Int. J. Production Economics 144 (2013) 200–208

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

journal homepage: www.elsevier.com/locate/ijpe

# Performance analysis and optimization of hybrid manufacturing systems under a batch ordering policy

Eman Almehdawe*, Elizabeth Jewkes

*Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*

ARTICLE INFO

*Article history:*
Received 7 January 2011
Accepted 6 February 2013
Available online 16 February 2013

*Keywords:*
Hybrid MTS–MTO manufacturing
Queueing theory
Matrix-analytic method

ABSTRACT

We consider a stochastic two-stage hybrid manufacturing system for a single product where semi-finished goods are Made-To-Stock (MTS) and then differentiated when demand is realized through a Make-To-Order (MTO) stage. Inventory for semi-finished goods is held between the two stages. We introduce a batch ordering policy to permit economies of scale in ordering due to a cost associated with each order placed. We use the Matrix-analytic method to evaluate system performance under this ordering policy. Afterwards, we develop an optimization model to find the optimal intermediate buffer size and the optimal replenishment order quantity for the system. We show that a base stock policy is sub-optimal in the presence of a replenishment cost for semi-finished goods. The savings from adopting the batch ordering policy can be high while the response time for most customer orders is not affected.


## 1. Introduction

Increased competition among manufacturing companies where flexibility, quality, cost and response time play a critical role is a major challenge in today's industry. Key to most companies is to increase product variety and yet maintain short delivery lead times. To achieve these goals, new manufacturing strategies have evolved. Hybrid manufacturing is one such trend where products are produced via a two-stage process: they are partially assembled using a Make-To-Stock (MTS) stage, and then customized based on customer requirements in a subsequent Make-To-Order (MTO) stage. Intermediate inventory of partially finished goods is stored in a buffer between the two stages.

Hybrid MTO–MTS systems are one of the many interrelated and complementary strategies for designing and managing supply chains identified by Mikkola and Larsen (2004). These strategies are mass customization (see, e.g. Silveira et al., 2001), postponement and modularization. Hybrid MTS–MTO utilizes certain aspects of each of these three strategies to deliver customers with a customized product at minimal cost.

MTS systems are usually suitable for high volume and low variety products, whereas MTO systems are suitable for low volume and high variety products. The hybrid manufacturing strategy, as suggested by Youssef et al. (2004), combines the benefits of both MTS and MTO. It reduces the order fulfillment delay relative to MTO because finished goods are made from partially completed items. At the same time, it can lower inventory costs since inventory is held only for semi-finished goods and raw components, rather than finished goods (Gupta and Benjaafar, 2004). The hybrid MTS–MTO strategy is widely used in the electronics industry and other similar markets where many product configurations can be produced from intermediate interchangeable modules (Gupta and Weerawat, 2006). A well-known and successful example of a company that has adopted this strategy is Dell Computer Corporation (Serwer, 2002).

Research on MTS–MTO systems falls into four primary categories: The first is research on inventory management for the semi-finished goods between the two stages, e.g. Gupta and Benjaafar (2004). Second, research on the optimal point of differentiation, e.g. Gupta and Benjaafar (2004) and Jewkes and Alfa (2009). Third, research that explores the optimal configuration for semi-finished products from the point of view of commonality among different final products, e.g. Swaminathan and Tayur (1998). Finally, there is research that looks into coordination between the two stages, e.g. Gupta and Weerawat (2006). Our work falls in the first category.

Most research on hybrid MTS–MTO manufacturing systems use a base stock, or one for one replacement policy for replenishing the intermediate buffer. That is, the MTO stage is triggered to replenish the semi-finished goods buffer by one unit each time a semi-finished item is removed from the intermediate inventory as a result of a customer order. This policy is widely adopted in the literature because of its ease of modeling, although such a policy is not necessarily optimal when there is an ordering cost associated with each replenishment. Veinott (1965) showed that when an ordering cost exists, a batch replenishment policy is

* Corresponding author. Tel.: +1 519 888 4567.
*E-mail addresses:* ealmehda@uwaterloo.ca (E. Almehdawe), emjewkes@uwaterloo.ca (E. Jewkes).

0925-5273/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpe.2013.02.005

optimal. In an MTS–MTO context, batch replenishment occurs when the MTS stage is signaled to restock the intermediate inventory only after a specified number of semi-finished items have been removed from the buffer. Compared to a base stock policy, batch replenishment may expose the system to longer customer lead times if the intermediate buffer is empty when a customer order is placed. This is because the customization process cannot begin until a unit of semi-finished goods is available. However, in the presence of replenishment costs, it is worth considering the impact of this tradeoff.

The primary contribution of this work is to introduce a batch replenishment policy to hybrid MTS–MTO manufacturing systems. The second contribution is to provide an *exact* analysis for various performance measures that capture the tradeoffs between replenishment costs and customer lead times. Furthermore, we formulate an optimization problem to find the optimal values of the intermediate goods buffer and batch size decision variables. Our batch ordering policy is a more realistic generalization of the base stock policy and captures a common manufacturing problem faced by decision makers.

While one might conclude that batching customer orders prior to issuing a replenishment order for intermediate goods would lead to increased customer lead times, our work shows that batching orders affect only a small percentage of customer orders. We are also able to quantify when the base stock policy produces sub-optimal solutions. To our knowledge, our paper is the first to examine the effects of batching replenishment orders when there are ordering/setup costs associated with each order placed for the semi-finished goods.

The rest of the paper is organized as follows: In Section 2, we review some of the relevant literature. We describe the model in detail and develop a Markov chain representation in Section 3. In Section 4, we develop an optimization model to find the optimal intermediate buffer size and order batch size. Numerical analysis is reported in Section 5. Finally, we conclude our work with the main findings and point out future research opportunities in Section 6.

## 2. Literature review

Hybrid MTS–MTO systems have been studied in the literature from several different perspectives, e.g. multi-echelon inventory systems, tandem queues, and delayed product differentiation. We focus on two streams of literature that provide a base for our research. First, multi-echelon inventory systems, particularly the serial inventory system or multi-stage Production/Inventory systems (PI) systems. These models generally consist of multiple production stages separated by inventory buffers. Whereas most multi-stage PI systems have a buffer for finished goods inventory at the final stage to immediately fulfill customer orders, MTS–MTO systems have no finished goods inventory. When customer orders are placed, there is a lead time while the semi-finished items are customized to fulfill the order.

The research that is closest to ours is Lee and Zipkin (1992). They study a multi-stage production system with limited intermediate buffers for semi-finished goods. It was modeled using a series of single server queues with exponential service times and a base-stock replenishment policy for replenishment of the intermediate buffers. While exact solutions for the system performance measures were available for either infinite or zero intermediate buffer sizes, they develop an approximation scheme for the expected number of backorders, and the expected number of units in semi-finished goods inventory. In contrast, our model provides exact results for the system performance measures.

Gupta and Selvaraja (2006) extend Lee and Zipkin's model by providing a near-exact solution for the system performance measures for a capacitated serial supply system. In further related work, Gupta and Weerawat (2006) investigate the coordination effects between a manufacturer and a supplier and Liu et al. (2004) present a decomposition approach to find the optimal inventory buffer sizes for a multi-stage inventory system while maintaining a certain service level for customer orders.

None of the above papers assume a setup time or setup cost to place an order. In contrast, Li and Liu (2006) use an (s,S) policy in a two-stage tandem production system, for which they assume that stage 1 produces semi-finished products to stock, and stage 2 produces finished products to stock. With a significant setup time at stage 1, they seek an optimal reorder point for the semi-finished goods, and an optimal order size for stage 1 that minimizes the WIP between the two stages while keeping a high utilization at stage 2. Their model is similar to ours in that it considers setup costs for stage 1. Their decision variables (s,S) are similar to our decision variables (B, S), however they do not consider the potential for customer order lead time delays. Instead, they focus on maintaining high utilization of stage 2.

Berman and Sapna (2001) study a (s,Q) policy for a service facility with Poisson arrivals and exponential replenishment lead times. They assume that there is limited space for waiting customers, and that any arrival beyond the limit is lost. They use linear programming to find the optimal server rate that minimizes the long run inventory holding, ordering, waiting and lost sales costs. Their model is similar to ours in which they consider ordering and delay costs. However, they assume that no queue is formed in front of the service stage, which implies that items are not released from the inventory buffer unless the server is idle. Our model assumes that customer orders can be queued. As a result, our model captures two types of customer order delays; delay due to customization of semi-finished goods, and delays if the order is backlogged.

Some of the work related to hybrid MTS–MTO system address the question of whether to produce a product using a pure MTS strategy, or a pure MTO strategy, or using a hybrid MTS–MTO policy, e.g. Zaeprpour et al. (2009), and Rafiei and Rabbani (2009) and others.

Another stream of research that is closely related to the hybrid MTS–MTO system is on Delayed product Differentiation (DD). DD is a hybrid system in which a common product platform is built to stock and then differentiated by assigning customer specified features after the demand is realized (Gupta and Benjaafar, 2004). Most of the analytical models developed for DD systems focus on the optimal intermediate buffer size as well as the optimal point of differentiation. Swaminathan and Tayur (1998) exploit the component commonalities in the case of IBM when end products are differentiated based on customer orders. They find the optimal configuration and inventory levels of vanilla boxes (semi-finished products) using a two-stage stochastic program. However, their model does not capture the production or the procurement process for the semi-finished goods, they assume that those components are instantaneously available. In their optimization problem, the costs considered are holding inventory costs and stock out costs when customers orders are not satisfied. In our work, we model the production/procurement process for the semi-finished goods. We also consider ordering costs in addition to holding inventory costs and delay costs.

Gupta and Benjaafar (2004) model a two-stage production system in which both the characteristics of the MTS–MTO system and delayed differentiation are considered. They assume that inventory for semi-finished and finished products are controlled via a base-stock policy. They use Lee and Zipkin's (1992) approximation scheme to develop performance measures for the system