Improving Production Logistics Through
Materials Flow Control and Lot Splitting

Catarina Gomes[1], Andreia Ribeiro[1], João Freitas[1], Luís Dias[1],
Guilherme Pereira[1], António Vieira[1], Nuno O. Fernandes[2(✉)],
and Sílvio Carmo-Silva[1]

[1] Department of Production and Systems, University of Minho,
Campus de Gualtar, 4710-057 Braga, Portugal
`catarinafrlgomes@gmail.com`,
`andreia.queiroz.ribeiro@gmail.com`,
`jtp_l993@hotmail.com`,
`{lsd,gui,antonio.vieira,scarmo}@dps.uminho.pt`
[2] Instituto Politécnico de Castelo Branco,
Av. Empresário, 6000-767 Castelo Branco, Portugal
`nogf@ipcb.pt`

Abstract. Competitive advantage of make-to-order manufacturing companies is highly dependent on their capability to offer short delivery times and on time delivery. This calls for effective production and materials flow control – a core part of production logistics. This paper applies discrete simulation to study the delivery performance of a make-to-order manufacturing system configured as a general flow shop, when operated under two card-based material flow control mechanisms: CONWIP and GKS. The influence of two lot splitting strategies on the performance of these mechanisms is also evaluated. Results show that GKS clearly outperforms CONWIP and that splitting strategies have a positive impact on the performance of both mechanisms. GKS also showed to be particularly robust to the variation of the number of production authorisation cards used. This, together with the fact that the card-based mechanisms require little data handling and simplify production control, makes GKS attractive for practical application in make-to-order companies.

Keywords: CONWIP · GKS · Lot splitting · MTO · Simulation

1 Introduction

Production and materials flow control are important functions of production logistics. Production logistics fundamentally pursues high delivery capability and reliability with minimum logistic and production costs [12]. Delivery capability expresses the degree to which a company can commit itself to customer desired delivery dates. Delivery reliability, on the other hand, expresses the extent to which the order due dates are met. To achieve this, production and materials flow control must organize and manage the entire production and material flow, from the acquisition of raw materials to the delivery of end products to customers, ensuring that each machine or workstation of the production system is fed with the right jobs at the right time.

Card-based materials flow control mechanisms such as the Generic Kanban Systems (GKS) [4] and CONstant Work-In-Process (CONWIP) [14] can be simple and yet effective means of controlling production and materials flow. Such mechanisms have been proposed as alternatives to the Toyota Kanban system (TKS), which do not typically apply to the make-to-order (MTO) and high-variability production: the former as a way of applying Kanbans (i.e. production authorization cards) to control production and materials flow in dynamic environments; the latter as a simplified alternative to TKS for the MTO production.

Chang and Yih [4] compared these mechanisms in a pure flow shop. GKS was shown to be more flexible in that by manipulating the number of kanbans at each workstation the performance of GKS could be improved beyond that achieved by CONWIP. To the best of our knowledge no study was carried out on the influence of lot splitting on the performance of these two material flow control mechanisms, neither were them evaluated for manufacturing systems more suited to satisfy demand under MTO. Thus, this research work gives a contribution to fill this research gap, using discrete event simulation to model and analyse the performance of a general flow shop under MTO production, when operated by these two card-based materials flow control mechanisms. In particular, the following research questions are addressed:

1. How materials flow control mechanisms perform in the context of make-to-order and general flow shops?
2. How lot splitting impacts the performance of these mechanisms?

Lot splitting allows accelerating the flow of work by splitting job lots into sublots. The basic idea is not to process the whole job at one workstation and then move them to the next, but to move smaller quantities (sublots) to the next workstation as soon as they are completed. This may result in the overlapping of operations, shortening throughput times and thus improved logistic performance, as pointed out by Jacobs and Bragg [9] and Wagner and Ragatz [15], among others. It also reduces the amount of storage space as well as the capacity of material handling equipment required. Chang and Chiu [3] and Cheng et al. [5] make comprehensive literature reviews on lot streaming, i.e. lot splitting for operations overlapping.

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

2   Simulation Study

A discrete event computer simulation model was developed using Arena® software to model the system under study and carry out experimentation towards answering the above research questions.

2.1   Simulation Model

In this study, we consider a six-stage general flow shop (GFS) manufacturing system with one workstation in each processing stage. The GFS is seen as a more realistic model of the flow structure of job shops than pure job shop [6, 13]. A representation of the GFS used in the study is shown in Fig. 1, while Table 1 summarises the characteristics of the simulation model.
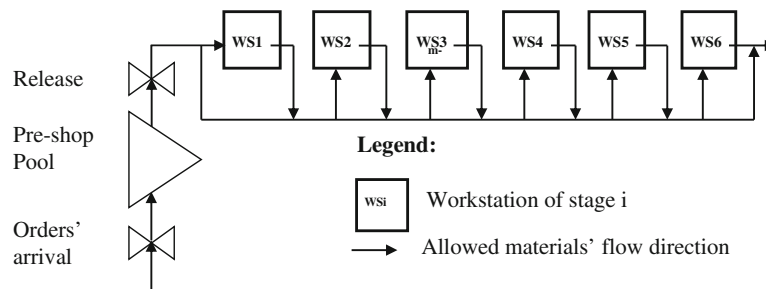

Fig. 1. The material flow structure of a general flow shop with six workstations (adapted from [2])

Table 1.   Job and shop characteristics

Shop configuration	General flow shop; no re-entrant flows
No. of workstations	6
Workstation capacities	All equal and constant over time
Workstation utilisation	90 %
Inter-arrival times	Exponentially distributed; mean = 0.647 time units
No. of operations per job	Discrete uniformly distributed [1, 6] operations
Job or order size (quantity)	Discrete uniformly distributed [1, 4] units
Unit processing times	2-Erlang, mean of 0,4 time units; truncated at 1.6 time units
Set-up time	Sequence independent, included in the processing times

As customer orders arrive to the manufacturing system, their operation times and due dates are established. It is assumed that all orders are accepted and enough raw materials inventory is always available. Orders' inter-arrival times follow an exponential distribution with mean 0.647 time units. The inter-arrival times when combined with the orders routings and operations times will result in an average utilisation of 90 % at all workstations. This is adopted in our model once it is a utilization level that usually is aimed at in practice and it allows having a good insight on the performance behaviour of the production control mechanisms tested.

Market driven due dates are set by adding a uniformly distributed time allowance to order arrival time. In this study, the allowance varies between 35 and 55 time units. The minimum value was set to cover a minimum shop floor throughput time corresponding to a planned operation lead time of 5 time units plus an allowance for the pool waiting

time. The maximum value was chosen in order to get a good perception of the relative performance of the control strategies on tardy jobs. This leads to approximately 18 % of orders being tardy under immediate release if lot splitting is not applied. To reflect the environment where customers demand unique or small quantities of products, the order quantity is randomly generated according to a discrete uniform distribution between one and four product units.

Customer orders, here also referred as jobs, can be split into sublots of equal size, which can be processed separately in the manufacturing system. The number of possible sublots in each job is directly related with the job size, i.e. the order quantity. Thus, the number of sublots for each job equals the order quantity and therefore, there are jobs with different number of sublots, varying from one to four.

In our simulation model, the processing times of operations of each product unit are drawn from a truncated 2-Erlang distribution with a mean of 0.4 time units and a maximum of 1.6 time units. These are then multiplied by the job size, i.e. order quantity, to obtain the processing time of each job operation. Additionally, the following assumptions are adopted:

- 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 and the values of system parameters were set to ensure easy and correct interpretation of the effects of the experimental factors as a contribution for the understanding of the performance behaviour of material flow control mechanisms when applied to the widely used in practice general flow shop manufacturing system configuration.

2.2   Materials Flow Control: Order Release and Dispatching

Materials flow control, an important part of production logistics, addresses two main production control functions: order release and priority dispatching. Order release determines the time and the orders to be released to the system, authorizing production to start. Release decisions are usually based on the orders' urgency and on their influence on the current shop floor situation [8]. Priority dispatching selects the job or the sublot to be processed next at a workstation that becomes free, from those waiting in workstations' queues.

In the manufacturing system considered, an arriving order immediately flows into a pre-shop pool, waiting its release to the system, i.e. to shop floor, for processing. This means that orders are not immediately released to the system as they arrive. Rather, they wait until release required conditions are met. The controlled release, associated to the use of a pre-shop pool, is expected to reduce the level of work-in-process (WIP) in

the system and allow better control over the flow of work or materials through the system. Orders in the pool are sequenced for release according to their urgency, i.e., a planned release date (Eq. 1), and are released under the control of two possible card-based materials flow control mechanism: CONWIP (CONstant-Work-In-Process) and GKS (Generic Kanban System).

$$\tau_j = d_j - \sum\nolimits_{k \in R_j} b_k \tag{1}$$

Where:

τ_j is the planned release date of job j;
d_j is the due date of job j;
b_k is the lead time at workstation k;
R_j is the set of workstations in the routing of job j.

Lead times at each workstation are fixed at 5 time units based on the throughput times observed in preliminary simulation runs of this study;

In both CONWIP and GKS mechanisms the cards, which are used for authorizing production, are not part or product number specific and therefore can be acquired, for production and materials flow control purposes, by any job in the pre-shop pool waiting release. CONWIP cards are all identical, i.e. of the same type, but GKS cards are not: they are workstation specific.

Cards for each sublot of a job are attached to the job at release. Detached cards from jobs are sent back to the pre-shop pool, where they can be attached to new production job when released into the system. CONWIP cards, as many as the number of sublots, are attached to the job at release and detached when the job (or a sublot, when lot splitting is performed), completes processing at its last production stage. GKS works in different way. Since GKS cards are associated to each workstation, then GKS cards from each workstation in the routing of the job are attached to the job at release and detached, and sent back for new releases, when the job or each one of the sublots, depending on the splitting policy used (see Sect. 2.3), completes its operation at the corresponding workstation.

The role of priority dispatching is a very moderate one when order release control is applied, because the choice among jobs is limited due to short queues [1]. Thus, in this study, shop floor dispatching is based on the *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   Control Policy - Lot Splitting

Regarding lot splitting three alternative policies are analysed to determine whether splitting should be considered or not and when, i.e. before or after release to the shop for processing:

- Policy P0: The job (or order) is not split, and thus released to the shop floor and processed as a whole.

- Policy P1: The job (or order) is split before release and split sublots are released in an independent manner to the shop floor.
- Policy P2: The job (or order) is released as a whole and then split in sublots on the shop floor for independent processing.

2.4   Experimental Design and Performance Measures

The experimental factors and simulated levels of the study are summarised in Table 2. Two material flow control mechanisms, namely CONWIP and GKS, are applied to release jobs from the pre-shop pool to the shop floor. Both were tested at five card counts, i.e. number of production authorisation cards, and for the three lot-splitting policies referred above. Thus, 30 simulation cases are tested (2 release mechanisms $\times$ 5 card counts $\times$ 3 lot splitting policies). Each test case runs 100 replicates. The time horizon for a simulation case is 13 000 time units and only data of the last 10 000 time units are collected, i.e., a warm-up period of 3 000 time units is considered.

Table 2. Experimental factors and levels

Experimental factor	Levels		
Material flow control mechanisms	CONWIP		GKS
Lot splitting policies	P0	P1	P2
Number of production authorisation cards	5 levels of WIP restriction		

The number of production authorisation cards is an experimental factor in our study. CONWIP uses a single-type production authorisation card. GKS, on the other hand, requires one card type per workstation. We define the number of cards per workstation to be different in GKS. The number of cards at workstations two to six equals those of workstation one multiplied by the workstation number. We adopt this on the assumption that cards for downstream workstations of the GFS are likely to remain longer in the system than for upstream workstations and because we have a balanced GFS with identical workstation throughput times.

Concerning system performance, two types of criteria are used: (1) the ability to deliver orders (jobs) on time, and (2) the ability to provide short delivery times. To measure performance with regard to the former, the percentage of tardy jobs and the standard deviation of lateness are recorded. To measure performance with regard to the latter, the shop throughput time and the total throughput time are used. The shop throughput time refers to the time that elapses between job release and job completion. The total throughput time is the shop throughput time plus the job delay in the pre-shop pool. Note that a job is not completed until all the lots that belong to it are fully processed. Thus we can define the synchronization time of a job as the time that elapses between the completion of the first and last lot of the job.

3 Simulation Results and Discussion

Here we present and discuss the results of the simulation study described in the previous section. Section 3.1 studies the impact of the CONWIP and GKS mechanisms under lot splitting policy P1. A detailed analysis of results under the three policies, P0 to P2, is given in Sect. 3.2.
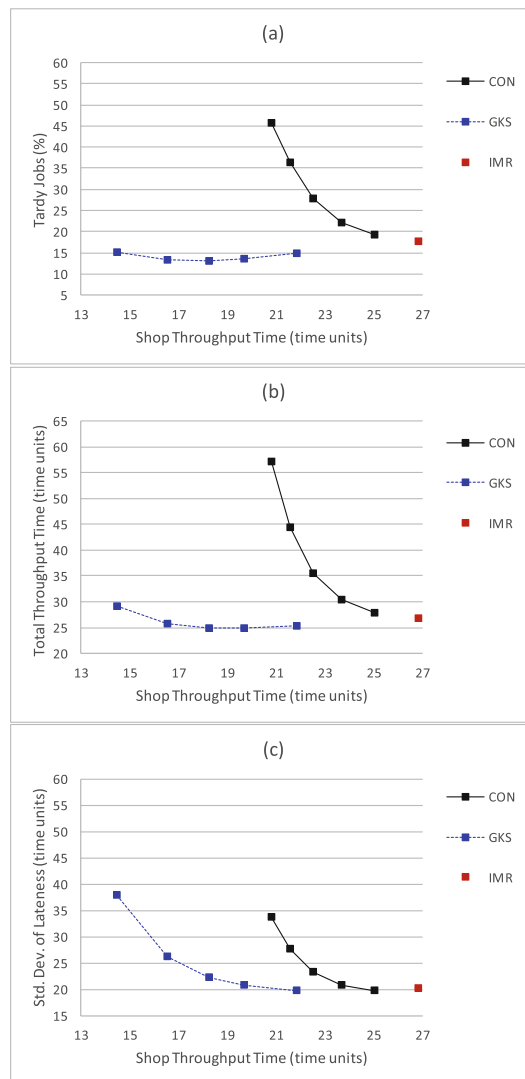

Fig. 2. Release results of release methods for: (a) percentage of tardy jobs; (b) total throughput time; (c) standard deviation of lateness.

3.1   Impact of Release Methods on Performance

Figure 2(a)–(c) plot the percentage of tardy jobs, total throughput time, and the standard deviation of lateness, respectively, against the shop throughput time for the two material flow control mechanisms. By comparing plotted curves, we can determine their performance differences for different values of card counts. A marker on a curve is the result of simulating a release method at a specific card count. Five card counts have been simulated.

Additionally, we also collect and show results for immediate release (IMR) as it was used as a base line for performance comparisons in this study. Immediate release means that when orders arrive they are immediately released to the shop floor without any restriction. The number of production authorisation cards decrease along the curve from right to left, leading to less work-in-process, i.e. less jobs on the shop floor, and therefore, according to Little's law [11], to lower shop throughput time.

As expected, IMR results in the highest level of shop throughput time. Accompanying an initial reduction in the number of cards available at the pre-shop pool is a reduction in the percentage of tardy jobs and also on the total throughput time for GKS. However, CONWIP behaves in a different way. In fact, for this mechanism, reducing the number of cards immediately leads to deterioration, i.e. to an increase of all performance measures. Germs and Riezebos [7] already concluded about the lack of balancing capability of CONWIP, expressed by the increase of total throughput time when CONWIP cards are continuously restricted. This is not the case in GKS. GKS clearly outperforms CONWIP and IMR for the percentage of tardy jobs, total throughput time and standard deviation of lateness. This can be explained by the better workload balancing capability that results from GKS controlling workload in each workstation of the system.

We can also see that the GKS performance starts deteriorating only for very low levels of card counts. This makes GKS more robust to changes in the number of card comparatively to CONWIP, which can be seen as an attractive feature for practical application.

3.2   Impact of the Lot Splitting Policy

Figure 3(a)–(d) plot the percentage of tardy jobs, total throughput time, standard deviation of lateness and the synchronisation time, respectively, against the shop throughput time for different combinations of the experimental factors.

It can be observed that splitting policies P1 and P2 perform better than the non-splitting Policy P0. This confirms our expectations that splitting jobs decrease throughput times through operations overlapping. This is independent of the material flow control mechanism applied, i.e. GKS or CONWIP.

Comparing splitting policies P1 and P2 it can be seen that policy P2 leads to better percentage of tardy jobs and total throughput time than P1. However, this is obtained at the cost of a higher standard deviation of lateness. This behaviour can be explained by the fact that releasing jobs without first splitting them requires a larger number of cards to be available at release. This means that larger jobs have a less streamlined release
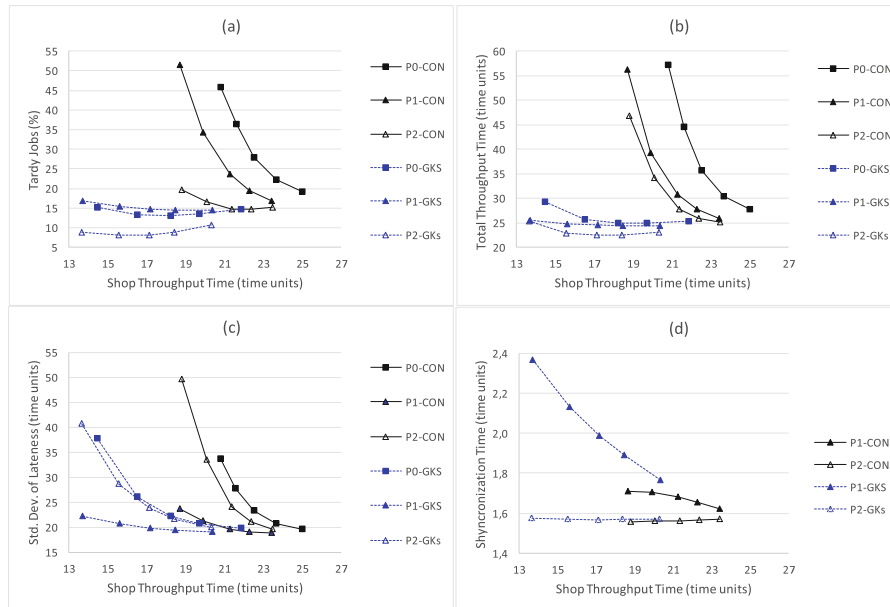
Fig. 3. Performance results for the impact of the splitting policy for: (a) percentage of tardy jobs; (b) total throughput time; (c) standard deviation of lateness and (d) synchronization time.

than smaller jobs, i.e. sometimes they can be release in due time and sometimes they cannot, due to the shortage of cards. This tends increase the standard deviation of lateness. This problem is mitigated by policy P1 that splits jobs before release. In this case, a single card per sublot is required, not the whole set of cards for a job, therefore facilitating the release of large jobs. However, this creates another problem, as we can see from Fig. 3(d): once lots are released independently they have to wait for each other after processing, to gather for the sublots of each job, increasing therefore the synchronization delay. This problem is more expressive under GKS than under CONWIP. This behaviour of the synchronization delay is the main reason why policy P1 performs worse than P2 in terms of tardy jobs and job throughput time. One explanation for the severe effects of policy P1, on the synchronization delay under GKS is the need to have available the right cards from the right workstations required by the sublot routing. This tends to make the time interval for the release of all lots of the same job to be highly extended in relation to the case of policy P2, which releases the whole job at the same point in time.

4 Conclusions and Managerial Implications

This study on production logistics compares two card-based material flow control mechanisms GKS and CONWIP when applied to a make-to-order manufacturing system configured as a general flow shop. GKS was shown to outperform CONWIP in

terms of the percentage of tardy jobs, system throughput time and standard deviation of lateness. It was also shown to be highly robust to the number of production authorization cards used to run the system. In fact, it sustained high levels of performance for a range of different numbers of cards used. This behaviour lends itself to practical application of the GKS. The reason for the good performance of GKS is its load balancing capability over workstations of the manufacturing system. This quality is not shared by the CONWIP system.

Lot splitting policies have shown to have a positive impact on both CONWIP and GKS performance. Better performance behaviour of GKS in relation to CONWIP was also verified under lot splitting.

We see some important managerial implications of this study. The main one is that under make-to-order and systems configured as general flow shop, card-based mechanisms can be used to achieve high levels of performance in terms of (1) the ability to deliver orders on time, and (2) the ability to provide short delivery times, particularly when combined with lot splitting policies. The use of these mechanisms has the advantages of requiring little data handling and allowing easy visual control of the flow of materials, which can be seen as attractive attributes for practical applications.

Although general flow shops have much in common to real world manufacturing systems' configurations, future research is planned to verify if the performance behaviour of the control strategies tested still applies under different shop configurations and manufacturing settings, e.g. considering set-up times since these have been proved to have an impact on throughput times as batch sizes change [10].

References

1. Bechte, W.: Load-oriented manufacturing control, just-in-time production for job shops. Prod. Planning Control **5**(3), 292–307 (1994)
2. Carmo-Silva, S., Alves, A.C.: A framework for understanding cellular manufacturing. In: Ferreira, J.J.P. (ed.) E-Manufacturing: Business Paradigms and Supporting Technologies, pp. 163–172. Springer, New York (2004)
3. Chang, J.H., Chiu, H.N.: A comprehensive review of lot streaming. Int. J. Prod. Res. **43**(8), 1515–1536 (2005)
4. Chang, T.M., Yih, Y.: Generic Kanban systems for dynamic environments. Int. J. Prod. Res. **32**(4), 889–902 (1994)
5. Cheng, M., Mukherjee, N.J., Sarin, S.C.: A review of lot streaming. Int. J. Prod. Res. **51**(23–24), 7023–7046 (2013)
6. Enns, S.T.: An integrated system for controlling shop loading and work flow. Int. J. Prod. Res. **33**(10), 2801–2820 (1995)
7. Germs, R., Riezebos, J.: Workload balancing capability of pull systems in MTO production. Int. J. Prod. Res. **48**(8), 2345–2360 (2010)

8. Henrich, P., Land, M., Gaalman, G.J.C.: Exploring applicability of the workload control concept. Int. J. Prod. Econ. **90**(2), 187–198 (2004)

9. Jacobs, F.R., Bragg, D.J.: Repetitive lots: flow time reductions through sequencing and dynamic batch sizing. Decis. Sci. **19**(1), 281–294 (1988)

10. Karmarkar, U.S.: Lot sizes, lead times and in-process inventories. Manage. Sci. **33**(3), 409–418 (1987)

11. Little, J.: A proof of the theorem $L = \lambda W$. Oper. Res. **9**(3), 383–387 (1961)

12. Nyhuis, P., Wiendahl, H.-P.: Fundamentals of Production Logistics: Theory Tools and Applications. Springer, Heidelberg (2009)

13. Oosterman, B., Land, M.J., Gaalman, G.J.C.: The influence of shop characteristics on workload control. Int. J. Prod. Econ. **68**(1), 107–119 (2000)

14. Spearman, M., Woodruff, D.L., Hopp, W.J.: CONWIP: a pull alternative to Kanban. Int. J. Prod. Res. **28**(5), 879–894 (1990)

15. Wagner, B.J., Ragatz, G.L.: The impact of lot splitting on due date performance. J. Oper. Manage. **12**(1), 13–25 (1994)