values of t_e 's. The code generates a new value for t_e (updates it) every deterministic time period t_u . In this study, the value for t_e is chosen to

be $E(t_e)$. However, if better information is known for a specific system, it should be used. This is also the case with deciding which distribution to use for t_e .

Now consider the same example from section 3. But now t_{e1} , t_{e2} and t_{e3} are random numbers generated from uniform distributions. The parameters of the uniform distributions for t_e 's are chosen such that the mean available capacity will yield the desired level of utilization and variability. Figures 8, 9 and 10 show graphical results for 3 different scenarios, which are the same scenarios seen in Figures 5, 6 and 7 respectively, but this time with stochastic t_e 's. The system parameters assumed are shown on each graph. At this relatively low level of variation ($cv = 0.05$), it can be seen that the graphs are similar to the no variability case. While the stations maintained average output flow rates close to the deterministic case, they accumulated higher levels of WIP.

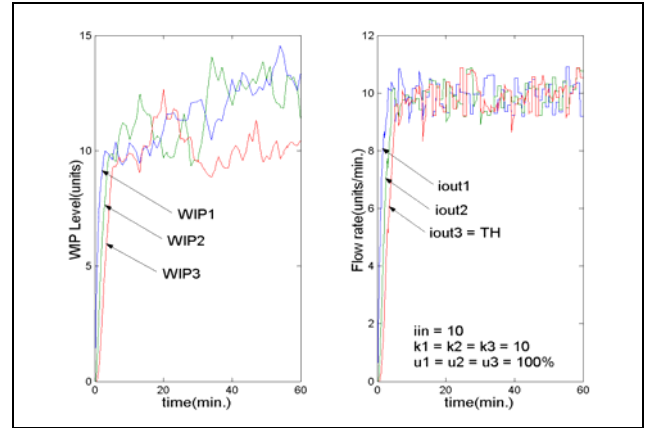


Figure 8: Stochastic ($cv = 0.05$) Three Station System without a Bottleneck

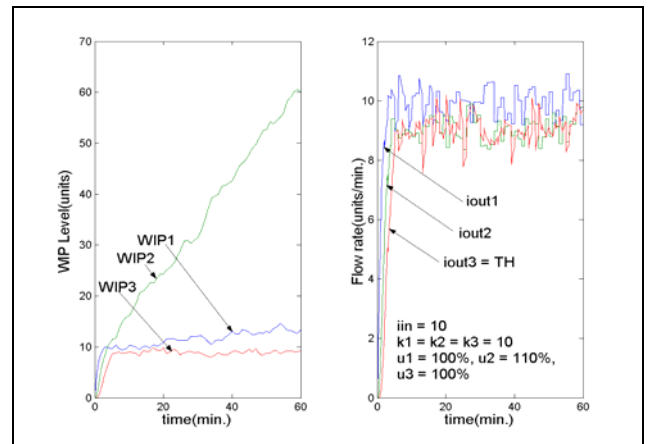


Figure 9: Stochastic ($cv = 0.05$) Three Station System with a Bottleneck

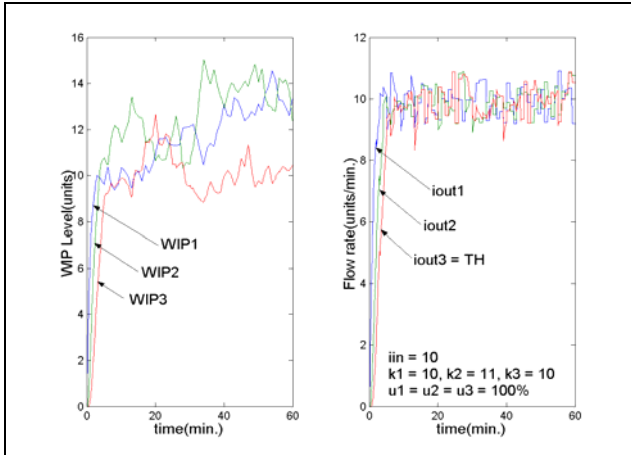


Figure 10: Stochastic ($cv = 0.05$) Three Station System with the Bottleneck Relieved

The effect of a higher variability level in the system can be shown for the same example with coefficients of variation of 0.25 at each station. We are still trying to achieve a 100% utilization level at each station. Figures 11, 12 and 13 represent the same scenarios seen in Figures 8, 9 and 10 respectively but at this higher variability level. The system parameters assumed are shown on each graph. Note that although the stations still maintained average output flow rates close to the previous cases, the WIP levels became considerably higher. Actually, at high levels of variability, the WIP levels may become so high that for practical reasons they can be thought of as infinite. This is in agreement with the literature that gives an infinite WIP level at a utilization level of 100% for exponentially distributed te 's ($cv = 1.0$) (Hopp and Spearman 2001). Figures 5 through 13 show that it is the combination of high levels of utilization and high variability levels that generates high levels of WIP. The problem of high WIP levels can be dealt with by reducing utilization (adding more capacity), reducing variability or a combination of both.

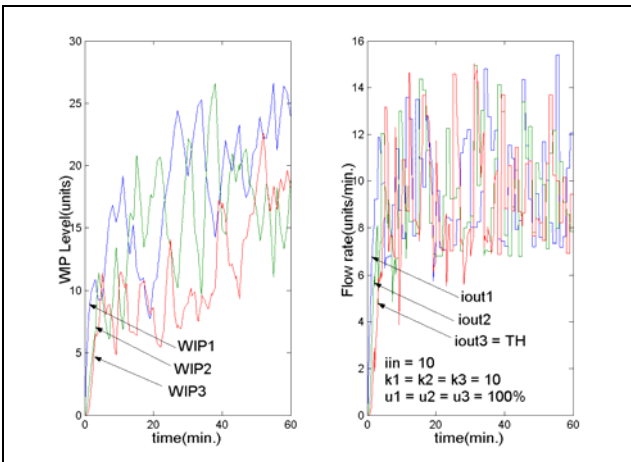


Figure 11: Stochastic ($cv = 0.25$) Three Station System without a Bottleneck

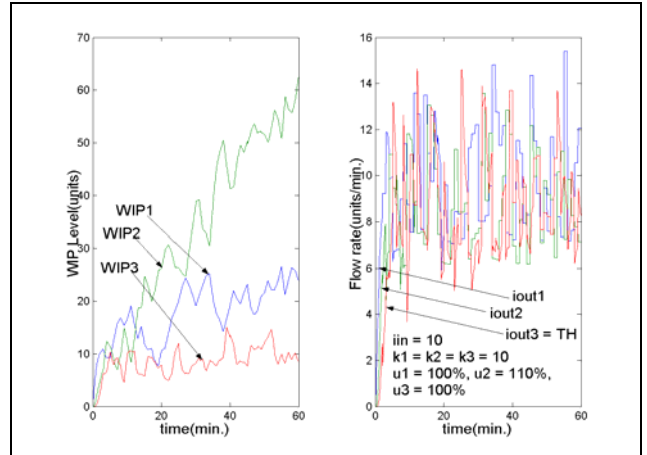


Figure 12: Stochastic ($cv = 0.25$) Three Station System with a Bottleneck

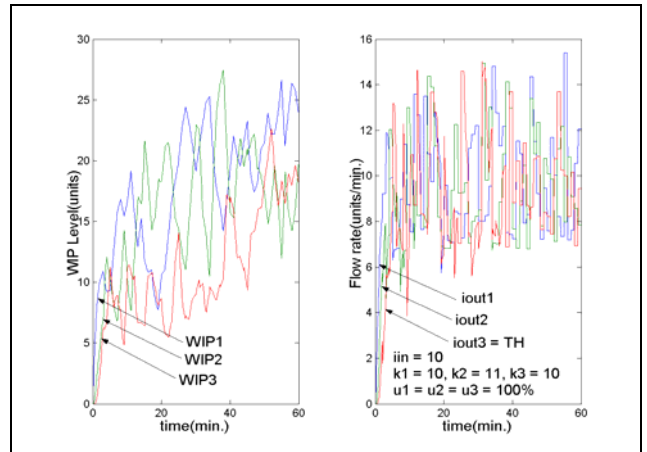


Figure 13: Stochastic ($cv = 0.25$) Three Station System with the Bottleneck Relieved

Although no closed-form solutions are found for the stochastic case, general trends (like the average rate of WIP increase) can be predicted from the corresponding deterministic case closed-form solution. For example, note that the average slope of WIP_2 increase in Figures 9 and 12 is the same as that found in Figure 6.

The same stream of random numbers was used in each of the simulations whose results are shown in this section. This was to make sure that differences in the results were due to the changes in system parameters and not a result of using a different set of random numbers. The goal was to see the effect of different levels of variability and not to accurately estimate the WIP levels and flow rates. To estimate the WIP levels and flow rates given any set of system parameters, a number of replications using different streams of random numbers should be used.

In the deterministic case (section 3), it was possible to validate the steady state response by simple calculations, and the results were intuitive. In the stochastic case, the re-

sulting WIP level at each station is not intuitive. So we decided to compare our model to a commonly used simulation tool, namely, ProModel. A formal comparison of discrete event simulation with continuous simulation is beyond the scope of this paper. However, we compared some results from ProModel to results from our model to see how different they would be from each other.

A three-station manufacturing line model was built in ProModel. Each station had 10 parallel machines. The arrival rate was set to 1 unit per 0.1 minute. Note that in discrete event simulation this is not the same as 10 units per minute. Having 10 parallel machines (each with one tenth of the station capacity) instead of a one big machine, and the way arrival rate was defined helps make ProModel behave in a way that is more similar to continuous simulation. As a result of having 10 parallel machines at each station, a variability pooling adjustment was necessary in defining cv at each machine so that the required cv for the whole station can be achieved. We compared the results for different scenarios and two probabilistic distributions, uniform and gamma. The scenarios involved two levels of variability and different utilization levels. Each scenario was replicated 10 times and the averages of replications were used for the comparison. Each replication simulated a period of time of 10 hours. Figures 14 and 15 show the results of comparing WIP levels at each of the three stations. The percentage differences shown were based on using ProModel results as reference values. So a negative per-

centage difference means that our model predicted less WIP at that station, while a positive percentage difference means that ProModel predicted less WIP.

The WIP levels from the two models were compared using a paired t-test at the 95% confidence level. If a percentage difference is bolded in the figures below, then it is statistically significant at 95% confidence level. When the uniform distribution is used to generate te's, Figure 14 shows that all the differences are well below 10%. When the gamma distribution is used, statistically significant differences are up to 22.3%. See Figure 15.

The discrepancies noted in WIP levels between ProModel and our model may result from different sources:

1. The conceptual difference between discrete processing (where each unit has its own processing time) and continuous processing (where there is a flow rate of units). This factor is still important even if there is no variability in the system.
2. In ProModel, a random number is generated to determine the processing time of each individual unit, while in our model a random number determines the processing rate over a fixed period of time, t_u .
3. Since ProModel needs to generate more random numbers than our model, it is not possible to use the same set of random numbers in both models.

The first source of discrepancy can be minimized by increasing the number of parallel machines at each station in

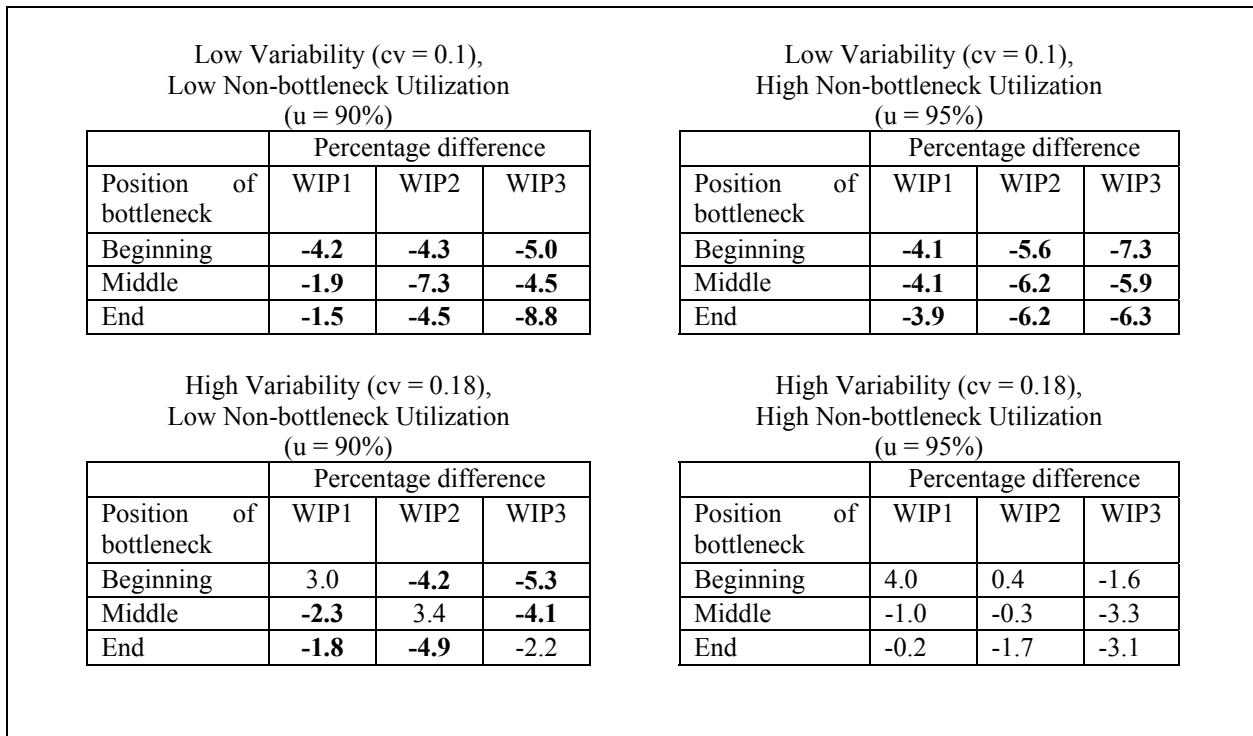


Figure 14: WIP Difference between ProModel and our Model Using Uniform Distribution and a Bottleneck Utilization of 97%

Low Variability (cv = 0.1), Low Non-bottleneck Utilization (u = 90%)				Low Variability (cv = 0.1), High Non-bottleneck Utilization (u = 95%)			
	Percentage difference				Percentage difference		
Position of bottleneck	WIP1	WIP2	WIP3	Position of bottleneck	WIP1	WIP2	WIP3
Beginning	-2.6	-2.9	-2.3	Beginning	-2.9	-5.0	-6.3
Middle	-0.7	-4.7	-2.5	Middle	-3.4	-6.1	-5.5
End	-0.4	-1.7	-7.1	End	-3.0	-6.2	-4.9

High Variability (cv = 0.25), Low Non-bottleneck Utilization (u = 90%)				High Variability (cv = 0.25), High Non-bottleneck Utilization (u = 95%)			
	Percentage difference				Percentage difference		
Position of bottleneck	WIP1	WIP2	WIP3	Position of bottleneck	WIP1	WIP2	WIP3
Beginning	7.8	18.1	14.5	Beginning	-5.5	22.3	22.2
Middle	14.2	25.1	14.4	Middle	13.7	13.7	20.0
End	14.9	19.2	0.8	End	6.8	17.3	13.9

Figure 15: WIP Difference between ProModel and our Model Using Gamma Distribution and a Bottleneck Utilization of 97%

ProModel. However, when we increased the number of parallel machines from 10 to 100, the effect was very small. The second source of discrepancy is inherent to the continuous modeling approach. We found that the choice of t_u has a significant effect on the WIP levels. The mean value of the processing times is used as a reasonable choice for t_u , however, the best choice for t_u depends on the real system being modeled and is subject of further research. The third source of discrepancy can be handled by increasing the number of replications and the simulated operation time. We used 10 replications and simulated an operation time of 10 hours. However, more replications and longer operation times can be used at the expense of longer simulation times.

5 CONCLUSION

This paper presents a dynamic system model of continuous manufacturing systems. It employs electrical analogies and components to build a visually understandable model of continuous manufacturing systems.

Compared to discrete event simulation and queuing theory approaches, the developed model has the advantage of providing physically meaningful equations (which is not the case with discrete event simulation) without limiting itself to any specific probabilistic distribution (as in queuing theory). It also has the advantage of producing closed form solutions that cover both the transient and steady states in the deterministic case. Although the deterministic case is theoretical and

can hardly exist in a real life manufacturing system, its closed-form solution provides a good insight.

In stochastic cases, the model is capable of handling variability in the processing and arrival rates. It should be noted here that the way we used to handle variability in the example mentioned in this paper is not unique. The best way to handle stochastic effects in one real system might not be the best way to handle it in another real system.

REFERENCES

- Baines, T. S., and D. K. Harrison. 1999. An opportunity for system dynamics in manufacturing system modeling. *Production Planning and Control* 10 (6): 542-552.
- Besombes, B., and E. Marcon. 1993. Bond-graphs for modelling of manufacturing systems. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* 3: 256-261.
- de Souza, R., R. Huynh, M. Chandrashekar, and D. Thevenard. 1996. A comparison of modelling paradigms for manufacturing line. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* 2: 1253-1258.
- Ferney, M. 2000. Modelling and controlling product manufacturing systems using bond-graphs and state equations: Continuous systems and discrete systems which can be represented by continuous models. *Production Planning and Control* 11 (1): 7-19.

Hopp, W. J., and M. L. Spearman. 2001. *Factory physics*. Irwin/McGraw-Hill.

Sader, B. H., and C. D. Sorensen. In preparation. Modeling Pull Control Schemes in a Continuous Dynamic Model Using Analogies to Electrical Systems.

AUTHOR BIOGRAPHIES

BASHAR H. SADER is a graduate student in the Mechanical Engineering Department at Brigham Young University. He received a B.S. from Mu'tah University, Jordan in 1997. He received an M.S. from Brigham Young University in 2000. He works on modeling manufacturing systems. He can be contacted by email at [<bashar_sader@hotmail.com>](mailto:bashar_sader@hotmail.com)

CARL D. SORENSEN is an Associate Professor in the Mechanical Engineering Department at Brigham Young University. He received a Ph.D. from the Massachusetts Institute of Technology, and has worked on modeling of manufacturing processes and systems since that time. In addition to the work described in this paper, he is also currently working on modeling a new solid-state joining process, Friction Stir Welding. You can reach him by email at [<carl_sorensen@byu.edu>](mailto:carl_sorensen@byu.edu)