Setting inventory levels of CONWIP flow lines via linear programming

Stefan Helber[a], Katja Schimmelpfeng[b] and Raik Stolletz[c]

[a]Department of
Production Management
Leibniz University Hannover, Germany
[b]Chair of
Accounting and Control
Brandenburg University of Technology, Germany
[c]Department of Management Engineering
Technical University of Denmark, Denmark

September 23, 2009

Abstract

This paper treats the problem of setting the inventory level of closed-loop flow lines operating under the constant-work-in-process (CONWIP) protocol. We solve a huge but simple linear program that models an entire simulation run of a closed-loop flow line in discrete time to determine a production rate estimate of the system. This new approach has been introduced in Helber et al. (2008) for open flow lines with limited buffer capacities. In this paper we present numerical results of the method for closed-loop CONWIP flow lines. The first part of the numerical study deals with the accuracy of the method. In the second part, we focus on the relationship

between the CONWIP inventory level and the short-term profit. In our numerical investigation we consider both limited and unlimited local buffer capacities between the machines.

1 Flow lines with stochastic processing times under CONWIP control

A flow line with **CON**stant **W**ork **I**n **P**rocess (CONWIP) is characterized by a constant number of work pieces (the CONWIP level) circulating in the line. This constant number can be due to a fixed number of pallets or production authorization cards (PACs), see Buzacott and Shanthikumar (1993, pp. 490).

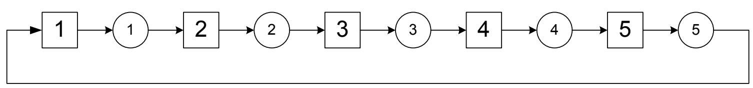


Figure 1: Example of a CONWIP system with 5 stations

Such a CONWIP system is specified by $k = 1, ..., K$ serially arranged work stations $M_1, \ldots, M_K$, each followed by a corresponding (downstream) buffer of size b_k. See as an example in Figure 1 a system with $K = 5$ stations depicted as squares and buffers represented by circles. It is assumed that in front of the first station an unlimited amount of raw material is available. Each work piece on a machine or in a buffer is attached to a pallet so that the total number of work pieces in the system is constant and equals the CONWIP level. When finished work pieces reach the buffer behind the last machine M_K, they are unloaded from the pallets. Then new work pieces are immediately loaded on the pallets to be next processed at the first machine M_1. The transportation times between the work stations as well as the times for (un)loading the pallets are negligible and are assumed to be zero.

In this paper we assume random effective processing times at each station, for example due to manual operations or machine failures. This can lead to blocking and starving and a production rate of the line below the capacity of the bottleneck station if it would operate in isolation. It is both economically important and scientifically challenging to quantify the impact of local buffer sizes and the global CONWIP inventory level on the production rate of the line.

Good surveys about methods for the analysis of flow lines are found in Dallery and Gershwin (1992) and Li et al. (2008). Recent literature surveys of closed loop systems are given in Gershwin and Werner (2007) and Resano and Luis (2008). There is a wide range of applications of closed loop flow lines in manufacturing. For example, Resano and Luis (2008) analyze real automobile assembly lines and preassembly lines as a network of closed loops. Li et al. (2008) give examples of different applications of closed-loop lines with a constant number of carriers for the automotive industry. Hopp and Roof (1998) review different methods of setting the work-in-process (WIP) level in pull systems and analyze a dynamic control of the WIP level to reach a target production rate within a given bound on the cycle-time. Onvural and Perros (1989) approximate the throughput of a CONWIP system numerically and present methods to optimize the CONWIP level.

In general, three approaches for the analysis of stochastic flow lines have been widely established: exact probabilistic analysis, decomposition methods, and discrete-event simulation (DES). Both exact methods and (approximate) decomposition approaches are typically very fast, but also inflexible as they rely on quite specific assumptions about the stochastic behavior of the production system. DES, on the other hand, is extremely flexible, but often requires a substantial computational effort to evaluate a single configuration precisely. Neither method can be easily combined with the powerful optimization methodology of linear programming, see Helber et al. (2008). Our objective in this paper is therefore to close this gap for the particular case of CONWIP flow lines.

The basic idea of our approach originally introduced in Helber et al. (2008) is to approximate the stochastic behavior of a discrete-material flow line operating in continuous time within a large discrete-time linear program (LP). An attractive feature of this approach is the possibility to combine simulation and optimization within a single linear optimization framework. Previous LP-based models of stochastic flow lines were formulated in continuous time, see Abdul-Kader (2006), Johri (1987), Matta and Chefson (2005) and Schruben (2000). Due to the continuous time modeling approach, they could not easily model buffer sizes as decision variables, which is possible in our approach and important in the context of flow line optimization.

The remainder of the paper is organized as follows. In Section 2, a discrete time linear program is developed to evaluate CONWIP systems with a given CONWIP level and either finite or infinite buffer capacities. This evaluation model is extended to an optimization model, where both the CONWIP level and the buffer spaces are decision variables. In both models, the objective is to maximize the respective production rate estimate. Another extension deals with an economic objective function, which is based on gross margins and holding cost per product unit. The numerical studies in Section 3 present results on the accuracy of the method as well as results for the economic optimization of CONWIP lines. In Section 4, we summarize the most important findings of the paper and give an outlook on further research topics.

2 Linear programming modeling of CONWIP flow lines

2.1 Outline of the approach

In our approach the behavior of a discrete-material flow line operating in continuous time is approximated by a linear program (LP) that includes a discrete-time dynamic production-inventory model with continuous production quantities. The number of work pieces that can be processed at a work station of the line during a period (i.e., the production capacity) results from a hypothetical simulation run in continuous time. In this hypothetical run we assume that the work station operates in isolation so that it can neither be blocked nor starved. The realizations of the stochastic processing times are transferred via sampling to realizations of maximum production capacities c_{kt} for each work station k and period t. If one considers a sequence of simulated processing times or durations d_{kw} to process an ordered set of work pieces w at a work station k, one just has to count the number of finished work pieces within period t. An example is shown in Figure 2, see Helber et al. (2008). In the upper part of Figure 2, three discrete time periods 1 to 3 are depicted. In the lower part the durations of seven consecutively processed work pieces are shown. Three work pieces are finished during period 1, one during period 2 and two during period 3. This procedure yields the capacity of the considered work station c_{kt} for the periods 1 to 3 as the realization of a stochastic count process.

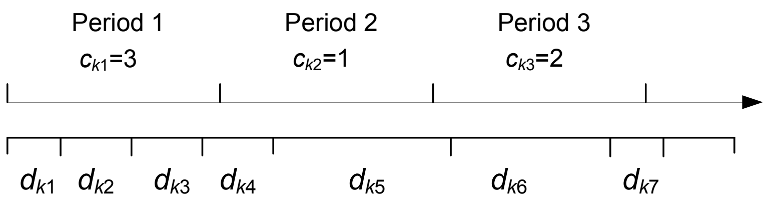


Figure 2: Sampling of discrete-time processing rates

Two conditions must hold so that this discrete-time modeling approach based on production capacities c_{kt} can provide a reasonable picture of the production process in continuous time. Firstly, the sampling frequency must be high enough, i.e., the time periods must be short enough to yield a reasonable representation of the individual processing times. Secondly, the simulation run must be long enough and represent processing of enough work pieces to get a stochastically valid picture of the randomness of the processing times. That means that a substantial number of periods is required within the linear program, each with a specific sampled production capacity c_{kt}.

To explore the accuracy of that approach for flow lines with limited buffer capacity (but without CONWIP control), Helber et al. (2008) analyzed a large and systematically created set of flow lines both via DES and the LP approach. That earlier paper also presents a detailed discussion of the errors induced by simulating a flow line that operates in continuous time within a discrete-time linear program, see Helber et al. (2008). The results showed that the method has a reasonable degree of accuracy unless buffers are very small and/or effective processing times (including possible repair times of unreliable machines) exhibit a high degree of variability of the processing times with a coefficient of variation greater than 1.

2.2 Performance evaluation of CONWIP systems via linear programming

To model a CONWIP flow line system within a linear program, the following assumptions are made:

- A single product type is produced by the flow line.
- The production system contains a cyclic transportation system.
- There is a constant number *pal* of work pieces in the system due to a fixed number of pallets or PACs.
- The production capacity c_{kt} for each station k and period t is a realization of a stochastic count process.
- A production quantity Q_{kt} at a station during a period can either be further processed at the next work station during the next period or be stored in the downstream buffer.
- The buffer behind work station k can hold up to b_k work pieces.
- Transportation times as well as (un)loading times of pallets are negligible.

Note that a CONWIP flow line with limited buffer capacities behaves exactly like one with unlimited buffer capacities if the smallest buffer in the line is large enough to hold all work pieces circulating in the line.

A simple approach to evaluate the performance of such a system using linear programming is to maximize the production rate estimate for a given number of work pieces in the system. The constraints which have to be respected concern the inventory stored in the system, the production quantity (capacity given by the production system), the buffer space, and the number of work pieces used in the system.

Given the notation in Table 1 and the indicator function

Table 1: Notation for the linear program

Indices	
$k = 1, .., K$	Workstations
$t = 1, .., T$	Periods
Input data	
b_k	Number of buffer spaces available behind station k
c_{kt}	Capacity, maximum number of work pieces that can be processed at station k in period t, provided that station k is neither blocked nor starved
pal	Fixed number of work pieces in the system (pallets or PACs)
T_0	Number of warm-up periods
Non-negative decision variables	
PR	Production rate estimate
Q_{kt}	Production quantity of station k in period t
$Y0_k$	Initial inventory behind station k
Y_{kt}	End-of-period inventory behind station k in period t

$$\mathbb{1}_{\{x\}} = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

the linear programming model can be stated as follows:

$$\text{Maximize } PR = \frac{1}{T - T_0} \cdot \sum_{t=T_0+1}^{T} Q_{Kt} \tag{2}$$

with respect to

$$Y0_k \cdot \mathbb{1}_{\{t=1\}} + Y_{k,t-1} \cdot \mathbb{1}_{\{1<t\leq T\}} + Q_{kt} =$$

$$Y_{kt} + Q_{k+1,t+1} \cdot \mathbb{1}_{\{k<K,t<T\}} + Q_{1,t+1} \cdot \mathbb{1}_{\{k=K,t<T\}} \qquad k = 1, .., K;\ t = 1, .., T \tag{3}$$

$$Q_{kt} \leq c_{kt} \qquad k = 1, .., K;\ t = 1, .., T \tag{4}$$

$$Y_{kt} \leq b_k \qquad k = 1, .., K;\ t = 1, .., T \tag{5}$$

$$Y0_k \leq b_k \qquad k = 1, .., K \tag{6}$$

$$\sum_{k \in K} (Y0_k + Q_{k,1}) = pal \tag{7}$$

$$\sum_{k \in K} (Y_{kt} + Q_{k,t+1}) = pal \qquad t = 1, .., T-1 \tag{8}$$

$$Y_{kt}, Y0_k, Q_{k,t} \geq 0 \qquad k = 1, .., K;\ t = 1, .., T \tag{9}$$

$$PR \geq 0 \tag{10}$$

The objective function (2) maximizes the production rate estimate at the last work station K. The production rate PR is determined by dividing the total production of the last work station K in periods $T_0 + 1$ to T by the length of that time span. Equations (3) are classical inventory balance equations. For each work station k and period t, the end-of-period inventory of the previous period $t-1$ plus the production quantity of the current period equals the current end-of-period inventory plus the production quantity of the following period $t+1$ at the next work station. Note that this "next" work station of the last work station K is station 1. Restrictions (4) state that the number of work pieces processed at station k must not exceed the maximum period-specific capacity c_{kt} from the count process described in Section 2.1. The inventory to be stored behind station k must not exceed the number of buffer spaces b_k as stated in Restrictions (5) and (6). Equations (7) and (8) ensure that the number of work pieces within the system meets the given total CONWIP level pal during each period. From a mathematical point of view Equations (8) are redundant because there are already implied by Equations (3) in combination with Equation (7). Considering the case of $k = K$ and $t > 1$, Equations (3) turn to $Y_{K,t-1} + Q_{K,t} = Y_{K,t} + Q_{1,t+1}$. This is equivalent to $Y_{K,t-1} - Y_{K,t} + Q_{K,t} = Q_{1,t+1}$. The left hand side represents the number of finished work pieces which leave the CONWIP line in period t. According to the CONWIP protocol, the same number of work pieces has to be sent to machine 1 in period $t+1$. As we already guarantee the proper initial inventory with Equations (7), Equations (8) are redundant, but help to explain the logic of the model. Last but not least all decision variables must be non-negative, see Restrictions (9) and (10).

2.3 Optimization-oriented models

The basic (evaluation) model presented in Subsection 2.2 can be extended in different ways.

1. In the formulation given above, the number of work pieces *pal* in the system is assumed to be a parameter, determined by the number of pallets or PACs. To be able to optimize this number, the parameter *pal* has to be replaced by a non-negative decision variable PAL in Equations (7) and (8). The resulting new constraints are given in Restrictions (11) to (13).

$$\sum_{k \in K} (Y0_k + Q_{k,1}) = PAL \tag{11}$$

$$\sum_{k \in K} (Y_{kt} + Q_{k,t+1}) = PAL \qquad t = 1,..,T \tag{12}$$

$$PAL \geq 0 \tag{13}$$

   Note that with this modeling approach, a single discrete time simulation run within a linear program can be used to optimize a stochastic CONWIP flow line, here with respect to the production rate.

2. The model presented so far yields a production rate estimate for the system characterized by a given CONWIP level *pal* or a production rate maximizing CONWIP level PAL. In a more economic perspective, it is interesting to find the profit-maximizing number of work pieces in the system. The solution of such a model depends on the cost of capital tied up in the work-in-process and the value of the produced work pieces. Let hc denote the holding cost per product unit and time period and gm the gross margin per product unit. Now the objective is to maximize the profit per time unit which depends on the gross margin of the finished work pieces and the holding cost of all work pieces in the system:

$$\text{Maximize } Profit = gm \cdot (\frac{1}{T - T_0} \cdot \sum_{t=T_0+1}^{T} Q_{Kt}) - hc \cdot PAL \tag{14}$$

The Restrictions (3) to (6) and (9) to (13) remain the same.

3. Similar to the first extension, the model can be used to optimize the distribution of buffers for a given total buffer capacity b_{tot}. The parameter b_k for local buffer capacities in Restrictions (5) and (6) has to be replaced by non-negative decision variables X_k as shown in Restrictions (15), (16) and (18):

$$Y_{kt} \leq X_k \qquad k = 1, .., K; \; t = 1, .., T \tag{15}$$

$$Y0_k \leq X_k \qquad k = 1, .., K \tag{16}$$

$$\sum_{k \in K} X_k = b_{tot} \tag{17}$$

$$X_k \geq 0 \tag{18}$$

Furthermore, Constraint (17) has to be added to guarantee that the total number of buffers allocated behind the different stations k in the system meets the total number of buffers available b_{tot}

Other modifications of this generic model are possible as well.

3 Numerical results

3.1 Accuracy of production rate estimates

3.1.1 Outline of the numerical study

In order to evaluate the accuracy and the numerical effort of our method, we performed a numerical study considering CONWIP lines. The design of this study is based on the results presented by Helber et al. (2008). One result of that paper is that an average processing rate of about 1.0 work pieces per (discrete) time unit combined with a length of a "simulation run" of 10,000 discrete time units yields a reasonable balance between the sampling frequency, the number of sampled events and the size of the linear program embedding such a "simulation run". A further result of that paper is that the method is not accurate for very small buffer sizes and/or coefficients of variation of the effective processing times greater than 1.

As the measure of accuracy in this current study we again use the relative deviation of the production rate estimates of the discrete-time linear program from the "true value" (gained by an extremely long and very precise discrete-event simulation). We compare the results of our method to those obtained from a discrete-event simulation model originally coded in C (Helber (1999)). The LP models were implemented in GAMS and Cplex 11.0.0 was used on a Dual Core Pentium IV machine with 2.8 GHz and 2 GB RAM to solve the models.

We investigate the impact of the following aspects of the problem instances on the accuracy of our method:

- Number of stations
- Number of buffer spaces for each buffer between the machines in the flow line
- Average processing rates at the machines
- Location of the bottleneck (if any) in the line
- Variability of the processing times
- Number of pallets (relative to the number of spaces for pallets in the system)
- Exogenously given even distribution of buffer spaces vs. endogenously determined (production rate maximizing) allocation of buffer spaces.

For reasons of transparency we first discuss these aspects briefly: We expect to find larger deviations for increasing **numbers of stations**, given the results for open lines in Helber et al. (2008). As buffers reduce blocking and starving and our method only approximates the true movement of finished work pieces in Equations (3), we expect to achieve more precise results for problem instances with larger **numbers of buffers**.

The number of periods of the machines in the discrete-time linear program and the **processing rates** determine the total production quantity. To create comparable conditions in our experiments, we set the number of simulated periods (after a warm-up period) and the processing rate of the machines in such a way that comparable expected numbers of approximately 10,000 work pieces could be processed by the line after an initial warm-up period of $T_0 = 500$ periods. Let μ^* denote the rate at which the bottleneck machine of the line operates. Then the number of periods in the discrete time linear program was determined as follows:

$$T = T_0 + T_1 = 500 + \lceil 10,000 \frac{1}{\mu^*} \rceil \tag{19}$$

Given the circular structure of CONWIP flow lines, we expect that the accuracy of the method does not depend on the location of a bottleneck in such a line.

With respect to the **variability of processing times**, we conjecture to find an increasing accuracy with decreasing variability as we did in the study for open flow lines.

As the number of work pieces in a CONWIP line is restricted by the number of pallets, we investigate their influence controlled by a so-called **pallets factor**. It represents the number of pallets related to the number of buffer spaces plus the number of spaces at the work stations. For a given pallets factor pf, the number of pallets in the system is therefore computed as follows:

$$pal = pf \cdot \left(K + \sum_{k=1}^{K} b_k \right) \tag{20}$$

The previous experiments presented in Helber et al. (2008) revealed that the accuracy of the production rate estimates for open flow lines is similar for an exogenously given (even) buffer allocation and an endogenously determined (uneven) **buffer allocation**. We wanted to check if this holds for CONWIP lines as well.

For all the parameter types described, we systematically explored a range of parameter values which we consider to be relevant, to find out under which conditions the method yields reasonably precise production rate estimates.

Table 2: Test Bed for the analysis of CONWIP lines (2430 cases); "f. m." means "first machine", "l. m." means "last machine","o. m." means "other machines"

Parameter type	Number of cases	Parameter value per case
Number of stations	3	5, 7, 9
Buffer spaces per buffer	3	4, 8, 16
Base processing rates	3	0.5, 1.0, 2.0
Bottleneck factor	3	(f. m.: 0.9; o. m.: 1.0), (balanced line, all machines 1.0), (l. m.: 0.9; o. m.: 1.0)
Processing time variability (SCV)	3	0.25, 0.5, 1.0
Pallets factor	5	0.2, 0.35, 0.5, 0.65, 0.8
Buffer allocation	2	even vs. production-rate maximizing

3.1.2 Comparison with continuous time simulation results

To evaluate the performance of our method for CONWIP lines, we used a test bed consisting of all possible combinations of the parameters described in Table 2. We compared the results of our method to those computed with a discrete-event simulation for the test bed consisting of 2430 ($= 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 5 \cdot 2$) cases. Given the results of the method for open flow lines, we considered in this paper minimum buffer sizes of 4 and maximum squared coefficients of variation (SCV) of processing times of 1.0. The last line in Table 2 indicates that for each line we first evaluated a given even distribution of the buffer spaces in the line and then sought the production-rate maximizing buffer allocation as described in Sections 2.2 and 2.3.

Figure 3 shows a diagram with the frequencies of absolute values of relative deviations of production rate estimates obtained by the LP approach from those of the DES. The maximum relative deviation is about 60%, the mean value of the relative deviation is about 5.7%. It also reveals that in 70.78% of the cases there is a deviation of less than 5%. Considering the average of the relative deviation of the production rate estimate as shown in Tables 3 to 8 our method tends to underestimate the production rate. This is a consequence of our discrete time approach which assumes in Equations (3) that work pieces always have to wait for the end of a period to move to the next work station.

To lay open the impact of the parameters listed in Table 2, their effect on the results is shown in the following Tables 3 to 8. We always report

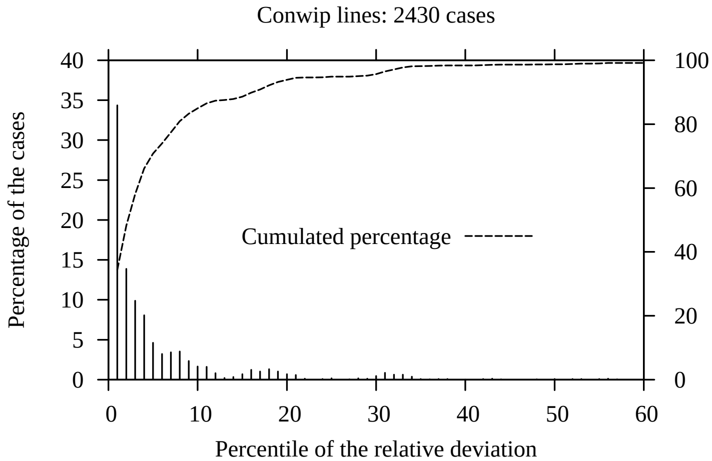


Figure 3: Percentage of cases over relative deviations for all 2430 cases of CONWIP lines

- *RelDev*, the average of the relative deviation of the production rate estimate,
- *AbsRelDev*, the associated average over absolute values of relative deviations, and
- *CPU*, the time (in seconds) to solve the linear program.

The upper part of the Tables is dedicated to the situation in which the buffer allocation is exogenously given and buffer spaces are evenly distributed. The lower part includes the results for the optimized (production rate maximizing) buffer allocation.

Table 3: Impact of the number of stations in the line

Stations	5	7	9
Even buffer allocation:			
RelDev [%]	-5.0	-5.0	-5.3
AbsRelDev [%]	5.4	5.7	5.9
CPU [sec.]	18.7	32.4	49.8
Optimized buffer allocation:			
RelDev [%]	-4.7	-4.8	-5.3
AbsRelDev [%]	5.7	5.9	5.8
CPU [sec.]	30.7	52.2	80.9

Against our expectations the impact of the number of stations in the CONWIP line does not appear to have a strong impact on the accuracy of the production rate estimate. Here the results differ from those for open flow lines in Helber et al. (2008). The results in Table 3 indicate that the CPU time rises as the size of the LP grows with the number of stations.

Table 4: Impact of the number of buffer spaces per buffer

Buffer spaces per buffer	4	8	16
Even buffer allocation:			
RelDev [%]	-10.9	-3.5	-1.0
AbsRelDev [%]	11.7	4.1	1.3
CPU [sec.]	31.4	33.6	35.9
Optimized buffer allocation:			
RelDev [%]	-10.7	-3.5	-0.6
AbsRelDev [%]	11.8	4.0	1.5
CPU [sec.]	53.5	53.7	56.6

The results in Table 4 reveal that the accuracy of the method increases with an increasing number of buffer spaces. The range of chosen parameter values does not seem to have a strong influence on the CPU times.

Table 5: Impact of the base processing rate

Base processing rate	0.5	1.0	2.0
Even buffer allocation:			
RelDev [%]	-6.1	-3.7	-5.5
AbsRelDev [%]	6.5	4.0	6.6
CPU [sec.]	53.0	32.1	15.8
Optimized buffer allocation:			
RelDev [%]	-6.0	-3.3	-5.5
AbsRelDev [%]	6.5	4.3	6.6
CPU [sec.]	92.3	48.4	23.1

Table 5 shows the impact of the base processing rate. For the production of one work piece per time unit, the method appears to yield the best results. The CPU time decreases with increasing base processing rates because the number of periods $(T - T_0)$ after the T_0 warm-up periods decreases (see Section 3.1.1).

Table 6: Impact of the bottleneck location

Bottleneck location	first machine	balanced line	last machine
Even buffer allocation:			
RelDev [%]	-5.0	-5.3	-5.0
AbsRelDev [%]	5.7	5.8	5.7
CPU [sec.]	35.3	30.4	35.2
Optimized buffer allocation:			
RelDev [%]	-4.8	-4.9	-5.1
AbsRelDev [%]	5.8	6.1	5.6
CPU [sec.]	56.8	48.0	59.0

As expected, the location of a bottleneck in a CONWIP line does not seem to have a strong influence on the accuracy of the results, see Table 6.

Table 7: Impact of the squared coefficient of variation

SCV	0.25	0.5	1.0
Even buffer allocation:			
RelDev [%]	-2.9	-4.6	-7.8
AbsRelDev [%]	3.8	5.1	8.3
CPU [sec.]	37.6	33.8	29.5
Optimized buffer allocation:			
RelDev [%]	-2.8	-4.1	-7.9
AbsRelDev [%]	3.7	5.5	8.1
CPU [sec.]	58.2	54.2	51.4

Table 7 reveals the strong impact of the variability of the effective processing times. The lower the SCV, the more accurate the production rate estimates are. The CPU times decrease as the SCV increases. We conjecture that a higher volatility of the sampled production capacities c_{kt} leads to more "extreme" restrictions of the solution space so that the optimum of the LP can be found more quickly.

Table 8: Impact of the pallets factor

Pallets factor pf	0.2	0.35	0.5	0.65	0.8
Even buffer allocation:					
RelDev [%]	-7.7	-3.8	-2.3	-2.6	-9.2
AbsRelDev [%]	8.5	4.0	2.4	3.0	10.5
CPU [sec.]	33.7	34.9	34.9	34.4	30.2
Optimized buffer allocation:					
RelDev [%]	-7.9	-3.8	-2.2	-2.3	-8.4
AbsRelDev [%]	8.4	3.9	2.4	3.3	10.9
CPU [sec.]	49.1	53.8	62.3	56.0	51.8

Table 8 indicates that the relative number of pallets in the system (related to the number of buffers and stations) is very important. In cases with a low (0.2) or a high pallets factor (0.8), the results are less accurate than in the other cases. As this seems to be a major finding of the study we used the method again for a subset of the test bed shown in Table 2. We created this subset by eliminating the two parameter values 0.2 and 0.8 of the pallets factor. The results obtained for the remaining 1458 cases ($= 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 2$) are shown in Figure 4. The method yields much more accurate results for this subset of the test bed. Now the maximum relative deviation is about 20%, the mean value of the relative deviation is 3.18%. In 79.7% of the cases, there is a deviation of less than 5%.

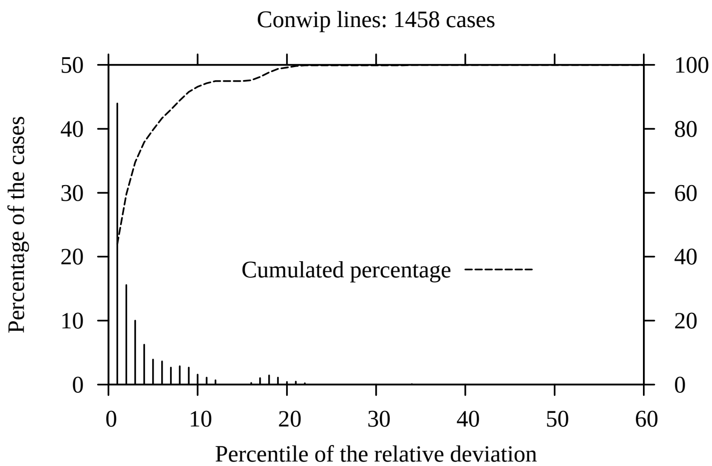


Figure 4: Percentage of cases over relative deviations for the subset of 1458 cases

The results show that our method works well for a wide range of relevant parameter settings.

3.2 Optimizing CONWIP levels via the linear programming approach

To study the problem of optimizing inventory levels with respect to the production rate or the profit, we consider a balanced five-machine CONWIP line as depicted in Figure 1. The gross margin gm per work piece is 100 monetary units and the holding cost hc per unit and (discrete) time period is 1.0 monetary units. The average processing rates of the machines are 1.0 work pieces per time unit and the squared coefficients of variation of the processing times at all machines in the line are either 0.1, 0.5 or 1.0, respectively. With respect to buffer sizes, we consider two cases: In the first case, we assume a buffer capacity of 10 work pieces behind each machine. In the second case, we set all the buffer capacities to 100. The CONWIP level varies from 1 to 50.

If each buffer can hold up to 10 work pieces, blocking can occur for CONWIP levels above 10 work pieces. However, deadlock cannot occur for the assumed maximum CONWIP level of 50 work pieces as machines and buffers can (together) hold up to 55 work pieces. The other case with buffer capacities of 100 work pieces behind each machine and a maximum CONWIP level of 50 work pieces models the infinite buffer capacity case as blocking cannot occur.

For each buffer capacity case and CONWIP level, we determine a production rate estimate via our LP method and via the discrete-event simulation (DES) in continuous time. Based on these production rate estimates for different CONWIP levels, the short-term profit is computed as specified in Equation (14).

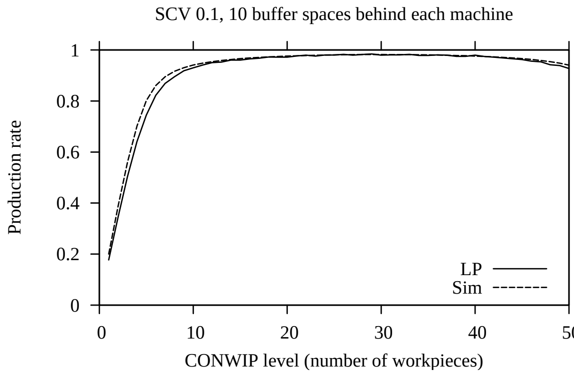
Now we first vary the CONWIP level from 1 to 50 to show the production rate and the profit as a function of the number of pallets. Then we ask how reliably our LP approach can find the number of pallets that maximizes the production rate or the profit. For that purpose, we repeatedly solved the models for different realizations of the simulated processing times.

The graphs for the production rate estimates as determined via our method and via a discrete-event simulation are depicted in Figures 5 to 7. They show that the production rate decreases as processing time variability increases and that CONWIP lines with unlimited buffer capacities are more productive than those with limited buffer capacities. They also show that for peak production rates, the results of the discrete time model (LP) are very close to those of the continuous time simulation (Sim).

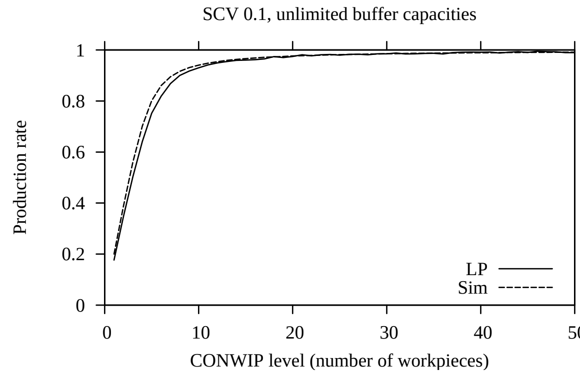
The results for the profit in Figures 8 to 10 show a very similar picture. The CONWIP line reaches its peak profitability in situations where blocking and starving rarely occur. Under these conditions, however, our method is apparently relatively accurate.

Note that the profit functions in Figures 8 to 10 exhibit a (from a practical point of view) extremely nice feature: As the variability of the effective processing time increases, the profit function becomes flatter around its maximum. While our method yields less accurate production rate estimates as the variability increases, the profit estimates are therefore still relatively exact and the solutions can be expected to be close to optimal. It should also be noted that our method always slightly underestimates the profit associated with a given CONWIP level as it tends to underestimate the production rate, see above.

In the last part of the numerical experiment we address the question how reliably our method can find the number of pallets that maximizes the production rate or the profit for a given line without enumerating all pallet levels as shown in Figures 5 to 10. We also want to quantify how precise the production rate or profit estimate is around the "true optimum". For that purpose we studied the same cases that led to the results presented

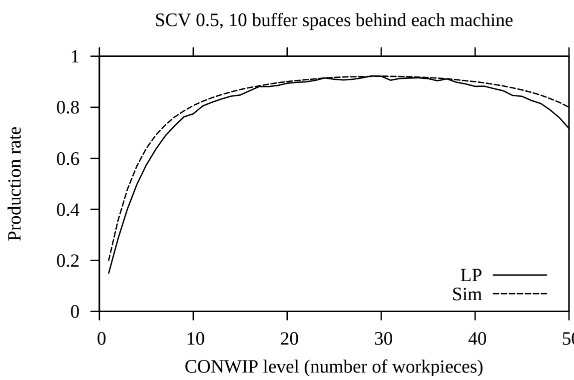


(a) Small buffers

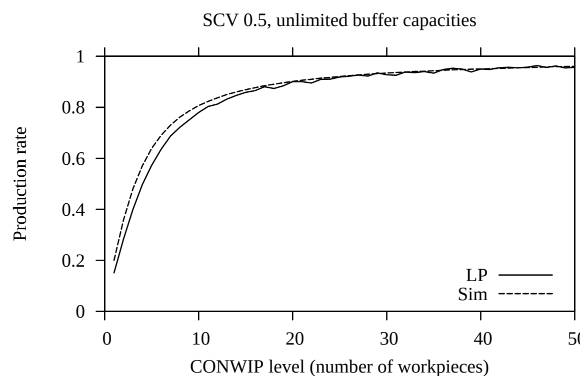


(b) Infinite buffers

Figure 5: Production rate for small resp. infinite buffers and low variability

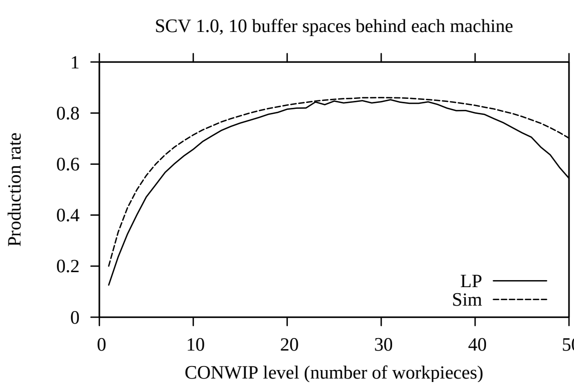


(a) Small buffers

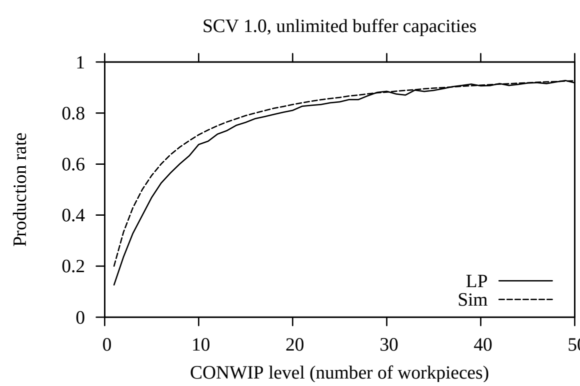


(b) Infinite buffers

Figure 6: Production rate for small resp. infinite buffers and moderate variability

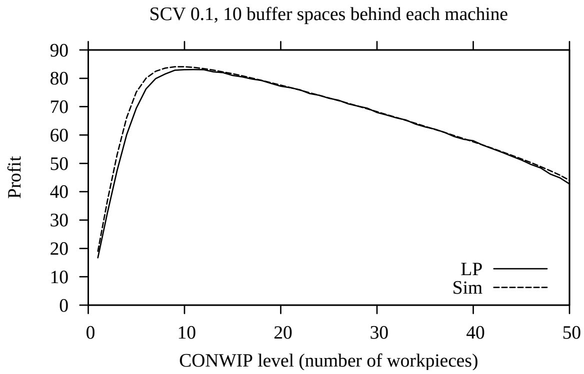


(a) Small buffers

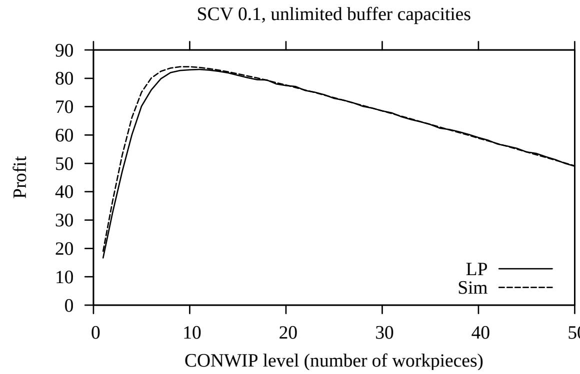


(b) Infinite buffers

Figure 7: Production rate for small resp. infinite buffers and high variability

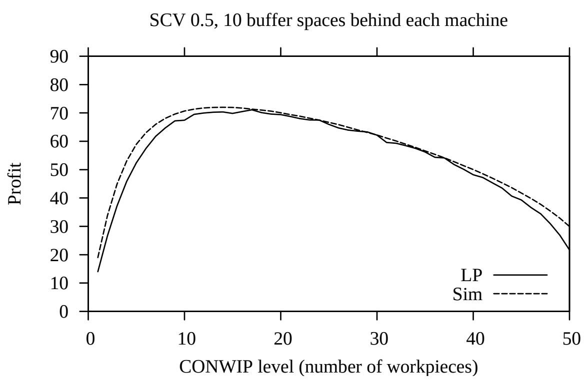


(a) Small buffers

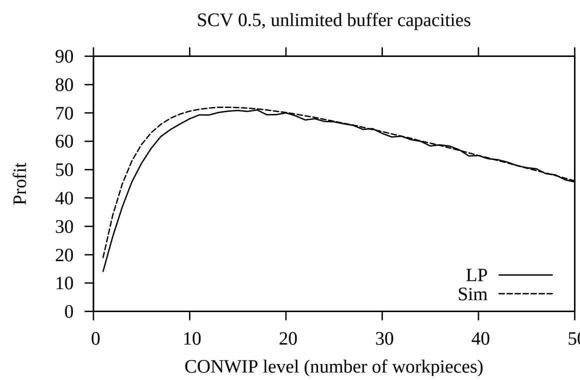


(b) Infinite buffers

Figure 8: Profit for small resp. infinite buffers and low variability

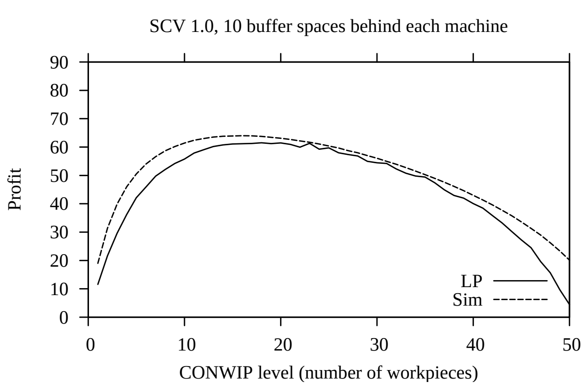


(a) Small buffers

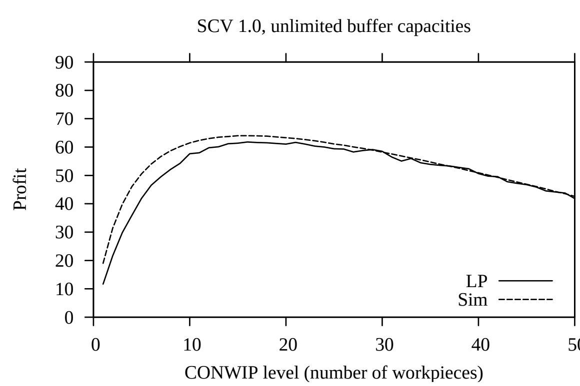


(b) Infinite buffers

Figure 9: Profit for small resp. infinite buffers and moderate variability



(a) Small buffers



(b) Infinite buffers

Figure 10: Profit for small resp. infinite buffers and high variability

Table 9: Production rate maximization for finite buffer cases: Accuracy of the estimated optimal number of pallets

SCV	0.1	0.5	1.0
Continuous time simulation:			
$PR^{Sim}(PAL^{Sim})$	0.982	0.922	0.861
PAL^{Sim}	30	30	30
LP-Approach:			
$PR^{LP}(PAL^{LP})$	0.980 $\cdots$ 0.983	0.912 $\cdots$ 0.925	0.841 $\cdots$ 0.854
$RelDev1$	-0.2% $\cdots$ 0.1%	-1.1% $\cdots$ 0.3%	-2.3% $\cdots$ -0.8%
PAL^{LP}	29 $\cdots$ 31	28$\cdots$ 30	28$\cdots$ 30
$PR^{Sim}(PAL^{LP})$	0.982 $\cdots$ 0.982	0.920 $\cdots$ 0.922	0.860 $\cdots$ 0.861
$RelDev2$	$\approx$ 0%	-0.2% $\cdots$ 0%	-0.1% $\cdots$ 0%

in Figures 5 to 10 and now make the number of pallets PAL a decision variable.

Remember that each single optimization for the linear program is based on different realizations of random variables for production capacities. Therefore, each optimization run leads to different estimates of the production rate and/or profit associated with a particular line. It also leads to different estimates of the respective optimal number of pallets. For this reason, we performed 10 independent optimization runs for each of the six systems and both objectives (production rate or profit maximization), leading to specific estimates of the optimal number of pallets. We then asked

- how strongly the estimated optimum objective function value (from the discrete time LP) deviates from the "true" optimum as obtained via the DES and
- how close the "optimal" number of pallets PAL as determined via the LP comes to the "true" production rate or profit-maximizing number of pallets and
- how much of the true optimum of the respective objective function is sacrificed if the pallet numbers as determined via the LP are implemented.

The results are presented in Tables 9 to 11. Note that maximizing the production rate as a function of the number of pallets is only a reasonable objective for the case of limited buffer capacities. Therefore we don't report results for unlimited buffer capacity.

Table 10: Profit maximization for finite buffer cases: Accuracy of the estimated optimal number of pallets

SCV	0.1	0.5	1.0
Continuous time simulation:			
$Profit^{Sim}(PAL^{Sim})$	84.1	72.0	64.0
PAL^{Sim}	10	14	16
LP-Approach:			
$Profit^{LP}(PAL^{LP})$	82.7 ⋯ 83.4	70.3 ⋯ 71.5	61.5 ⋯ 62.4
$RelDev1$	-1.7% ⋯ -0.8%	-2.4% ⋯ -0.7%	-0.2% ⋯ -2.5%
PAL^{LP}	10 ⋯ 11	15 ⋯ 16	17 ⋯ 18
$Profit^{Sim}(PAL^{LP})$	84.1 ⋯ 83.8	71.9 ⋯ 71.7	63.9 ⋯ 63.7
$RelDev2$	0% ⋯ -0.4%	-0.1% ⋯ -0.4%	-0.2% ⋯ -0.5%

Table 11: Profit maximization for infinite buffer cases: Accuracy of the estimated optimal number of pallets

SCV	0.1	0.5	1.0
Continuous time simulation:			
$Profit^{Sim}(PAL^{Sim})$	84.1	72.0	64.0
PAL^{Sim}	9	13	15
LP-Approach:			
$Profit^{LP}(PAL^{LP})$	82.9 ⋯ 83.3	69.9 ⋯ 71.1	60.8 ⋯ 63.2
$RelDev1$	-1.4% ⋯ -1.0%	-2.9% ⋯ -1.3%	-5.0% ⋯ -1.3%
PAL^{LP}	10 ⋯ 11	all 15	17 ⋯ 18
$Profit^{Sim}(PAL^{LP})$	84.1 ⋯ 83.8	all 71.9	63.9 ⋯ 63.8
$RelDev2$	0% ⋯ -0.4%	all -0.1%	-0.2% ⋯ -0.3%

For each squared coefficient of variation we first report the maximum of the considered objective value from the continuous time simulation, i.e., in Table 9 the production rate maximum PR^{Sim} from the continuous time simulation and the corresponding "true" optimal number of pallets PAL^{Sim}. The lower part of the table reports results from our linear programming approach. We start with the range of production rate estimates PR^{LP} over the 10 independent replications of the LP optimization. The respective range of relative deviations from the true optimum is labeled $RelDev1$. We next present the range of pallet numbers PAL^{LP} that were considered to be "optimal" within the 10 replications of the optimization. In the bottom part of the table we finally report for that range of pallet numbers the corresponding range of production rates $PR^{Sim}(PAL^{LP})$ from the DES and the respective range of relative deviations from the true optimum $PR^{Sim}(PAL^{Sim})$. This last number $RelDev2$ indicates by how many percent we actually miss the optimum value by setting the inventory level PAL via our LP method. The structure of Tables 10 and 11 related to profit maximization is identical. Tables 9 to 11 show that in our examples we never miss the optimum objective function value by more than about 0.5%. This is due to the fact that both the production rate and the profit are flat around the optimum number of pallets, see Hopp and Spearman (2000, p. 358).

Some results from these tables deserve a more detailed discussion. Note that in Table 9, the number of pallets that maximizes the production rate is always 30, for any variability of the effective processing times. This is plausible as in the balanced five-machine line with identical buffer sizes, five work pieces are required at the machines and the remaining 25 work pieces use exactly 50% of the $5 \cdot 10 = 50$ buffer spaces in the line and make blocking and starving of machines equally likely, and hence maximize the production rate.

As we expected, maximum production rates and profit levels decrease as the variability of processing times increases. It is also interesting to note that both the maximum profit and the corresponding number of pallets in Tables 10 and 11 are almost identical. This has a potentially important managerial implication. If pallets are expensive and one seeks a profit-maximizing configuration of a flow line, it might be worthwhile to study the infinite buffer case first and determine an estimate of the optimal number of pallets. This helps to set an upper bound on the number of buffer spaces that may be required between any two adjacent machines and thus speed up the search process for a good buffer allocation.

4 Conclusion and further research

In this paper we analyzed the performance evaluation and optimization of stochastic flow lines under the CONWIP protocol. We used a simple linear program that models an entire simulation run of the closed loop system in discrete time. This way, it is possible to evaluate and optimize the production rate and/or short-term profit of the CONWIP system.

Our approach offers the optimization power of (mixed-integer) linear programming in combination with the flexibility of stochastic simulation with respect to probability distributions of stochastic processing times. It avoids the disadvantages of the established approaches (e.g. time-consuming computation times within DES, special knowledge requirements and restrictive assumptions within queueing models). The accuracy of the method depends especially on the variability of the processing times and the number of pallets in the CONWIP line.

In particular for profit-maximizing CONWIP levels, the approach appears to be remarkably accurate unless buffers are very small and/or effective processing times are highly variably. As hard- and software continue to become more and more powerful, it will be possible to study longer lines and systems with higher degrees of variability of the effective processing times using our method. Our future work will address flow line configuration and design problems from an investment perspective.

References

Abdul-Kader, Walid. 2006. Capacity improvement of an unreliable production line–an analytical approach. *Computers & Operations Research* **33** 1695–1712.

Buzacott, John A., J. George Shanthikumar. 1993. *Stochastic Models of Manufacturing Systems*. Prentice Hall, Englewood Cliffs, NJ.

Dallery, Yves, Stanley B. Gershwin. 1992. Manufacturing flow line systems: A review of models and analytical results. *Queuing Systems Theory and Applications* **12**(1-2) 3–94. Special issue on queuing models of manufacturing systems.

Gershwin, S. B., L. M. Werner. 2007. An approximate analytical method for evaluating the performance of closed-loop flow systems with unreliable machines and finite buffers. *International Journal of Production Research* **45**(14) 3085–3111.

Helber, Stefan. 1999. *Performance Analysis of Flow Lines with Non-Linear Flow of Material*, *Lecture Notes in Economics and Mathematical Systems*, vol. 473. Springer-Verlag, Berlin et al.

Helber, Stefan, Katja Schimmelpfeng, Raik Stolletz, Svenja Lagershausen. 2008. Using linear programming to analyze and optimize stochastic flow lines. Tech. rep., Leibniz Universität Hannover, Institut für Produktionswirtschaft, Königsworther Platz 1, D-30167 Hannover. Available online at http://www.wiwi.uni-hannover.de/Forschung/Diskussionspapiere/dp-389.pdf.

Hopp, W. J., M. L. Roof. 1998. Setting WIP levels with statistical throughput control (STC) in CONWIP production lines. *International Journal of Production Research* **36**(4) 867–882.

Hopp, Wallace J., Mark L. Spearman. 2000. *Factory Physics. Foundations of Manufacturing Management*. 2nd ed. Irwin, Chicago et al.

Johri, Pravin K. 1987. A linear programming approach to capacity estimation of automated production lines with finite buffers. *International Journal of Production Research* **25** 851–866.

Li, D. E., J.and Blumenfeld, N. Huang, J. M. Alden. 2008. Throughput analysis of production systems: recent advances and future topics. *International Journal of Production Research* **DOI: 10.1080/00207540701829752**.

Matta, A, R. Chefson. 2005. Formal properties of closed flow lines with limited buffer capacities and random processing times. J. Manuel Felix-Teixera, A. E. Carvalho Brito, eds., *The 2005 European Simulation and Modelling Conference*. Porto, 190–198.

Onvural, R. O., H. G. Perros. 1989. Approximate throughput analysis of cyclic queueing networks with finite buffers. *IEEE Transactions on Software Engineering* **15**(6) 800–808.

Resano, A., C. J. Luis. 2008. Analysis of an automobile assembly line as a network of closed loops working in both, stationary and transitory regimes. *International Journal of Production Research* **46**(17) 4803–4825.

Schruben, Lee W. 2000. Mathematical programming models of discrete event system dynamics. J. A. Joines, R. R. Barton, K. Kang, P. A. Fishwick, eds., *Proceedings of the 2000 Winter Simulation Conference*. 381–385.