Advances in Manufacturing Technology XXXV
M. Shafik and K. Case (Eds.)

doi:10.3233/ATDE220613

Implementing Pull Manufacturing in Make-To-Order Environments

Mohammed ALMANEI, Omogbai OLEGHE, Mohamed AFY-SHARARAH and Konstantinos SALONITIS[1]
School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield MK43 0AL, UK

Abstract. The demand for increasing product variety and customization has forced many companies to adopt a make-to-order (MTO) strategy. Traditional push-type MTO companies suffer from unstable demands, struggling to deliver on time, making them consider the utilization of pull systems to control production. In the present paper, an overview of pull systems in MTO environments is presented. Moreover, a discrete event simulation (DES) model of an MTO company in the printing and packaging industrial sector was developed and validated, in order to identify areas for improvement. DES was also used in order to evaluate the feasibility of implementing three types of pull systems: kanban, CONstant-Work-In-Process (CONWIP) and Paired Overlapping Loops of Cards with Authorizations (POLCA). The main performance indicators measured were the average WIP and the average throughput time of parts. The key findings of this project for the case study were: a) kanban is inapplicable for the current routing of parts; b) a CONWIP strategy improves the shop floor performance, but only when extra capacity is added to the extrusion workstation; c) production based on POLCA leads to the blockage of the system due to the existence of multi-routes and undirected routing.

Keywords. Lean Manufacturing, Discrete event simulation, Make-to-order.

1. Introduction

Lean Manufacturing is well established in the manufacturing sector as a philosophy focused on reducing all types of waste, which consequently improves quality and reduces production time and cost. The ultimate objective of lean is to achieve a continuous flow [1]. Continuous workflow reduces many kinds of production waste, for example, product waiting times, and increases value to the customer. Practice has shown that lean is easily adapted to make-to-stock organizations where the production demand can be predicted quite accurately. However, the increasing demand for high mix and low volume products has forced many companies to adopt the make-to-order (MTO) approach [2]. Such companies must cope with unstable customer demand and their main challenge is to meet such demand on time. It has been shown that high mix and low volume producers find it difficult to adopt lean strategies, such as Kanban and one-piece flow effectively [3] and have to rely on more traditional practices based on push-type production systems and functional layouts and implementation of simple Material Resource Planning (MRP) [4].

However, implementing pull production helps keep under control inventory as well as controlling throughput times by establishing a work in progress (WIP) cap [5], [6]. Nowadays companies are fortunate to be able to opt from many different pull systems, such as kanban, CONWIP and POLCA. The benefits of pull systems are achieved with two main features, production control based on visual signals (mostly card-based) and the establishment of a WIP cap [4]. The visual signals are used by downstream processes to authorize the transportation of parts or upstream processes to produce [6], which means jobs are pulled according to the downstream demand. The use of visual signals combined with an established WIP cap facilitates workers to detect problems such as material build-ups and product defects [7]. However, MTO manufacturing systems can be very complex and implementing pull systems can be arduous or even impossible tasks.

In order to test the actual implementation of pull systems on the shop floor, more and more companies turn to simulation software. Simulation models, being dynamic representations of the real world, ensure that experiments using this model are adequately accurate predictors of reality. Moreover, the author includes in his quantitative benefits that simulation enables evaluation of how to reduce lead-time, reduce work-in-process and improve efficiency. He also points out qualitative benefits on decision making and action taking that include clear cause and effect, reduced risk and tested assumptions.

The aim of this paper is to analyze the general feasibility and effects of pull systems in MTO environments and, particularly, using a case study of a small company in the printing and packaging industrial sector using discrete event simulation.

2. Make-to-order environments

The demand for increasing product variety has forced many companies to adopt an MTO strategy. An MTO manufacturing is one that only starts manufacturing when a customer order is received and is often used to deal with product customization. Due to this demand for unique customized goods or services, companies have tended to adopt a mass customization approach, meaning they can create variety and customization by being flexible and quickly responsive, and at a cost similar to mass production.

By analyzing previous studies, it is possible to conclude that mass customization is justified by three major ideas [8]-[11]. First, the growing flexibility of manufacturing and the improving quality of information systems, allow the production of high variety products in shorter lead times and lower costs. Second, product variety and customization are increasing over time. And last, industrial competition and shorter product life cycles force companies to adapt to be competitive in their market sector.

MTO systems only produce final products when an order is placed, meaning customers have to wait an additional time to receive a final product when compared with a traditional make-to-stock (MTS) strategy where production is based on forecasts and stock of final product is held. However, MTO strategy's biggest issue is to meet customer demand on time.

In general push systems control throughput and measure work-in-process, while pull systems control WIP and measure throughput [12]. Three major motivators have been identified for pulling production rather than pushing it [13]: reduction of congestion, easier control, and the impact of WIP limitations than the act of pulling production itself. A pull system is defined as a strategy where no upstream process produce a good or service until a downstream customer requires it. Pulling production is key in keeping the workflow under control [7]. Furthermore pull systems are characterized by reduced WIP and cycle time, better quality and ultimately, reduced costs. A pull system maintains inventory at minimum levels and products can be delivered in shorter lead times.

Hopp and Spearman [5] try to clarify what they call the misconception of pull systems, stating that “a pull production system is one that explicitly limits the amount of work in process that can be in the system. By default, this implies that a push production system is one that has no explicit limit on the amount of work in process that can be in the system”. In order to carry on with this project, this was understood as the most suitable definition of a pull system to achieve its aim. During the review of current literature, it was noticed that over the years, the implementation of pull systems in MTO environments has been attempted and studied many times. Most of the time kanban, CONWIP and POLCA were the pull systems used (figure 1).

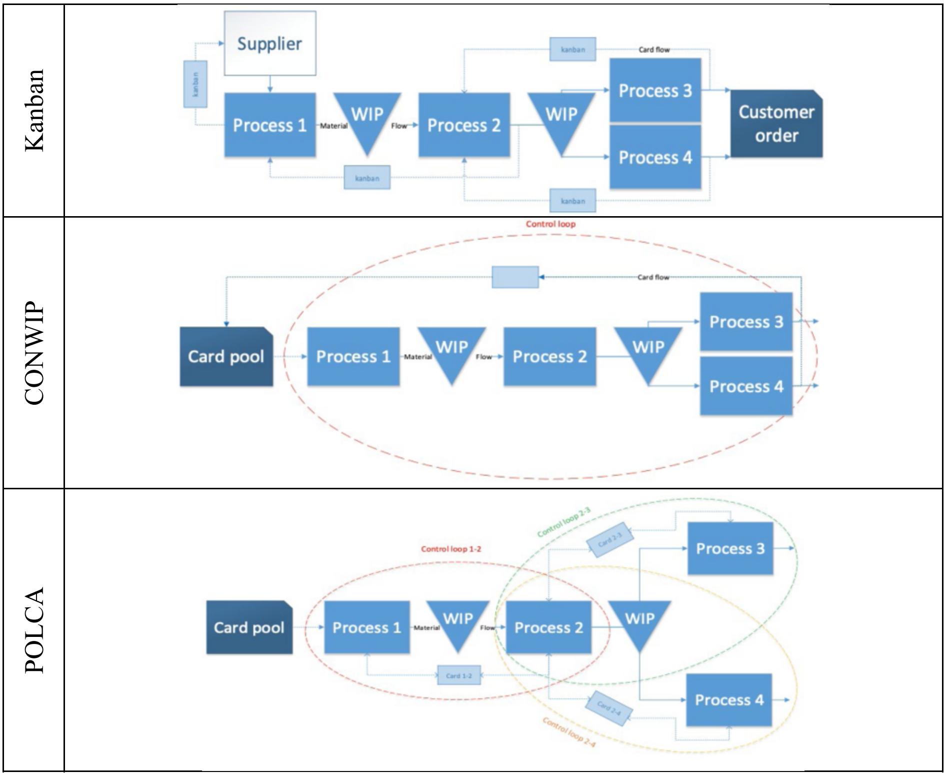

Figure 1. Kanban, CONWIP and POLCA systems (based on [14]).

3. Kanban system

Kanban is a visual signal that is used by downstream processes to trigger upstream actions in the just-in-time (JIT) process. Few attempts to use Kanban in a MTO environment have been reported. Gravel and Price [15] reported to have successfully adapted kanban to a job shop environment that manufactured outdoor sport clothing. Martin-Vega et al. [16] applied Kanban and a set of other JIT tools to a wafer fab. The achievements of the project were mostly related to applying JIT as a philosophy rather than a single tool. Rees et al. [17] compared an MRP lot-for-lot system with a kanban system. In this case, the kanban system was simulated in an improved manufacturing system, where cycle and set-up times were reduced. Obviously, major improvements were detected when comparing the simulation results with the normal production system. Finally, they appoint that in the same conditions where kanban was simulated, an MRP system could actually be the better option to reduce costs, because it required less inventory and fewer set-ups. Finally, Li [18] compared push and pull systems in a job shop environment using simulation. The author also found that a pull system should be implemented with the support of other JIT concepts, regarding a suitable shop layout and part flow. He concludes that the extraction of the overall benefits of a pull system depends mostly on the coordination and prioritization of the job shop JIT concepts, and not on one tool alone. In this case, although the effects of using a push system were always superior to when pull systems were simulated, the author realizes that, in the ideal conditions, a pull system could represent better improvements, and blames the negative results of previous studies on kanban on the ignorance of other JIT practices.

4. Constant work-in-process system (CONWIP)

An alternative pull system to Kanban is a constant work-in-process system (CONWIP). It was first proposed as a generalized form of kanban, or in other words, a set of standard production cards [7]. CONWIP production cards are assigned to the production line and are not part number specific. Moreover, when the production cards leave the system they return to the beginning of the production line where the part numbers are associated with them, according to the backlog list (a list of part types and quantities in need of production that is the responsibility of production and inventory control staff).

CONWIP performs better over simple kanban systems in MTO environments, as the part-specific WIP caps in the kanban systems have a tendency to increase the number and, consequently, the time spent on set-ups, which results in higher WIP levels [17], [18]. CONWIP cards are specific to a process and thus it is possible to sequence jobs strategically in order to reduce the number of set-ups [12]. Furthermore, CONWIP is more compatible with some production environments where Kanban is useless, due to too many parts or long set-up times [12].

As explained before, the objective of using kanban is to define an inventory limit between two workstations. Each workstation should then work in order to replace parts that have been withdrawn by the following downstream processes, keeping inventory levels as high as possible. Gaury et al. [19] describe the principle of CONWIP as the combination of the low inventory levels achieved with kanban, with the high throughput levels of push systems. This principle is the reason for CONWIP to be often designated as a hybrid push/pull system.

CONWIP has been compared with different order release strategies using simulation as well as compared to the performance of an MRP system with a CONWIP strategy [20]. In both cases, CONWIP outperformed the other production strategies in terms of mean WIP, mean throughput and percentage of late job fulfilments. Finally, Marek et al. [21] report a simulation study and concluded that most of the time, under the same circumstances, CONWIP systems outperform kanban systems. They also add that CONWIP systems are easier to implement and adjust, as only one set of cards is needed.

5. Paired-cell Overlapping Loops of Cards with Authorization (POLCA)

A different pull system approach to kanban and CONWIP is the Paired-cell Overlapping Loops of Cards with Authorization (POLCA). POLCA was firstly introduced by Suri

[22] to shorten lead times and react quickly to customer demand. POLCA is defined as a material control system designed for low volume and high-variety manufacturing. There is clearly a lack of previous studies in the current literature that discuss the implementation of POLCA systems. For example, Vandaele et al. [23] presented a case study regarding the design and implementation of a POLCA system and mentioned that practical issues had emerged but fail to provide a clear explanation of these problems.

Fernandes and Carmo-Silva [24] compared Generic POLCA (GPOLCA) with MRP and POLCA in a multi-station and high variety manufacturing environment under a changing product mix and variable customer demand. They concluded that GPOLCA was more robust to change in demand and that it outperformed the two remaining systems by achieving the same throughput while using lower levels of WIP. Later, Germs and Riezebos [25] compared POLCA, CONWIP, and multiple CONWIP (m-CONWIP) using simulation. Their findings state that the three pull systems could reduce the total throughput time in different magnitudes, depending on the pull system used. Finally, Mortágua et al. [26] compared card-based production control mechanisms in MTO environments using simulation, namely CONWIP, GPOLCA and Generic Kanban Systems (GKS). They concluded that GPOLCA and GKS could be effective in MTO environments but only if certain adaptions are implemented.

6. Case study and results

The case study of this project is based on a company that provides printed labels and flexible packaging wrapping solutions for food and powdered products. Products are developed in a multi-product manufacturing line with high flexibility to process many different products using the same machinery, which is laid out on the shop floor. Figure 2 presents the DES model that represents the actual shop floor. Products are finished in rolls, cut pieces, or ready pouches. The nature of this manufacturing process allows the production of an infinite number of products because all products undergo the customisation process, printing.

The customer needs are driven by short lead times, quality items and good customer service. The job orders are received directly from customers and can vary in quantity and product type. A total of seven different kinds of standard products, which vary in quantity, are provided and undergo a production process flow that can be composed of 6 different processes and their combinations.

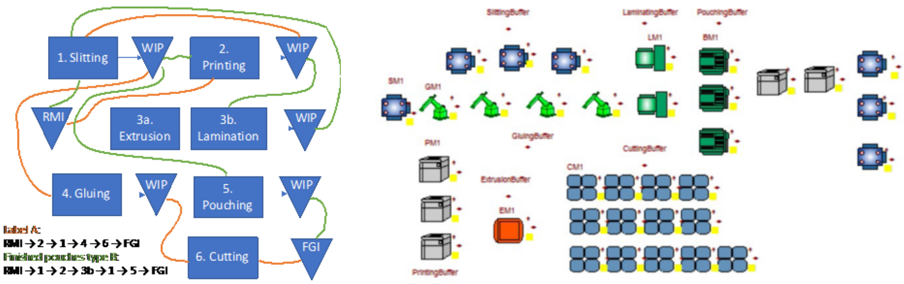

Figure 2. Left: process flow map for two example products, right: general shop floor layout in the DES model of the case study.

The first analysis focused in simulating the as-is situation, where the system is work based on pushing inventory. The simulation highlighted that printing, along with the extrusion process, are the major bottlenecks of the production. The printing buffer reaches queue levels of over one hundred parts, just for the first month of production. The simulation also indicated that the number of parts in the system (WIP) reach very high levels, sometimes larger than the number of shipped parts. This aggravates the size of queues waiting to be processed, and ultimately the throughput time of the parts. The push system was also tested under the scenario that setups are optimized (through improvements such as SMED and 5S) to almost zero times. Even under this scenario, the printing, and the extrusion processes are still the bottlenecks.

A production based on Kanban systems involves storing standard inventory between the processes of each route. Due to the nature of the case study's manufacturing process, which depends on an early customization printing process, it was acknowledged that this type of pull system is incompatible with this type of MTO strategy. The early customized printing process makes it impossible to store standard inventory downstream.

For assessing the idea of introducing a CONWIP strategy, various simulation models were built while varying the WIP cap. The extrusion station was still identified as the bottleneck. In order to achieve a working CONWIP strategy, it would be necessary to increase the extruding capacity of the system, and for this, an extra extruding machine was added to the simulation model. Adding an additional extruding station, the bottleneck station becomes the lamination process. The variation of the WIP cap has an impact on the number of shipped parts. By simulating the various scenarios according to the different WIP caps, it is possible to identify the minimum WIP cap that can lead to a steady-state production (figure 3-left).

The POLCA strategy is based on keeping inventory in buffers before and after each workstation. The implementation of POLCA for the present case study, resulted in slitting and printing machines being blocked. This happens due to a large number of routes and undirected routing of the parts. Concluding that CONWIP is the only applicable pull system for this case study, but only by neglecting setup times and raw material delays, and adding extra capacity to the extrusion workstation, it is only fair if this scenario is compared to the case study under the same conditions. The chosen CONWIP scenario establishes the WIP cap at 40, the minimum level for the maximum number of shipped parts is shown in figure 3-right.

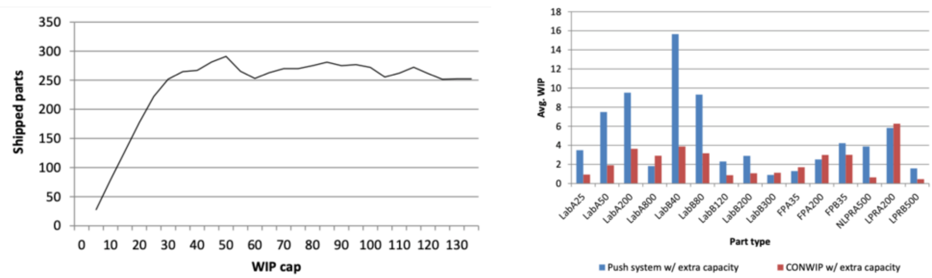

Figure 3. Left: ped parts and WIP cap using CONWIP with extra extrusion capacity, right: Avg. WIP of parts between the improved CONWIP system and push-type production for the first month of production.

7. Conclusions

The main aim of this project was to analyse the feasibility and effects of pulling production in MTO environments, particularly in a multi-routing MTO environment in the printing and packaging industrial sector, using discrete event simulation. The DES model of the as-is situation of the case study reveals that it is not possible for the manufacturing system to accept all customer orders per month. Over time, jobs pile up more and more before the printing and extruding workstations. Also, setup times and raw material delays aggravate the situation.

In this case study, a pull system using kanban was considered to be non-applicable due to the impossibility of keeping standard inventory between downstream workstations after an early customisation process.

When analysing the scenario of adapting the case study to a POLCA methodology, it was understood that the existence of multiple routes and undirected routing of parts caused the blockage of the entire production system, as two occupied workstations were waiting for each other to be available, and transfer the processed parts to each other.

The main output of this project is the result of using CONWIP as a pull manufacturing strategy in the case study, but only by increasing the extrusion capacity. This pull system scenario is compared to a push system with the same increased capacity. The established WIP cap in this scenario showed that a similar number of shipped parts can be achieved by rationalising customer orders, and so injecting fewer jobs into the system by following an adequate job order release.

Acknowledgements

The authors would like to acknowledge the work completed by Cranfield MSc student Joseph Tucker as part of this research.

References

[1] M. Rother, R. Harris, *Creating continuous flow*. Brookline, Mass.: Lean Enterprise Institute, 2001.
[2] M.R. Melchert, M.A. Mesquita, P.G. Francischini. Lean manufacturing on make-to-order suppliers: a case study. *ICIEOM* (2006)
[3] R. Suri, *Quick response manufacturing*. Portland, Or.: Productivity Press, 1998.
[4] J. Li, Comparing Kanban with CONWIP in a make-to-order environment supported by JIT practices. *Journal of the Chinese Institute of Industrial Engineers*, 28 (2011), 72-88.
[5] W. Hopp, M. Spearman To Pull or Not to Pull: What Is the Question? *Manufacturing & Service Operations Management*, 6 (2004),133-148.
[6] T. Cheng, S. Podolsky, *Just-in-time manufacturing*. London: Chapman & Hall, 1993.
[7] W. Hopp, M. Spearman, *Factory physics*. Chicago: Irwin, 1996.
[8] C. Hart, Mass customization: conceptual underpinnings, opportunities and limits. *International Journal of Service Industry Management*, 6 (1994), 36-45.
[9] S. Kotha, Mass customization: Implementing the emerging paradigm for competitive advantage. *Strategy Management Journal*, 16 (1995), 21-42.
[10] P. Åhlström, R. Westbrook. Implications of mass customization for operations management. *International Journal of Operations & Production Management*, 19 (1999), 262-275.
[11] B. Pine, *Mass customization*. Boston, Mass.: Harvard Business School Press, 1993.
[12] M. Spearman, D. Woodruff, W. Hopp. CONWIP: a pull alternative to kanban. *International Journal of Production Research*, 28 (1990), 879-894.
[13] M. Spearman, M. Zazanis. Push and Pull Production Systems: Issues and Comparisons. *Operations Research*, 40 (1992), 521-532.
[14] M. Thurer, M. Stevenson, C. Protzman, *Card-based control systems for a lean work design*. Routledge, 2016.

[15] M. Gravel, W. Price. Using the Kanban in a job shop environment. *International Journal of Production Research*, 26 (1988), 1105- 1118.
[16] L. Martin-Vega, M. Pippin, E. Gerdon, R. Burcham. Applying just- in-time in a wafer fab: a case study. *IEEE Transactions on Semiconductor Manufacturing*, 2 (1989), 16-22.
[17] L. Rees, P. Huang, B. Taylor B. A comparative analysis of an MRP lot-for-lot system and a Kanban system for a multistage production operation. *International Journal of Production Research*, 27 (1989), 1427-1443.
[18] J. Li. Simulation-based comparison of push and pull systems in a job- shop environment considering the context of JIT implementation. *International Journal of Production Research*, 41 (2003), 427-447
[19] E.G.A. Gaury, H. Pierreval, P.C. Kleijnen PC. An evolutionary approach to select a pull system among Kanban, Conwip and Hybrid. *J. Intell. Manuf.*, 11 (2000): 157-167.
[20] L. Roderick, J. Toland, F. Rodriguez. A simulation study of conwip versus MRP at Westinghouse. *Computers & Industrial Engineering*, 26 (1994), 237-242.
[21] R.R. Marek, D.A. Elkins, D.R. Smith, D.R. Understanding the fundamentals of Kanban and CONWIP pull systems using simulation. In Proceedings of the 2001 Winter Simulation Conference, ed. B.A. Peters, J.S. Smith, D.J. Medeiros & M.W. Rohrer, p.921-929, Arlington, 2001.
[22] R. Suri. *Quick response manufacturing*. Portland, Or.: Productivity Press, 1998.
[23] N. Vandaele, I. Van Nieuwenhuyse, D. Claerhout, R. Cremmery. Load-Based POLCA: An Integrated Material Control System for Multiproduct, Multimachine Job Shops. *Manufacturing & Service Operations Management*, 10 (2008), 181-197.
[24] N.O. Fernandes, S. Carmo-Silva. Generic POLCA: A production and materials flow control mechanism for quick response manufacturing. *International Journal of Production Economics*, v. 104 (2006), 74-84.
[25] R. Germs, J. Riezebos. Workload balancing capability of pull systems in MTO production. International Journal of Production Research, 48 (2009), 2345-2360
[26] J. Mortágua, N.O. Fernandes, S. Carmo-Silva. Comparing Card-Based Production Control Mechanisms in MTO Production. 28th European Simulation and Modelling Conference - ESM'2014. October 22-24, 2014. FEUP - University of Porto, Porto, Portugal