

Materials Flow Control in Hybrid Make-to-Stock/Make-to-Order Manufacturing

Filipa Rocha¹, Emanuel Silva¹, Ângela Lopes¹, Luis Dias¹, Guilherme Pereira¹,
Nuno O. Fernandes², and S. Carmo-Silva^{1(✉)}

¹ Department of Production and Systems, University of Minho,
Campus de Gualtar 4710-057, Braga, Portugal
{lsd, gui, scarmo}@dps.uminho.pt

² Instituto Politécnico de Castelo Branco, Av. Empresário 6000-767, Castelo Branco, Portugal
nogf@ipcb.pt

Abstract. Today's company competitiveness is favoured by product customisation and fast delivery. A strategy to meet this challenge is to manufacture standard items to stock for product customisation. This configures a hybrid environment of make-to-stock and make-to-order. To explore the advantages of this requires good understanding of production control. Thus, we study production under hybrid MTS-MTO, organising the system in two stages. The 1st manufactures items to inventory, which are then customised in the 2nd. We analyse how the percentage of tardy orders is affected by the inventory of items required to achieve a given fill rate. The impact of two mechanisms for releasing orders to both stages is also analysed. Results of a simulation study indicate that most of the reduction on the percentage of tardy orders is achieved by a moderate increase in the stock level of semi-finished products. Moreover the percentage of tardy orders decreases if suitable controlled release of orders is exerted.

Keywords: Make-to-Stock/ Make-to-Order (MTO-MTS) · Base-stock · CONWIP · DBR, simulation

1 Introduction

For many manufacturing companies, the possibility of having competitive advantage in relation to their competitors frequently resides in having shorter delivery times and higher delivery reliability of orders [1]. This requires firmly controlling the internal throughput times and schedule adherence, while at the same time, devising low inventory levels and adequate resources' utilisation. This well-known dilemmatic problem can be solved by the Production Logistics function of a company, which organizes and manages the entire material flow, from the acquisition of raw materials to the delivery of end products to customers. Organizing production in a hybrid make-to-stock/make-to-order (MTS-MTO) manufacturing environment is a strategy that contributes to solve the dilemma. In fact it allows companies to exploit the benefits of delayed differentiation, reducing delivery times and inventory costs in comparison to the pure MTO and MTS

© Springer International Publishing Switzerland 2015
F. Corman et al. (Eds.): ICCL 2015, LNCS 9335, pp. 559–568, 2015.
DOI: 10.1007/978-3-319-24264-4_38

gui@dps.uminho.pt

strategies, respectively. It also allows customers to get what and when they want and companies to avoid the costs of shortages and overages. Its application is widely seen as a growing trend in manufacturing [2].

This type of production is common, for example, in the aluminium profiles manufacturing. Companies of this industrial sector instead of making-to-stock, start the production of standard and regular aluminium profiles without confirmed orders, holding a stock of semi-finished products. These are then assigned to customer orders and customized according to specifications. This, for example, can be for a specific surface finishing material or and profiles machining to fit the required application. This customization process is also common in other industries for several different types of products.

In this paper, we focus on the above dilemma in the context of the hybrid MTS-MTO manufacturing environment. The paper specifically deals with a two-stage manufacturing system with unidirectional production flows. In the MTS stage, standard semi-finished products are firstly manufactured and stocked as intermediate inventory. In the MTO stage, semi-finished components are assigned to customer orders for customisation according to specific requirements. The intermediate inventory acts as a buffer or decoupling point between the two stages. High inventory of semi-finished products means high holding costs, whereas low inventory may increase the waiting time of customer orders for semi-finished components, and thus the delivery time. Therefore, in addition to inventory decisions at the MTS stage, order release and dispatching decisions at the MTO stage should determine the company capability to quote short and reliable delivery times to their customers, and thus to remain competitive in the market. In spite of this, most research on inventory location ignores the intricacies of scheduling, typically assuming that orders are processed in the sequence in which they arrive to the production system [3]. Thus, in this paper, we investigate how materials flow's control strategies impacts system performance. The release of orders to the MTS stage is based on inventory replenishment. Two alternative mechanisms have been studied, namely Base-stock [4, 5] and CONstante Work-In-Process (CONWIP) [6]. The release of orders to the MTO stage is based on the capacity availability at the bottleneck workstation. Here Drum-Buffer-Rope (DBR) [7, 8] was applied. The objective is to satisfy the MTO demand within competitive delivery times, while keeping work-in-process (WIP) low at the MTS stage and avoiding stock-outs of semi-finished products at the intermediate buffer. The controlled release of orders to the MTO stage must ensure that orders are not released too early or too late, and maintaining the workload at workstations low and stable. Inventory replenishment at the MTS stage ensures that the semi-finished products' buffer is filled to the required level, without a rigid order release plan.

Research efforts are restricted in the area of materials flow control for the hybrid MTS-MTO manufacturing. Most of the research literature in two-stage production systems assume that the semi-finished products inventory is managed according to a base-stock policy. Under this policy, each demand arrival triggers the immediate release of raw materials to the MTS stage and the immediate release of orders to the MTO stage as soon as semi-finished products become available at the intermediate buffer (see e.g. [3, 9, 10, 11, 12]).

This research work gives a contribution to fill the research gap in the area, using discrete event simulation to model and analyse the performance of a two-stage manufacturing system when operated under the above-referred materials flow control mechanisms. In particular the following research question is addressed: How controlled release of orders to both stages of the production system impacts system performance?

The remainder of the paper is organized as follows. In Section 2, we present the simulation study carried out, including the simulation model, the experimental set-up and the performance measures considered. In Section 3, we discuss the results of the simulation study, and finally, in Section 4 of the paper, we summarize key results and managerial implications.

2 Simulation Study

Simulation will next be used to model building and experimentation towards answering the above research question. A discrete event simulation model was developed using Arena® software.

2.1 Manufacturing System

In this study, we consider a hypothetical two-stage manufacturing system with an intermediate buffer of semi-finished products (see Figure 1). Stage one consists of workstations 1, 2 and 3, and manufactures standard components from raw materials. Stage two consists of workstations 4, 5 and 6, and manufactures end products to order from the components made at stage one, i.e., according to the customer specifications. Production flows at the system are unidirectional with each production order having exactly the same routing.

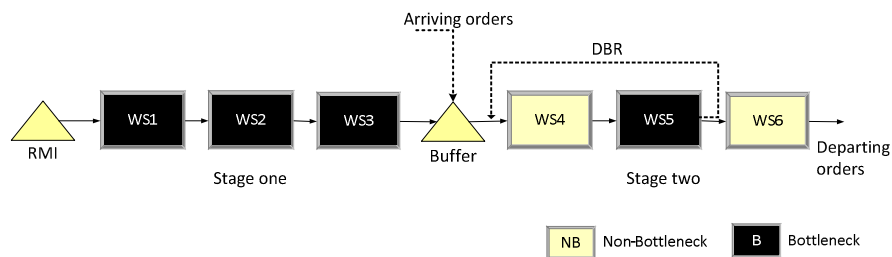


Fig. 1. Two-stage manufacturing system

Whenever a customer order arrives at the production system, a standard component from the intermediate buffer is allocated to the order. We assume that each customer order requires just one unit of the standard components. However, when a customer order arrives and finds the intermediate buffer empty, the order is backordered. Backorders are filled after the standard components are made available at the intermediate buffer following processing at stage one. The first stage is capable of manufacturing

two different types of standard components or semi-finished products that are then customised in the second stage into, virtually, an infinite number of end products.

As customer orders arrive to the production system, their operation times are identified and due dates established. It is assumed that all orders are accepted and enough raw materials inventory is always available in the beginning of the first stage. Orders inter-arrival times follow an exponential distribution, due dates are market driven and set by adding a uniformly distributed time allowance to the time of order arrival. In this study, the allowance varies between 25 and 45 time units. This leads to approximately 12% of orders being tardy when immediate release of orders to the second stage is used, for a fill rate level of 80%. This has been verified through preliminary simulation tests. This fill rate level means that 80% of arriving customer orders are filled from the semi-finished products inventory and then made immediately available for release into the second stage.

In the simulation model, operations' processing times follow an exponential distribution, with a mean of one time unit. The arrival rate combined with the routings and processing times ensures that utilisation is 90% at all workstations, except at workstations four and six. These have 20% of protective capacity. Protective capacity is defined as a given amount of extra capacity at non-constraints above the system constraint's capacity, used to protect against statistical fluctuations [13]. Workstations with protective capacity are non-bottleneck. This means that on stage one the bottleneck will shift across workstations 1 to 3, since utilisation is equal across workstations, while on stage two workstation five (WS5) is the bottleneck for most of the time. Additionally, following assumptions are adopted:

- Workstations operate asynchronously, so production orders can be loaded whenever material or inventory is available and the required production authorisations have been received. These take the form of available production control cards when required by the production control mechanism.
- Workstations capacity remains constant over time and no breakdowns have been modelled.
- Set-up times are assumed to be sequence-independent and included in the operation processing times.
- Distances and transportation times between workstations and between production stages are assumed to be negligible.
- Information of production control events and production control cards are transmitted instantly.

The simulation model presented here was kept simple in order to ensure easy and correct interpretation of the effect of combining MTO and MTS and getting results that may contribute for the understanding of the behaviour of more complex production systems.

2.2 Materials Flow Control

Materials Flow Control addresses two main functions: order release and priority dispatching. Order release determines the time and the orders to be released into each

production stage. Release decisions are usually based on the orders' urgency and on their influence on the current shop floor situation [14]. Priority dispatching determines which orders in queue should be selected next for processing once a workstation becomes idle.

In the manufacturing system considered, an arriving customer order will consume a standard component from the intermediate buffer of semi-finished products and immediately flows into a pre-shop pool, waiting its release to the second stage of processing for customisation. This means that orders are not immediately released to the shop as they seize the semi-finished products that they need. Rather, they must wait for capacity availability at the second stage. The use of a pre-shop pool is expected to reduce the level of work-in-process (WIP) and allow better control over the flow of production orders through the shop. Orders in the pool are sequenced according to their urgency, i.e., due date, and released under the drum-buffer-rope control mechanism (DBR), see Figure 1.

There are three major components of DBR: the drum; the buffer; and the rope. The drum represents the bottleneck workstation (WS5 in our manufacturing system), which defines the overall pace of the system, the buffer is a time mechanism that protects the bottleneck against starvation and the rope drives the release of orders to the shop. For this purpose, production authorisation cards are used, which are attached to production orders at the beginning of stage two and detached from the order after processing at the bottleneck workstation. Cards are not part number specific and can be acquired by any order waiting release in the pre-shop pool. Detached cards are sent back to the first workstation of the second stage, where they can be attached to new production orders entering the system. In our study we consider that one order requires just one production authorisation card from each pair of workstations in the routing of the order.

When a customer order arrives at the production system or when a semi-finished product is consumed from the intermediate buffer, depending on the mechanism applied (Base-stock or CONWIP, see section 2.3), a MTS order is also released to the first stage for the replenishment of the standard component that will be (or was) consumed from the semi-finished product buffer.

The role of priority dispatching is a very modest one when order release control is applied, because the choice among jobs is limited due to short queues [15]. Thus in this study shop floor dispatching at both stages is based on *first-come-first-served* (FCFS) priority dispatching rule that supports the natural flow of the orders through the shop, stabilizing operation throughput times.

2.3 Experimental Design and Performance Measures

The experimental factors and simulated levels of the study are summarised in Table 1. Thirty simulation cases are tested (2 production control strategies; 5 card counts; and 3 levels of the fill rate), and each test case runs 100 replicates. The time horizon for a simulation case is 200 000 time units and only data of the last 190 000 time units are collected, i.e., a warm-up period of 10 000 time units is considered.

Table 1. Experimental factors and levels

Experimental factor	Levels		
Materials Flow Control Strategies	B-DBR	C-DBR	
Card counts	10, 12, 15, 18 and infinity		
Fill rate	80%	90%	99%

Two materials flow control strategies are applied to release orders to the shop, namely, B-DBR and C-DBR. The former combines Base-stock with DBR (B-DBR) to release orders to first and second stages, respectively. The latter combines CONWIP with DBR (C-DBR) for the same purpose.

Under Base-stock control, demand information is instantaneously transmitted to all workstations of the first stage when a customer order arrives to the manufacturing system. The idea is that all workstations know about each customer order as it arrives to the system and start immediately to manufacture a replacement item. For that, Base-stock maintains a certain amount of inventory in the input buffer on each workstation. This amount of inventory is called base stock level. Thus the particularity of this mechanism is their immediate reactivity to demand. In our study all base stock levels are zero, except for the semi-finished products. This complies with previous studies e.g. [16].

Under CONWIP control, each time a semi-finished product is seized by a customer order, the release of a new job (i.e., MTS order) to the first stage is authorized. CONWIP uses cards from a single card type to control the total amount of work-in-process (WIP) permitted in the first processing stage. Cards are attached to the order at the beginning of the first stage and detached from the order (i.e., semi-finished product) when it is seized by an arriving customer order.

The mechanism applied to release orders to the second stage of processing, DBR, was tested at 5 card counts, including infinity. Card counts are sets of cards to operate the manufacturing system. When infinity is assumed, this means that the number of cards in the set imposes no restriction on the number of jobs, and thus on the workload, that can be released to the shop.

The order fill rate, defined as the percentage of orders that is immediately filled from the semi-finished products buffer when a customer order arrives, was tested at three levels: low (80%), medium (90%) and high (99%). Note that the fill rate is expected to approach 100% as the semi-finished products inventory increases and tends to infinity. The inventory replenishment mechanisms used aim at determining the base-stock level or the number of CONWIP cards, depending on to the mechanism applied at first stage, to achieve the desired fill rate, i.e., 80%, 90% or 99%. This was determined through exhaustive searching using pre-test simulation runs.

We use three types of criteria to evaluate the system's performance: (1) the ability to deliver orders on time, (2) the ability to provide short delivery times, and (3) the ability to keep the inventory levels low. To measure performance with regard to the first, the percentage of tardy orders (P_{tardy}) and the standard deviation of lateness (SD_{lateness}) are recorded. To measure performance with regard to the second, the shop

throughput time (TT) and the total (or system) throughput time (TTT) are used. The shop throughput time refers to the time that elapses between order release to the second stage and order completion. The total throughput time is the shop throughput time plus the pool delay and plus the time that orders wait for semi-finished products availability. To measure performance with regard to the third, work-in-process (WIP) at stage one and the average inventory of semi-finished products (SFPI) at the intermediate buffer are used.

3 Simulation Results and Discussion

This section discusses the results of the simulation study described in the previous section. Comparisons between materials flow control strategies are based on the Student paired t-test with $\alpha=0.05$.

The overall results are summarised in Table 2. In what concerns the C-DBR strategy only the results of an infinity number of cards are shown as it was observed that there are no statistical performance differences between the two materials flow control strategies, i.e. C-DBR and B-DBR, for TT, TTT, SDlateness and P_{tardy}.

Figure 2 plots the percentage of tardy orders against the shop throughput time for different combinations of the experimental factors under the B-DBR strategy. By comparing plotted curves we can determine performance differences for different values of card counts. A marker on a curve is the result of simulating DBR at a specific card count. Five card counts have been simulated, including infinity. Infinity means unrestricted release of orders to the second stage of shop and refers to the rightmost mark on each curve.

From Figure 2 it can be observed that increasing the fill rate from 80% to 99%, leads to a lower percentage of tardy orders. A higher fill rate means a higher probability of orders being filled from the intermediate buffer of semi-finished products, and therefore, the time that orders wait for semi-finished products availability tends to be lower. This makes the total throughput time lower and percentages of tardy orders also lower (see Table 2). Note that a lower percentage of tardy orders may result from a lower average lateness (as a results of lower total throughput times), but also from a lower variance of the lateness. In this case it results from both, as can be observed from Table 2. However, the lower percentage of tardy orders is obtained at the cost of having a higher inventory of semi-finished products between production stages and a higher WIP at the MTS stage, as Table 2 shows.

Increasing the fill rate from 80% to 99% allows a reduction on the total throughput time in more than 13% and on the percentage of tardy orders in more than 38%, under unrestrictive release, depending on the materials flow control strategy. However, this requires increasing the average inventory of semi-finished products in more than 231%, and WIP at stage one in more than 87%, depending on the materials flow control strategy. Most of the reduction on the percentage of tardy orders is achieved when the fill rate increases from 80% to 90%. This requires a moderate increase in WIP and in the stock level of semi-finished products, sees Table 2.

Table 2. Performance Results

Control strategy	DBR card counts	Fill rate	WIP	SFPI	TT	TTT	SD _{lateness}	P _{tardy}
C-DBR	infinity	80%	41.00	15.91	15.24	18.48	15.24	12.17
		90%	51.00	24.78	15.48	16.81	13.41	8.95
		99%	80.00	52.69	15.75	15.87	12.29	7.23
B-DBR	infinity	80%	42.69	15.38	15.36	18.34	14.87	11.94
	18				13.58	18.40	14.71	11.70
	15				12.86	18.54	14.87	11.69
	12				11.88	18.89	14.95	12.21
	10				11.01	19.53	15.42	13.53
B-DBR	infinity	90%	51.23	23.93	15.51	16.88	13.36	9.13
	18				13.60	16.93	13.33	8.68
	15				12.88	17.11	13.45	8.74
	12				11.90	17.42	13.54	9.28
	10				11.07	18.24	14.07	10.78
B-DBR	infinity	99%	80.12	52.85	15.79	15.92	12.40	7.34
	18				13.71	16.02	12.29	6.84
	15				12.94	16.14	12.34	6.87
	12				11.91	16.47	12.48	7.42
	10				11.09	17.40	13.13	9.01

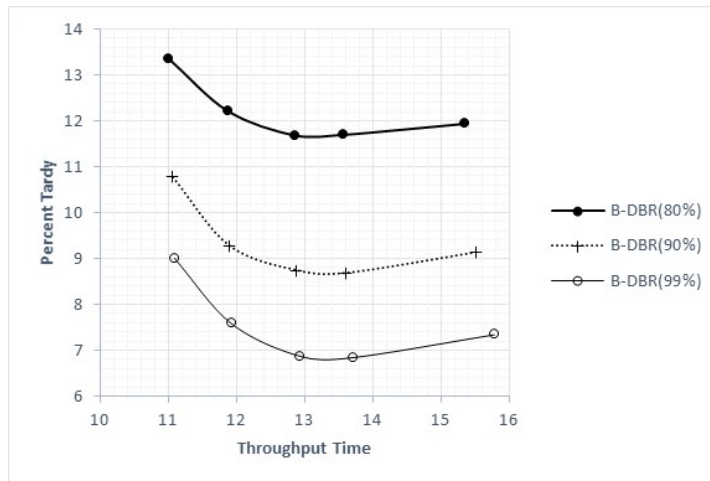


Fig. 2. Performance results for the percentage of tardy orders under the B-DBR strategy

Figure 2 also shows that when card counts are reduced, i.e. moving from right to left along the curve, the percentage of tardy orders decrease. Thus we may conclude that controlled release of orders allows improving performance. However, card counts cannot be set excessively low in order to avoid deteriorating performance. We also may observe in Table 2 that C-DBR performs slightly better than B-DBR in what concerns WIP at stage one, but only when the fill rate is low. This may be attributed to the CONWIP part of this control strategy, which only releases an MTS order to the first stage when a semi-finished product is consumed by a customer order, therefore avoiding increasing WIP. On the contrary, Base-stock releases MTS orders as soon as customer orders arrive to the manufacturing system.

4 Conclusions and Managerial Implications

This study investigates the impact of two materials flow control strategies in the hybrid MTS-MTO manufacturing. The study also analyses how the percentage of tardy orders is affected by the semi-finished products inventory between the MTS and the MTO stages of manufacturing required to achieve a given fill rate. Results led us to conclude that: (1) controlled release of orders to both stages contributes to reduce the percentage of tardy orders; (2) CONWIP may result in a lower work-in-process at stage one, but only when a low fill rate is allowed; (3) A trade-off exists between the percentage of tardy orders and the semi-finished products inventory required to achieve a given fill rate. However, most of the reduction on the percentage of tardy orders is achieved for a moderate increase in the stock level of semi-finished products.

Our future research work will extend the study to other materials flow control strategies and shop configurations in context of this hybrid production environment.

Acknowledgements. This study had the financial support of FCT-Fundação para a Ciência e Tecnologia of Portugal under the project PEst2015-2020: UID/CEC/00319/2013.

References

1. Nyhuis, P., Wiendahl, H.-P.: *Fundamentals of Production Logistics: Theory, Tools and Applications*. Springer-Verlag, Heidelberg (2009)
2. Åhlström, P., Westbrook, R.: Implications of Mass Customisation for Operations Management. *International Journal of Operations and Production Management* **19**(3), 262–274 (1999)
3. Kaminskaya, P., Kayab, O.: Inventory positioning, scheduling and lead-time quotation in supply chains. *International Journal of Production Economics* **114**(1), 276–293 (2008)
4. Kimball, G.: General principles of inventory control. *Journal of Manufacturing and Operations Management* **1**(1), 119–130 (1988)
5. Lee, Y.J., Zipkin, P.: Tandem queues with planned inventories. *Operations Research* **40**(5), 936–947 (1992)

6. Spearman, M.L., Woodruff, D.L., Hopp, W.J.: CONWIP: a pull alternative to kanban. *The International Journal of Production Research* **28**(5), 879–894 (1990)
7. Goldratt, E.M., Fox, R.E.: *The Race*. North River Press, New York (1986)
8. Fry, T.D., Karwan, K., Steele, D.: Implementing drum-buffer-rope to control manufacturing lead time. *The International Journal of Logistics Management* **2**(1), 12–18 (1991)
9. Aviv, Y., Federgruen, A.: Capacitated multi-item inventory systems with random and seasonally fluctuating demands: implications for postponement strategies. *Management science* **47**(4), 512–531 (2001)
10. Gupta, D., Benjaafar, S.: Make-to-order, make-to-stock, or delay product differentiation? A common framework for modeling and analysis. *IIE Transactions* **36**, 529–546 (2004)
11. Almehdawe, E., Elizabeth Jewke, E.: Performance analysis and optimization of hybrid manufacturing systems under a batch ordering policy. *International Journal of Production Economics* **144**(1), 200–208 (2013)
12. Fernandes, N.O., Silva, C., Carmo-Silva, S.: Order release in a hybrid MTO-MTS two-stage production system. In: *Pre-prints of the 18th International Working Seminar on Production Economics*, Innsbruck, Austria, February 24–28, 2014
13. Cox III, J.F., Blackstone Jr., J.H. (eds.): *APICS Dictionary* Alexandria. APICS, VA (2002)
14. Henrich, P., Land, M., Gaalman, G.J.C.: Exploring applicability of the workload control concept. *International Journal of Production Economics* **90**(2), 187–198 (2004)
15. Land, M.J., Gaalman, G.: The performance of workload control concepts in job shops: Improving the release method. *International Journal of Production Economics* **56–57**(1), 347–364 (1998)
16. Bonvik, A.M., Couch, C.E., Gershwin, S.B.: A comparison of production-line control mechanisms. *International Journal of Production Research* **25**(3), 789–804 (1997)