Doctoral Thesis Proposal

Performance analysis of manufacturing systems under hybrid control policies

Asbjoern M. Bonvik
September 22, 1995

1 Introduction

The purpose of the proposed research is to gain better understanding of the performance of repetitive manufacturing systems that are operated under a wide class of control policies. This class contains classical control disciplines like basestock, kanban, and CONWIP, as well as hybrid policies that combine aspects of these classical approaches.

As a result of this understanding, I will construct tools for fast performance assessment of manufacturing systems operated by these policies.

This proposal has the following structure: I first survey some of the previous work that has major relevance for this project, before outlining the steps of the proposed research. Finally, I present some preliminary results to demonstrate the feasibility of the proposed research.

2 Previous work

2.1 Control policies

The simplest control policy for a repetitive manufacturing system is to let it run at full speed. For a tandem system with finite buffer capacities between the stages, this leads to a *flow line* or *transfer line* model. It is usually assumed that the first stage is never starved of raw materials and the last stage is never blocked. That is, the line produces to an infinitely large stock or is trying to keep up with an unlimited demand. For a given system, the control parameters of this policy, even if they are not usually acknowledged as such, are the capacities of buffers along the line. This control policy reduces to a simple rule for each machine: Produce whenever there are parts to process and a place to put the finished product. There are several variations of this control policy, such as blocking before service and blocking after service. See Dallery and Gershwin (1992) for a review.

If the last machine can be blocked, for instance by servicing a demand rate that is less than the production capacity of the line, the above control policy becomes a variation of *kanban control*. The essence of kanban control is that parts are only made as replacements for parts consumed by the downstream process. This can either be the demand process or the next downstream machine in a line. There are a multitude of kanban variations, see Berkley (1992) for an overview. Two main classes of kanban systems are those that are equivalent to tandem queues of finite capacity (Berkley 1991), like the flow line models just described, and those that conform to a different model called *minimal blocking* (So and Pinault 1988). It is interesting to note that even if kanban controls are usually associated with so-called just-in-time production systems (Womack, Jones, and Roos 1990; Monden 1983), it actually requires some amount of inventory between each production stage, since it is the consumption of these inventories that signals upstream stages to produce.

A classical Western approach to controlling a repetitive manufacturing system is *basestock control* (Kimball 1988). This control passes a work order to every manufacturing stage in the line whenever a demand is received. Initial inventories in each buffer (the base stocks) control how much material is held in the line when waiting for another demand. It can be shown that this is an optimal control policy for an uncapacitated manufacturing system (Clark and Scarf 1960). However, in a two-machine line with finite capacities and unreliable machines, Veatch and Wein (1994) demonstrated that the choice between a basestock or a kanban control depends on the location of the system bottleneck. If the upstream machine is slower, basestock control is preferred, otherwise kanban control is better. This seems to be due to the different information flows in these two control disciplines: Basestock control transmits demand information to all machines simultaneously, while kanban control lets the information percolate up the line one stage at at time.

It is interesting to note that for a single stage production system, kanban and basestock control become equivalent since both attempt to maintain a certain inventory ahead of demand. It is also interesting to note that this is also the same as the *hedging point policy* that can be derived from an optimal control viewpoint (Bielecki and Kumar 1988). The relationship between basestock and hedging point control can be extended further: For a multiple stage system, the base stock level of a buffer is equal to the difference in hedging points of the machines on either side of the buffer. A difference is that the hedging point policies are derived for continuous material systems serving a constant demand rate, while the basestock control is defined in terms of discrete material and some stochastic demand process.

Another control policy, CONWIP, was proposed by Spearman, Woodruff, and Hopp (1990). This policy fixes a population constraint for the system, and when this constraint is reached, orders are only released to the system when a finished product departs. This control can be seen as a single kanban cell encompassing the entire manufacturing system.

Recently, several authors have proposed control schemes that are combinations and supersets of the control policies mentioned above. One general scheme is the *production authorization card* control proposed by Buzacott and Shantikumar (1992), which contains all the above mentioned policies as special cases. Similar frameworks have more recently been proposed by Frein, Di Mascolo, and Dallery (1994) and Dallery and Liberopoulos (1995). The latter is essentially a superposition of the basestock and minimal blocking kanban controls.

Exact optimal control policies are not available for multi-stage production systems with unreliable capacitated machines and finite buffer capacities. A near-optimal version of the hedging point policy based on a hierarchy of events at different timescales was developed by Kimemia and Gershwin (1983), and implemented in a discrete event simulation model by Violette (1993). Bai (1991) proposed an approximate algorithm to simultaneously find good hedging points and buffer sizes from a non-linear optimization model representing a semiconductor factory.

Close approximations to the optimal control can be found numerically for small systems. Studies of two-machine systems (Van Ryzin, Lou, and Gershwin 1993; Veatch and Wein 1994) reveal some qualitative properties of the optimal control. Consider a two-stage tandem line with a buffer between the stages and a finished goods buffer separating the line from the demand process. Figure 1, adapted from Van Ryzin, Lou, and Gershwin (1993), shows an optimal control policy in the space defined by the level x of the finished goods buffer and the level y of the intermediate buffer. The policy is to produce at full speed at both machines whenever the system state is inside the solid *switching curve*, then, as the switching curve is reached, to adjust the machine speeds as appropriate to follow the curve to the point marked by a black disk. Thereafter the system produces at the demand rate,

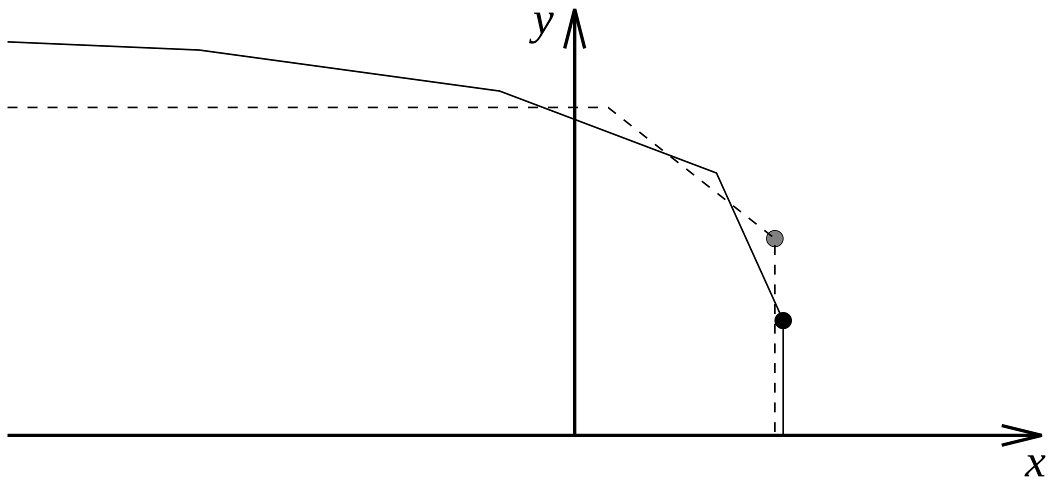

Figure 1: An optimal control policy and its two-boundary approximation

so as to remain at this point.

Two observations are important to make: This control policy is hard to implement in practice, since the optimal size of the intermediate and finished goods buffers depend on the level of the other buffer. And, as the dashed line shows, this control policy can be closely approximated by a *two-boundary control policy* where each machine has a hedging point and the machines are separated by a buffer of constant size. The horizontal part of the switching curve for the two-boundary policy is such that the intermediate buffer is full, the vertical part is such that the second machine is at its hedging point (or equivalently, the finished goods buffer is full), and the 45 degree line is where machine 1 is at its hedging point. The gray disk is the system resting state, where both machines are at their hedging points and produce to demand rate, while the buffer level is constant at the base stock level. This control can then be expressed as a superposition of the basestock and finite buffer kanban controls. The two-boundary control policy was first proposed by Van Ryzin (1987).

2.2 Performance analysis

In general terms, performance analysis has lagged behind the introduction of new control disciplines. The general purpose tool for performance analysis

Policy	Service level	Inventory
Kanban	0.99916 ± .00006	15.82 ± .05
Minimal blocking	0.99909 ± .00005	15.62 ± .05
Basestock	0.99918 ± .00006	14.60 ± .02
CONWIP	0.99922 ± .00005	14.59 ± .02
Hybrid	0.99907 ± .00007	13.93 ± .03

Table 1: Performance of best configurations

remains discrete event simulation, which has the inherent disadvantage of very large computational requirements to give reliable results. One recent simulation study (Bonvik, Couch, and Gershwin 1995) considered a variety of control policies for a subassembly production line at an automobile assembly factory. This study used a lost sales demand model for the assembly line behavior (if the line stocks out, it stops), and posed a very high service level requirement on the subassembly line. As table 1 shows, a CONWIP/kanban hybrid control policy achieved the service requirement with the least inventory. One remarkable result was the ease of implementation of this policy: Instead of passing kanban from the assembly line to the last machine in the subassembly line as in kanban control, these kanban are passed to the *first* machine to regulate the release of work to the system, while all the intermediate kanban cells work as before.

Some faster analytical models have been developed. One approach that has been proved useful in a number of contexts is that of *decomposition*. The basic idea is to break a complex manufacturing system down into a collection of simple building blocks that can be analyzed individually. The interactions between the building blocks are captured in a set of approximate equations that can be solved numerically. One early paper was Sevast'yanov (1962), but the most influential results were those of Gershwin (1987). This paper analyzed the performance of a synchronous transfer line with unreliable machines and finite buffers from the following principle: Imagine an observer sitting in a buffer. The observer will sometimes see material entering the buffer, and sometimes not. Sometimes material will be leaving the buffer, and sometimes not. To the observer, it is very hard to discern whether these processes are due to a single unreliable machine on either side of the buffer or to a longer line of interacting machines. Therefore, the local behavior at each buffer can be approximated by a two-machine, finite buffer model, for which a closed form solution exists. The parameters of these two-machine building blocks are then iteratively adjusted to maintain conservation of material flow across all two-machine lines. It is assumed that all machines have exponential up- and down-times.

This approach was improved upon by Dallery, David, and Xie (1988), and extended to tree-structured assembly/disassembly networks by Gershwin (1991). A similar approach has been used to analyze continuous material tandem lines where all machines have equal processing rates (*homogeneous* systems) by Dallery, David, and Xie (1989), which also was extended to tree-structured networks by Di Mascolo, David, and Dallery (1991). Homogeneous continuous material closed networks were analyzed by Frein, Commault, and Dallery (1994). Most recently, the continuous material decomposition has been extended to inhomogeneous lines (where the machines can have different processing rates) by Burman (1995). This continuous material model is commonly used as a fluid approximation to a discrete material manufacturing system where the operation times are deterministic or nearly deterministic, but where the machines are not bound to a common synchronous operating cycle.

The decomposition method was also used to analyze discrete material lines where all machines have exponentially distributed processing times with different rates by Choong and Gershwin (1987).

Among other control policies, systems controlled by a kanban discipline were analyzed by Mitra and Mitrani (1990) and Mitra and Mitrani (1991) using a decomposition approach, and the basestock discipline was studied by Lee and Zipkin (1992) using a different approximation.

All of these decomposition papers use a one-moment approximation, that is, the machines in the building blocks match only the mean of the time to resume material flow in the original line when a disruption occurs. If machines in the line have different repair rates, the actual distribution of flow interruption durations is hyper-exponential, where each machine contributes to one phase. A two-moment approximation using *generalized exponential* distributions has recently been proposed by Dallery (1994) and implemented in a decomposition procedure by Dallery and Le Bihan (1995). Initial results of this approach seem promising, especially for systems where the repair rates differ by several orders of magnitude between machines.

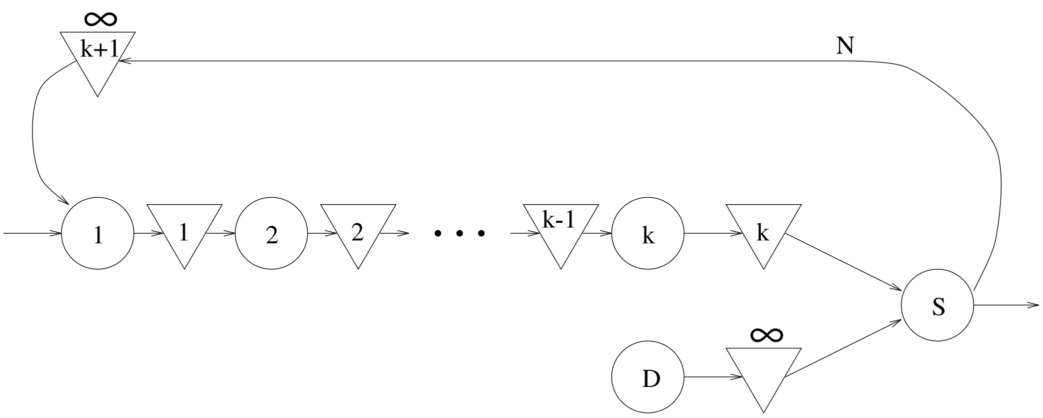

Figure 2: A fluid model representation of the CONWIP/finite buffer hybrid control policy

3 Outline of proposed research

I propose to study this class of policies through a fluid approximation, where the discrete parts in the system are approximated by a continuous material stream. The machines are considered unreliable with exponentially distributed times to fail and repair. The demand process can be modeled by a machine, as shown in figure 2, where a hybrid CONWIP/finite buffer control is represented. Machine D is the demand process, feeding an infinite buffer that represents the system backlog. Demand from this buffer is then joined with finished products by the reliable and infinitely fast synchronization machine S, which causes work orders to flow back to the first machine in the actual line through another infinite buffer.

This model can be decomposed as shown in figure 3. The variation from previous decompositions is that we now need to solve the synchronization stage equations, yielding a different type of building block, and allow for infinite buffers in the two-machine lines. This decomposition can then be solved by a procedure similar to that of Frein, Commault, and Dallery (1994). Because the conservation of flow equations are linearly dependent (wrapping around the loop), we need to fix some solution variable, solve, adjust the variable, resolve, and so on until a solution is found that satisfies the loop population constraint to the desired level of accuracy. Frein, Commault, and Dallery chose to fix the ratio of failure and repair rate at the first machine, after obtaining an upper and lower limit on this ratio, and then do a binary

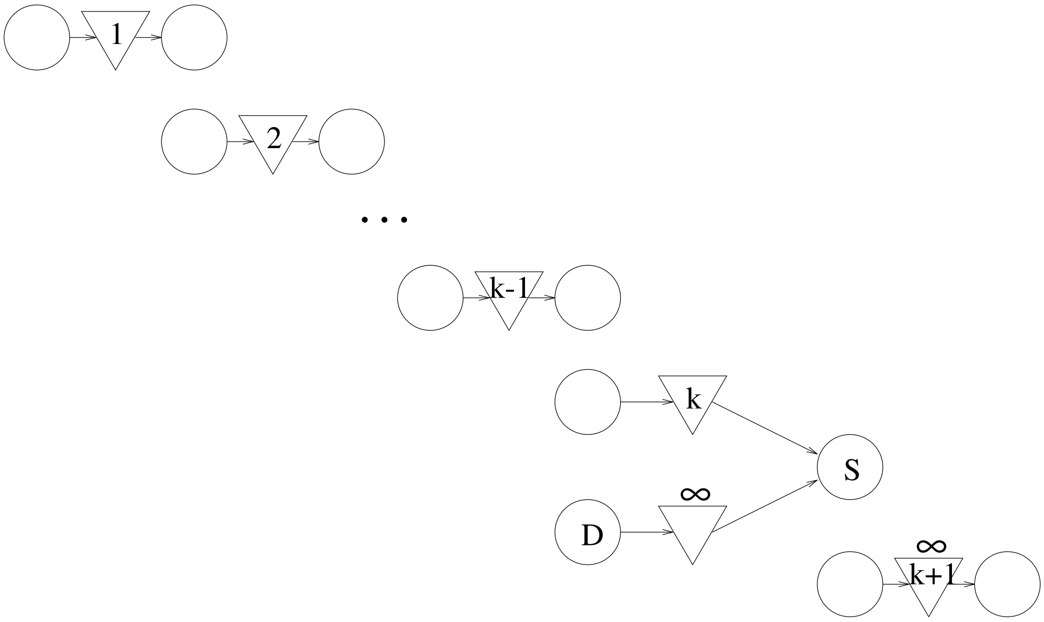

Figure 3: A decomposition of the hybrid controlled system

search in the value of this ratio until the loop population came sufficiently close to the constraint.

I also propose to develop a similar model for the basestock/finite buffer hybrid, with the restriction that all basestock levels (except that of the finished goods buffer) be zero. It is then possible to represent this control discipline as shown in figure 4. The difference from the previous case is that the loop invariant now is the summed loop inventory levels *minus* the system backlog. A similar decomposition procedure is expected to work.

For both cases, I propose to explore a number of decomposition methods, including the two-moment approximations of the generalized exponential decomposition (Dallery and Le Bihan 1995) and the decomposition method for inhomogeneous systems (Burman 1995).

Based on the results obtained by Frein, Commault, and Dallery and preliminary results reported below, I expect the decomposition approach to give satisfactory results for sufficiently large systems (more than about 10 machines in the line). However, for smaller systems, the results are likely to be inaccurate, because the production authorizations returning to machine 1 introduce significant correlation between the machines at the ends of the

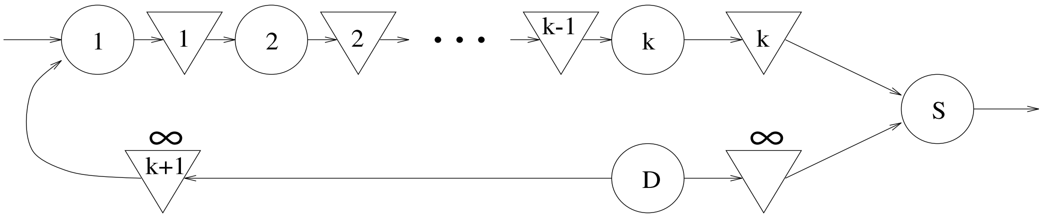

Figure 4: A fluid model representation of the restricted basestock/finite buffer hybrid control policy

system. This is borne out by the preliminary results. For this reason, I also propose to study the fluid approximation to a two-machine line operated by a two boundary control policy in detail. This system can be described by a set of partial differential equations with complicated boundary conditions. A closed form solution is not likely to be obtainable, but the system can be solved numerically by a method of finite differences. This is hoped to give additional insight into the reasons for inaccurate decomposition results for small systems.

4 Preliminary results

I have implemented the decomposition described for the CONWIP/finite buffer hybrid control policy for the homogeneous case. The demand process is represented as an unreliable machine with average production rate equal to the demand rate. The decomposition uses the "classical" decomposition equations of Dallery, David, and Xie (1989). (The decomposition produces the exact production rate of the system, which is not as impressive as it seems, since it is given a priori by the demand process.) The estimates of average buffer levels compare quite well with simulations of the same fluid system model, as shown in figure 5. In this figure, buffer 15 (marked FG) is the finished goods buffer, buffer 16 (BK) is the system backlog, and buffer 17 (FK) are "free kanbans" or production authorizations headed back to machine 1. As expected, smaller systems give less accurate results than those shown here.

As a stepping stone to the analysis of the continuous material two-machine line and its partial differential equations, I have investigated the behavior of a discrete material two-machine system controlled by a two-boundary policy

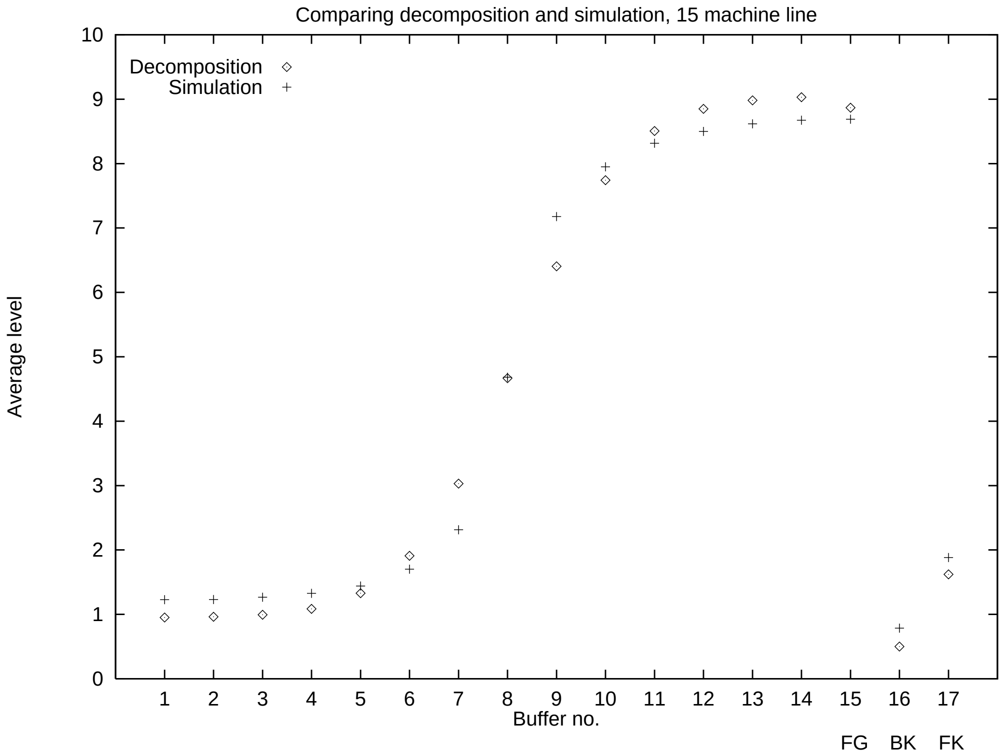

Figure 5: Accuracy of buffer level estimates

when demand is a Poisson process, the operation times are exponentially distributed, and the failure and repair times are exponential. In this case, the system can be modeled as a continuous time discrete state Markov process, albeit one with an infinite state space in the direction of increasing backlog. An approximate solution to this system has been developed by assuming asymptotical independence of the backlog and intermediate buffer level distributions as the system backlog goes to infinity, and using this to find a product-form approximation to the joint distribution in the deep backlog region. This gives some qualitative insight into the structure of the solution, as demonstrated by the probability distributions plotted in figure 6. In this figure, the "Surplus" axis measures the level of finished goods inventory when positive and the system backlog when negative.

The partial differential equations for the continuous material two-machine line and their boundary conditions have been derived for the case where the basestock level for the intermediate buffer is less than the buffer size, but

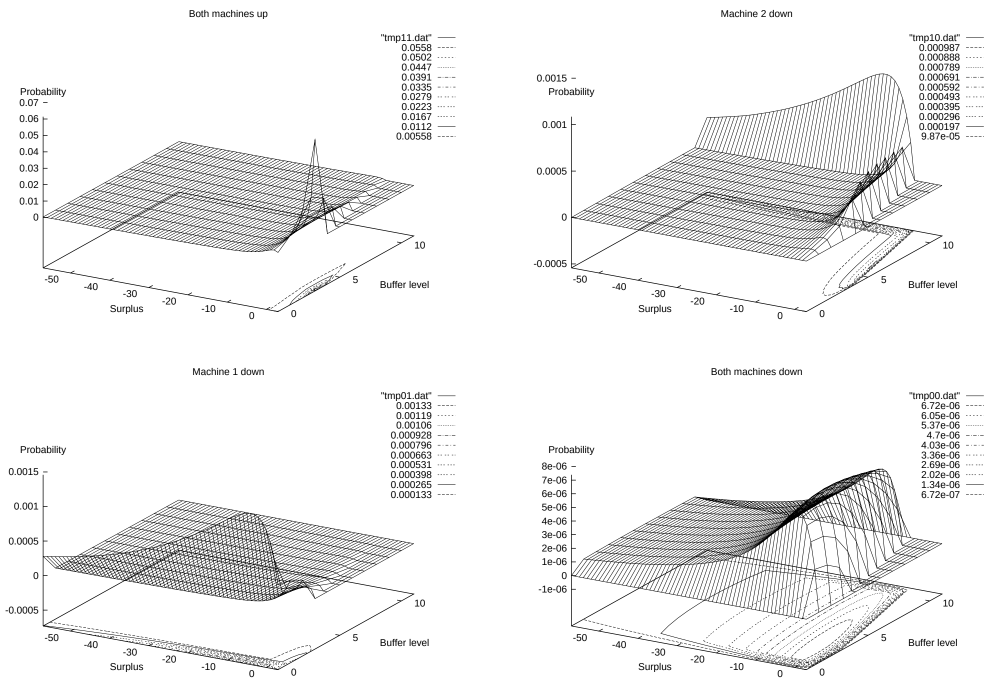

Figure 6: Markov solutions with approximated tail distribution

greater than zero. These equations will not be repeated here due to space considerations (there are 171 equations), but are available on request.

5 Timetable

The following is a tentative timetable for the research:

Sep 1, 1995	This proposal ready.
Oct 1, 1995	Inhomogenous and generalized exponential decompositions implemented.
Nov 1, 1995	Basestock/finite buffer decomposition implemented.
Dec 1, 1995	Numerical solution to the continuous material two-machine line implemented.
Feb 1, 1996	Numerical results complete.
April 1996	Thesis defense.
May 3, 1996	Thesis due date.

Acknowledgements

Thanks to Yves Dallery, who gave me many of these ideas.

References

Bai, S. X. (1991, Sep). *Scheduling manufacturing systems with work-in-process inventory control.* Ph. D. thesis, Massachusetts Institute of Technology.

Berkley, B. J. (1991). Tandem queues and kanban-controlled lines. *International Journal of Production Research 29*(10), 2057–2081.

Berkley, B. J. (1992). A review of the kanban production control research literature. *Production and Operations Management 1*(4), 393–411.

Bielecki, T. and P. R. Kumar (1988, Jul-Aug). Optimality of zero-inventory policies for unreliable manufacturing systems. *Operations Research 36*(4), 532–541.

Bonvik, A. M., C. Couch, and S. B. Gershwin (1995). Improving the performance of a production line controlled by kanban with rapid information dissemination. Submitted for publication.

Burman, M. H. (1995, Jun). *New results in flow line analysis*. Ph. D. thesis, MIT, 77 Massachusetts Ave., Cambridge MA 02139. Also available as Report LMP-95-007, MIT Laboratory for Manufacturing and Productivity.

Buzacott, J. A. and J. G. Shantikumar (1992). A general approach for coordinating production in multiple-cell manufacturing systems. *Production and Operations Management 1*(1), 34–52.

Choong, Y. F. and S. B. Gershwin (1987). A decomposition method for the approximate evaluation of capacitated transfer lines with unreliable machines and random processing times. *IIE Transactions 19*(2), 150–159.

Clark, A. J. and H. Scarf (1960, Jul). Optimal policies for the multi-echelon inventory problem. *Management Science 6*(4), 475–490.

Dallery, Y. (1994). On modeling failure and repair times in stochastic models of manufacturing systems using generalized exponential distributions. *Queuing Systems 15*, 199–209.

Dallery, Y., R. David, and X.-L. Xie (1988, Sep). An efficient algorithm for analysis of transfer lines with unreliable machines and finite buffers. *IIE Transactions 20*(3), 280–283.

Dallery, Y., R. David, and X.-L. Xie (1989, Sep). Approximate analysis of transfer lines with unreliable machines and finite buffers. *IEEE Transactions on automatic control 34*(9), 943–953.

Dallery, Y. and S. B. Gershwin (1992, Dec). 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.

Dallery, Y. and H. Le Bihan (1995, Oct). An improved decomposition method for the analysis of production lines with unreliable machines and finite buffers. In *Proceedings of the Emerging Technology for Factory Automation Conference (ETFA)*, Paris.

Dallery, Y. and G. Liberopoulos (1995, Jun). Extended kanban control system: A new kanban-type pull control mechanism for multi-stage manufacturing systems. Unpublished manuscript.

Di Mascolo, M., R. David, and Y. Dallery (1991, Dec). Modeling and analysis of assembly systems with unreliable machines and finite buffers. *IIE Transactions 23*(4), 315–330.

Frein, Y., C. Commault, and Y. Dallery (1994, Nov). Modeling and analysis of closed-loop production lines with unreliable machines and finite buffers. Technical Report LAG No. 92-123, revised Nov 1994, Laboratoire d'Automatique de Grenoble, URA 228 CNRS, ENSIEG, BP 46, 38402 Saint Martin d'Hères. To appear in IIE Transactions.

Frein, Y., M. Di Mascolo, and Y. Dallery (1994). On the design of generalized kanban control systems. To appear in *International Journal of Operations and Production Management*, special issue on Modelling and Analysis of Just-in-Time Manufacturing Systems.

Gershwin, S. B. (1987, Mar-Apr). An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking. *Operations Research 35*, 291–305.

Gershwin, S. B. (1991, Dec). Assembly/disassembly systems: An efficient decomposition algorithm for tree-structured networks. *IIE Transactions 23*(4), 302–314.

Kimball, G. (1988). General principles of inventory control. *Journal of Manufacturing and Operations Management 1*(1), 119–130.

Kimemia, J. and S. B. Gershwin (1983, Dec). An algorithm for the computer control of a flexible manufacturing system. *IEE Transactions 15*(4), 353–362.

Lee, Y.-J. and P. Zipkin (1992, Sep-Oct). Tandem queues with planned inventories. *Operations Research 40*(5), 936–947.

Mitra, D. and I. Mitrani (1990, Dec). Analysis of a kanban discipline for cell coordination in production lines, I. *Management Science 36*(12), 1548–1566.

Mitra, D. and I. Mitrani (1991, Sep-Oct). Analysis of a kanban discipline for cell coordination in production lines, II: Stochastic demands. *Operations Research 39*(5), 807–823.

Monden, Y. (1983). *Toyota Production System: Practical approach to production management*. Atlanta: Industrial Engineering and Management Press.

Sevast'yanov, B. A. (1962). Influence of storage bin capacity in the average standstill time of a production line. *Theory of Probability and Its Applications 7*, 429–438.

So, K. C. and S. C. Pinault (1988, Dec). Allocating buffer storages in a pull system. *International Journal of Production Research 15*(12), 1959–1980.

Spearman, M. L., D. L. Woodruff, and W. J. Hopp (1990). CONWIP : a pull alternative to kanban. *International Journal of Production Research 28*(5), 879–894.

Van Ryzin, G. (1987, Jun). Control of manufacturing systems with delays. Master's thesis, Massachusetts Institute of Technology. Also available as MIT Laboratory for Information and Decision Systems Report LIDS-TH-1676.

Van Ryzin, G., S. X. C. Lou, and S. B. Gershwin (1993, Sep). Production control for a tandem two-machine system. *IIE Transactions 25*(5), 5–20.

Veatch, M. H. and L. M. Wein (1994, Mar-Apr). Optimal control of a two-station tandem production/inventory system. *Operations Research 42*(2), 337–350.

Violette, J. D. (1993, April). Implementation of a hierarchical controller in a factory simulation. Master's thesis, Massachusetts Institute of Technology.

Womack, J. P., D. T. Jones, and D. Roos (1990). *The machine that changed the world : The story of lean production.* Rawson Associates.