Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

HOW FACTORY PHYSICS HELPS SIMULATION

Charles R. Standridge

School of Engineering
Grand Valley State University
301 West Fulton
Grand Rapids, MI 49504, U.S.A.

ABSTRACT

Factory physics provides a systematic description, expressed as laws, of the underlying behavior of a system. These laws can provide important assistance in performing simulation studies. They help in deciding what performance measures to collect and what alternatives to evaluate as well as in interpreting simulation results. The laws help identify the properties of systems that may be important to include in models. They provide an analytic foundation that helps in understanding the behavior of systems as well as giving insight into the types of issues addressed in simulation studies. Verification and validation evidence can be collected based on these laws. This paper examines the application of specific factory physics laws to the activities of a simulation project. Examples showing the application of these principles in industrial projects, masters level student projects, and application studies used in undergraduate and graduate simulation classes are given.

1 INTRODUCTION

Analytic models and simulation models are two approaches to explaining and understanding the behavior of systems in general as well as manufacturing systems in particular. Analytic models include relationships between components of a system expressed in mathematical form. Building an analytic model often involves conceptualizing the system in a predefined structure with some details omitted. Solution of such models typically yields information about long term or steady state average behavior. Analytic models and their solutions provide, at least, a starting point for gaining valuable information about system structure and behavior even if additional information is required. These models help in gaining explicit mathematical insight into the cause and effect relationships that can govern, or at least influence, the behavior of a system.

Simulation models track system behavior over time. They can include mathematical relationships and logical relationships (if-then-else). Simulation models place operational details precisely in time and allow their transient effects to be observed. A simulation model can be used to determine any computable measure of system performance. Simulation models can include all details of system behavior.

Analytic and simulation models can each be used in studying the same kinds of systems and addressing the same kinds of issues. Analytic models may be faster to build and yield exact mathematical solutions. Simulation models can accommodate details analytic models cannot and meet project specific requirements for information.

Thus, it seems useful to explore the ways in which simulation and analytic models should be used together for problem solving. One set of analytic models and related principles was developed by Hopp and Spearman (2000) and is referred to as factory physics. Some of these laws are expressed as equations while others state general principles concerning how production systems operate. The concurrent use of factory physics laws and simulation for problems solving is the subject of this paper.

Each factory physics law that experience has shown to be helpful in conducting a simulation project is discussed in turn. Examples of the use of each law within a simulation project are given. When not restricted by the confidentiality of industrial applications, quantitative results are shown. Factory physics laws are shown to impact many aspects of a simulation study including verification and validation, performance measure selection, alternative selection, and the scope of models.

2 CONSERVATION OF MATERIAL LAW

The conservation of material law is stated as: In a stable system, over the long run, the rate out of a system will equal the rate in, less any yield loss. This law can be expressed in the form of Equation (1).

$$\text{Rate}_{\text{out}} = \text{Rate}_{\text{in}} + \text{Rate}_{\text{Loss}} \qquad (1)$$

This law has application in a simulation project as a way of obtaining verification and validation evidence. It may be restated for this purpose as: The number of entities created during a simulation run must equal the number destroyed plus the number remaining in the system at the end of the run. The number of entities created includes those initially in the system at the start of the run, that is the initial conditions. This idea is also expressed in Equation (2).

$$\text{Created} + \text{Initial} = \text{Destroyed} + \text{Remaining} \qquad (2)$$

An entity is a modeling construct that represents the part, customer, information, etc. that is acted upon or transformed by processing done by a system. Thus, the conservation law provides a balance equation that when shown to be true may provide evidence that a model is correctly implemented and valid.

Consider a simple system: two workstations in a sequence with a buffer preceding each workstation. Entities arrive to the buffer preceding the first station and are destroyed after processing at the second station. Thus, in each replicate of a simulation of this system the number of entities arriving plus the number initially present must equal the number destroyed plus the number in the system at the end. This latter quantity is equal to the number in each buffer plus the number in processing at each station.

Note that this approach involves specifying each place where an entity could be. This helps verify that entities are moving correctly through the model and not being misrouted or otherwise lost.

It is possible for the conservation law to be obeyed and evidence that the model is incorrectly implemented or invalid obtain. For example, for the two workstations in a series system suppose that a simulation run had 100 entities arriving, 10 destroyed and 90 remaining in the system at the end. The fact that 90% of the entities remained in the system at end indicates a problem in processing, perhaps a lack of capacity at one of the workstation.

This law can be applied to other types of systems as well as more complex systems. Consider the production and inventory management system described by Maas (2003) as shown in Figure 1. The numbers in parentheses indicate the batch size.

There are 26 products produced by this system using a single foam line and multiple injection modeling machines. Each product has its own inventory. Customer demands are filled from the inventory. When the inventory level falls below a pre-specified target value, an order is placed with the production system to restore the inventory level to the target value.

In this case, the conservation law is applied separately to each product and may be rewritten as: beginning inventory + production volume = shipments to customers + ending inventory. The use of the conservation law helps assure that the conversions between batch sizes between each component of the system were done correctly.

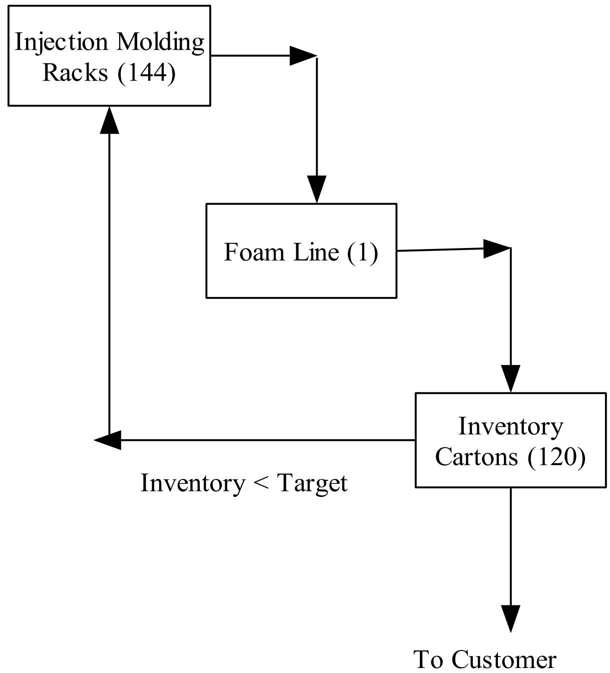

Figure 1: Production and Inventory System

3 STATION CYCLE TIME DEFINITION

The station cycle time definition of factory physics states that average cycle time at a station can be obtained by adding the average time spent in each component of the operation of a station such as movement to the station, time in the buffer, setup time, processing time and so forth.

This law can be restated for use in simulation projects as: total time at station is equal to the sum of the waiting time, processing time, and other non-productive time. The latter term usually involves the unavailability of a resource that is being repaired or busy performing a task at another station. This term is particular onerous and simulation is uniquely able to evaluate it. This is expressed in Equation (3).

$$\text{Station Time} = \text{Waiting Time} + \text{Processing Time} + \text{Non-Productive Time} \qquad (3)$$

As an example, consider a work cell consisting of multiple, semi-automated stations. The principle of one-piece flow is used to operate the cell. Each station has three processing steps: load on a machine with worker assistance, process without worker assistance, and unload from the machine with worker assistance. Non-productive time for a machine occurs when a part is ready to be loaded or unloaded but no worker is available to do the task. Non-productive time for a worker includes the time to walk from station to station without moving a part.

Non-productive time effectively reduces the capacity of a station even to the point of making systems operation infeasible. Infeasible means that the effective processing rate, reduced from the nominal rate by non-productive time, is less than the arrival rate.

Only simulation can be used to quantify the non-productive time. Alternatives that potentially reduce the non-productive time, such as alternative worker priorities for multiple waiting tasks, can be evaluated.

4 LITTLE'S LAW

In factory physics, Little's Law (Little 1961) is stated as Equation (4):

$$TH = {}^{WIP}\!/_{CT} \tag{4}$$

or in words: throughput is equal to work in process divided by cycle time in the long term. In the factory physics context, cycle time is defined to be the time between entering the system and departing the system.

Little's Law can be used in the following ways in a simulation study. Throughput is generally thought of as the rate at which entities complete processing. However, by the conservation of material law throughput is also the rate at which entities arrive, usually expressed as the time between arrivals. Thus, throughput is not an effective performance measure for most simulation studies. The clear exception is when the purpose of the study is to evaluate whether a system has sufficient capital equipment or other resources to do all of the work it is assigned to do as was illustrated in the discussion of the station cycle time definition.

Cycle time, or time in the system, and WIP are common simulation experiment performance measurement. However, the ratio of WIP to cycle time is a constant in the long term. Thus, the information provided by WIP and cycle time for conclusion drawing is equivalent.

Little's Law can be employed in verification and validation. Changes to a model intended to reduce WIP should also reduce cycle time. If not, there is evidence that the model is not valid. The ratio of WIP to cycle time should be approximately equal to the arrival rate. If either of these is not true, there is evidence that the model is not implemented correctly or is not valid.

5 VARIABILITY BUFFERING

The variability buffering law may be stated as follows: variability will be buffered by some combination of inventory, capacity, and lead time. Spearman (2003) has stated that one meaning of "lean" is that a system has minimal variability buffering.

The need for variability buffering results only from variability in a system. There are two types of variability: random and structural. Sources of random variability are typically modeled as random variables and include quantities such as times between entity arrivals, operations times, customer demands for product, and times between breakdowns. Structural variability arises whenever something is not done in the same amount of time or the same way all the time even if no random variables are involved. For example, a machine processes two types of parts, one in exactly one minute and the other in exactly two minutes.

Simulation helps evaluate system behavior when all the "bad" events happen at once. For example, a production line goes down due to a random failure immediately after a shutdown for maintenance when the inventory of the finished goods it produces is low and random customer demand is near the maximum.

Consider a chemical plant production system of the kind described by Standridge and Heltne (2000). Product is made in a continuous reactor and flows to a mixing tank. When the mixing tank is full, product is transfer to a shipping tank, if space is available. Customer demands are filled from the shipping tank by trucks that are filled from a single load spot. This system is shown in Figure 2.

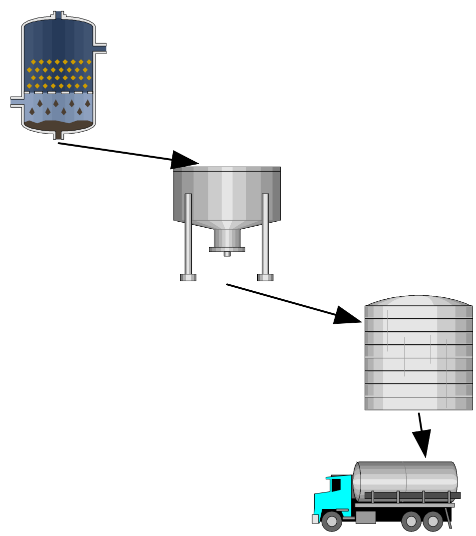

Figure 2: Continuous Production System

Variability buffering is provided by the shipping tank. Random variability is due to events that changes flow rates from the reactor over a wide range as well as unscheduled load spot downtimes. Structural variability is due to scheduled flow rate changes as well as to a mismatch between schedules for production (every day) and shipping (5 days per week). Simulation was used to evaluate the effect of the reduction in structural variability due to shipping every day on the required capacity of the shipping tank (the variability buffer).

6 PRACTICAL WORST CASE DEFINITION

Hopp and Spearman (2000) define minimally acceptable system behavior. This practical worst case can be viewed as the maximum randomness case, that is an entity is equally likely to be at any workstation. Assuming that all workstations have the same average process time and each can work on one entity at a time, this definition leads to processing times as well as the time between arrivals being exponentially distributed.

This definition has implications for assumptions about the distribution of times between arrivals, and less commonly, operation times made in building simulation models. Use of the exponential distribution is seen as a conservative assumption that may add more variability to a model than resides in the corresponding system and thus could result in concluding that a larger variability buffer than necessary is needed. Conducting experiments using other distributions with less variability or constants may be necessary for comparison purposes. This helps ensure that conclusions are not overly dependent on the distribution assumptions that were made.

7 UTILIZATION LAW

The utilization law is as follows: If a station increases utilization without making any other changes, average WIP and cycle time will increase in a highly non-linear fashion. This law is embodied in the VUT equation which can be stated as Equation (5) for the case of one machine at a station:

$$CT_q = \frac{\left(c_a^2 + c_e^2\right)}{2} * \frac{u}{1-u} * t_e = V * U * T \qquad (5)$$

where CT_q is the average cycle time in the queue or buffer, c_a is the coefficient of variation of the time between arrivals, c_e is the coefficient of variation of the processing time (the standard deviation divided by the mean), μ is the utilization (percent busy time) and t_e is the processing time. Note that the average number of entities in the queue can be computed using Equation (5) and Little's Law.

In simulation studies, this equation is useful for the relationships it expresses. The V term has to do with the average of the squared coefficient of variation of the time between arrivals and the processing time. The exponential distribution, often used to model the time between arrivals and corresponding to the practical worst case, has $c_a = 1$. For other distributions, such as those used to model processing times, $c_e < 1$. If the processing time is constant or has little variation ($c_e << 1$), then the V term can be dominated by c_a. Thus, it is important to model random quantities precisely with regard to the mean and standard deviation of the distribution employed. Failure to do so may lead to imprecise estimates of performance measures such as cycle time and WIP as is also discussed by Law and McComas (2003).

Figure 3 shows a graph of the U term versus the utilization to show the non-linear increase in cycle time as the utilization increases.

Suppose one goal of system operation is to keep the cycle time less than a predefine quantity at least a given percent of the time. Meeting this goal may require additional capacity to keep the utilization sufficiently low. Assessing the need for such additional capacity is a common objective of a simulation study. Alternatively, suppose it is an operational goal to have high utilization. This implies that the V term must be small enough to keep the cycle time low by undoing the effect of a large U term. Thus,

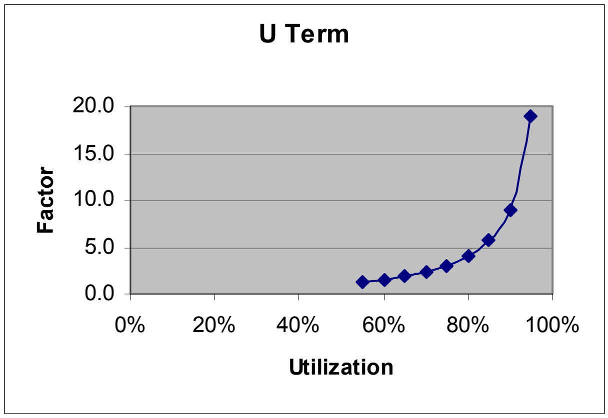

Figure 3: U Term versus Utilization

variation in processing time and the time between arrivals must be controlled.

Consider the logistics system discussed in Standridge and Heltne (2000) and shown in Figure 4.

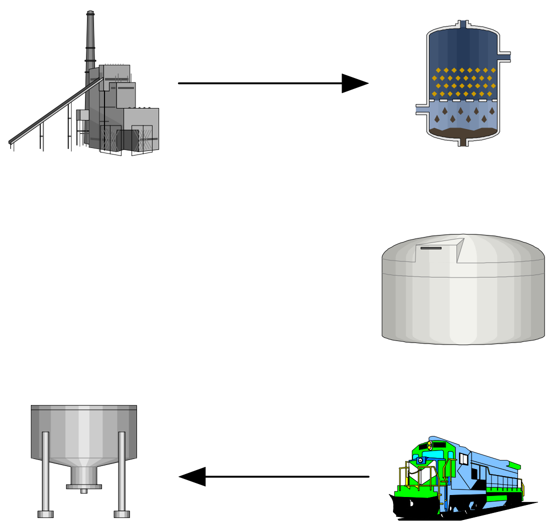

Figure 4: Logistics System

Product is made in a reactor and stored in a tank. From the tank, product is loaded in rail cars for shipment to customers. After rail cars are unload at the customer site, they return to the plant to be reloaded.

Based on the relationships expressed in Equation (2), a utilization standard of 75% was established for the rail car loading facilities. This was done to aid the plant in meeting its overall cycle time standard for loading product shipments bound for customers. Furthermore, rail shipment times have large variance. Thus, additional rail cars and rail yard capacity is needed to keep the utilization of rail cars and rail yards low enough to prevent long waiting times for these resources. These insights based on Equation (2) were confirmed and quantified in detail using simulation.

8 CONWIP LAWS

ConWIP stands for Constant WIP. ConWIP means that the maximum amount of WIP allowed at a workstation or a group of workstations is a pre-specified parameter. Hopp and Spearman (2000) argue that kanban systems are a special case of ConWIP systems. ConWIP and kanban systems are pull systems where product is produced in response to customer demand as opposed to push systems where production is scheduled to meet anticipated demand.

Most simulation languages are oriented to modeling push systems while most modern manufacturing systems are pull systems. Thus, modeling pull systems in a push system framework must be accomplished.

One way of accomplishing this for a ConWIP system is as follows. Arrivals correspond to customer demands for product that is removed from an inventory. This triggers demand for additional production. However, to start production requires permission from the ConWIP control system. This permission is modeled as a resource with a number of units equal to the number of entities that can be concurrently processed.

Grimard (2003) used these ideas to model and simulate the production and calibration system shown in Figure 5.

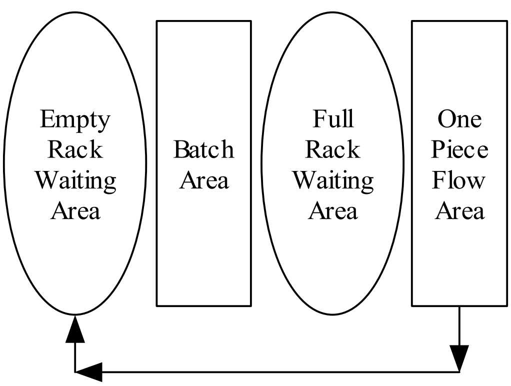

Figure 5: Production and Calibration System

ConWIP controls are implemented using WIP racks. One production batch in the batch area fills one rack that must be available before production can begin. Only one rack at a time may be in the one piece flow area where calibration is performed.

Resources are used to model WIP racks and to control the number of WIP racks in the one piece flow area. Simulation experiments were used to determine the minimum number of WIP racks needed to avoid constraining throughput. This quantity was determined to be 14 which was in agreement with the initial system design. Animation was used to check that the one piece flow logic was correctly implemented in the model and would prove effective in operating the system.

9 SUMMARY

Factory physics laws aid in performing simulation projects. These law help with diverse activities such as validation and verification; defining performance measures and interpreting results; and defining alternatives for evaluation. Modeling issues are raised by the need to model ConWIP and other pull systems.

Teaching factory physics and simulation concurrently in the context of relevant applications in manufacturing and other areas seems worthy. Such courses are under development and will be taught in the near future. Standridge (2004) is developing a text book that helps serve this purpose.

REFERENCES

Hopp W. J. and M. L. Spearman. 2000. *Factory physics, 2nd edition*. Boston, MA: McGraw-Hill.

Grimard, C. 2003. EMD assembly line. Unpublished masters project, School of Engineering, Grand Valley State University, Grand Rapids, Michigan.

Law, A. M. and M. G. McComas. 2003. How the Expertfit distribution-fitting software can make your simulation models more valid. In *Proceedings of the 2003 Winter Simulation Conference*, ed., S. Chick, P.J. Sanchez, D. Ferrin, and D. J. Morris, 169-174. Institute of Electrical and Electronics Engineers, Piscataway, NJ.

Little, J. D. C. 1961. A proof of the queueing formula $L = \lambda W$. *Operations Research* 9 (3).

Mass, S. 2003. A procedure for multi-product scheduling and inventory management. Unpublished masters project proposal, School of Engineering, Grand Valley State University, Grand Rapids, Michigan.

Spearman, M. 2003. Measures, models, and factory physics. Presentation to the Michigan Simulation Users Group 2003 Annual Conference. Available online via <www.m-sug.org> [accessed August 6, 2004].

Standridge, C. R. and D. R. Heltne. 2000. An MSE-based simulation capability for strategic and tactical logistics. In *Proceedings of the 2000 Winter Simulation Conference*, ed., J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 1107-1113. Institute of Electrical and Electronics Engineers, Piscataway, NJ.

Standridge, C. R. 2004. *Simulation in practice: an introduction*, under development.

AUTHOR BIOGRAPHY

CHARLES R. STANDRIDGE is a professor in the School of Engineering, Padnos College of Engineering and Computing, at Grand Valley State University. He has 30 years of simulation experience in academia and industry. He has performed many simulation applications, developed commercial simulation software, and taught simulation at three universities. His current research interests are in the development of modular simulation environments (MSE). He is working with industry on the application of MSE to supply chain management problems emphasizing strategic and tactical logistics as well as inventory management. His teaching interests are in the concurrent use factory physics, lean manufacturing, six sigma and simulation in introductory undergraduate and graduate courses using a case-based approach. He also teaches in the area of engineering measurement and data analysis. He has a Ph.D. in Industrial Engineering from Purdue University. His email address is <standric@gvsu.edu>.