Improving the Efficiency of TOC Demand-Pull and Buffer Management by Incorporating Demand Information Using EWMA in Semiconductor Manufacturing

Yung-Chia Chang [1,a,*] and Shu-Wen Liang[2,b]

[1,2]Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan

[a] yungchiachang@mail.nctu.edu.tw, [b] newwen68@gmail.com

*Corresponding author

Keywords: Theory of Constraints, Demand-pull, Buffer management, Exponentially Weighted Moving Average, Rolling forecast.

Abstract. Products in the semiconductor manufacturing industry usually have characteristics of having short life-spans, volatility in demands and long lead times. Applying the demand-pull approach and buffer management (DPBM) suggested by the Theory of Constraints (TOC) to manage the inventory of such products does not efficiently respond to the unpredicted demand change thus could potentially result in either excessive inventories or shortages. This research proposed a new buffer adjustment mechanism to improve DPBM approach by applying the methodology of Exponentially Weighted Moving Average (EWMA) to integrate the true demands and the rolling forecasts of market in order to make the stock replenishing decision. Real product demand data provided by a wafer foundry in Taiwan is analyzed to demonstrate the effectiveness of the proposed approach. The study results show that the proposed idea can indeed keep a lower inventory level and provide a higher service standard in contrast to the traditional DPBM method.

1. Introduction

Theory of Constraint [1]suggested to use demand-pull along with buffer management (DPBM) to decide on when and how much to replenish inventory. DPBM is a theory that is easy to understand for replenishing stocks. The required replenished amount depends on the demand, and the safety inventory is based on the inventory level. It is different from the traditional replenishing strategy such as fixed period ordering system and fixed quantity system. This theory is being appreciated by the industry as it is easy to understand and implement. There are some studies that applied this methodology for inventory management and, furthermore, found that the efficiency has been greatly improved[2-4].

However, Chang et al. [5] discovered that the application with the traditional method of DPBM in managing semi-conductor inventory does not result in great efficiency, especially with respect to products with fluctuating demand. On top of that, Hung et al. [6] claims that different parameters should be used accordingly with different type of products while using DPBM in inventory management and they have designed a decision support system in order to search for different

ISBN: 978-605-64453-0-9

parameters. Chang et al. [7] found that if a certain product has a short life span with long lead times, there is often times where there are surplus in inventory or a shortage due to the fact that the inventory cannot immediately respond to sudden rapid changes in demand when using classical DPBM to make replenishment decision. They proposed a method of replenishing stocks by using information of future demands forecasted by the downstream market by in order to manage the inventory buffer effectiveness. On the other hand, Chang and Lee [8] proposed to use the exponential weighted moving averages (EWMA) of actual demands to predict the demand trend in order to improve the effectiveness of DPBM while dealing with long-lead time and fluctuating demands. They designed a set of rules to follow when making inventory replenishment.

This research is aiming to combine the information of actual demands along with the market demand forecasts by EWMA technique and rolling forecasts. With this newly proposed method, we proposed a new set of rules for replenishing stocks in each replenishing cycle and also when to adjust the inventory buffer. The rest of this paper is organized as follows. Section 2 described the proposed approach. Real data provided by a famous wafer foundry in Taiwan is analyzed to verify the feasibility and effectiveness of the proposed approach in section 3. Section 4 concludes this study.

2. Methodology

2.1 Problem description and assumption

With respect to the unique attributes of semi-conductors, this research aims to use EWMA by combining the information of past demands and future forecasts of demands then propose different methods to manage inventory buffer, lower the average inventory level and also increase the service level.

Assume that there are only two roles in the supply chain, a semi-conductor manufacturer and the customer, as can be seen in fig.1. By the theory of demand-pull, the warehouse would be the buffer within the supply chain in order to fulfill customers' demands. Without adjusting the inventory buffer, the number of stocks being reordered will be the number of stocks being sold to customers. Standing on the semi-conductor manufacturer's point of view, this number of stocks being replenished will be looked at as number of batches ready to be produced.

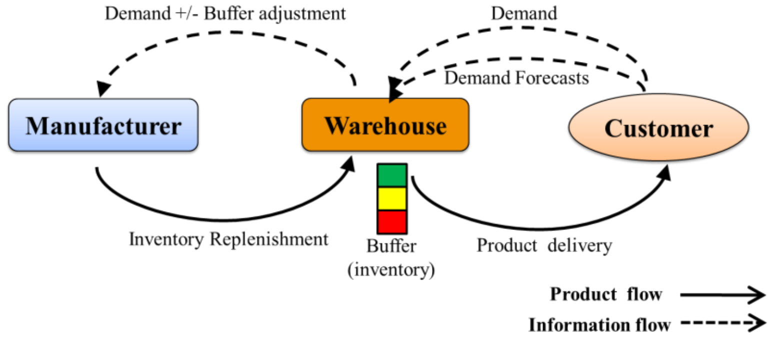

Fig. 1 Conceptual model of the DPBM considered in this study

The assumptions of our model are listed as below:

1. There is only one semi-conductor manufacturer and only on customer.
2. The replenishment decision is made on a weekly basis.
3. The customer will provide the information of current demand and also the information on future demands of every period within that specific lead time.
4. The replenished (production) lead -time is known and fixed.
5. The mount of replenished requests will be filled and arrived all together after one replenished lead-time without any defective items.
6. The ordering cost is very low and thus can be ignored.
7. Backlog is allowed. The backlogged demand needs to be satisfied before regular demand.

2.2 Buffer adjustment by combining EWMA with forecasted information of demand

An initial target buffer is set within the traditional method of DPBM in order to fulfill the demand during the lead time. There are several alternatives to determine the initial target buffer per the traditional DPBM. This research considers setting the initial target buffer as the total demand during the previous lead time and multiplies it by 1.5. As suggested by the conventional buffer management, the target buffer is divided three equal sections: the top 1/3 is called the *green* zone, the middle 1/3 section is the *yellow* zone, and the bottom 1/3 is the *red* zone. Under the circumstances of not adjusting the target buffer, the sum of the inventory on hand (*OH*) and the inventory in-transit is the target buffer. On-hand inventory is computed by Eg.1.

$$OH_t = OH_{t-1} + POR_t - D_t \quad (1)$$

In Eq.1, POR_t is the t^{th} period's Planned Order Receipts (*POR*); D_t is the t^{th} period's actual demand.

In order to enhance the ability of controlling the inventory level in the buffer zone and also to accurately decide on how much to replenish by the trend of demand, we combine the information of past demand and market demand forecasts to establish a new rule to adjust the target buffer and also to decide on the quantity for replenishing. The entire process is carried out via demand-pull that is based on the actual demand within that specific period; at the same time, the decision regarding replenish quantity is also based on the trend of market demand forecast provided by the customer.

2.2.1 Parameter and Formulae

There are two values of EWMA in the proposed method. The first EWMA is calculated by the actual demand; the second EWMA is generated by forecasted information from downstream. The two values of EWMA are then integrated with the inventory level with respect to the buffer zone in that specific period to decide on the timing for buffer adjustment. The following is the brief introduction of the parameters and formulae within this research paper:

1. EWMA smoothing constant (λ): Chen *et al.* [9] proposed a method to determine the smoothing constant (λ) to be used in calculating EWMA per the length of lead time, as shown in Eq. 2.

$$\lambda = \frac{2}{1 + LT} \quad (2)$$

LT in Eq. 2 represent the lead time which is the time in between orders placed and orders received.

2. EWMA value of actual demand ($EWMAD_{t+1}$): Once the actual demand of t^{th} period (D_t) is known, Eq. 3 can be applied to calculate the value of EWMA in the period of t+1.

$$EWMAD_{t+1} = \lambda \times D_t + (1 - \lambda) \times EWMAD_t \quad (3)$$

3. The change in ratio of EWMA value of actual demand ($\Delta EWMAD_t$): The ratio of the change in value of EWMA in the Period *t* is calculated by Eq. 4

$$\Delta EWMAD_t = \begin{cases} \frac{EWMAD_{t+1} - EWMAD_t}{EWMAD_t} & ,if\ t > 1 \\ 0 & ,if\ t = 1 \end{cases} \quad (4)$$

4. The sum of change in ratio of EWMA value of actual demand (*sum of* $\Delta EWMAD_t$) is the sum of proportional change in value of EWMA from the previous lead time to the current period, as calculated by Eq. 5. This term is used to observe the trend of actual demand.

$$Sum\ of\ \Delta EWMAD_t = \sum_{t-LT}^{t} \Delta EWMAD_i \quad (5)$$

5. EWMA value of forecasted demand ($EWMAF_{t,t+i}$) is obtained by Eq. 6.

ISBN: 978-605-64453-0-9

$$EWMAF_{t,t+i} = \begin{cases} \lambda \times F_{t,t+i} + (1-\lambda) \times EWMAD_i, & if\ i = 1 \\ \lambda \times F_{t,t+i} + (1-\lambda) \times EWMAF_{t,i}, & if\ i > 1 \end{cases} \tag{6}$$

In Eq. 6, $F_{t,t+i}$ is the forecasted demand of period $t+i$ provided in the t^{th} period in order to calculate $EWMAF_{t,t+i}$, the EWMA value of forecasted demand of period $t+i$ based on the rolling forecast provide in period t, to observe the future trend.

6. The change in ratio of EWMA value of forecasted demand: By using the EWMA value of actual demand of period t as the initial value, the relative difference between the two consecutive EWMA values of forecasted demand is calculated by Eq. 7.

$$\Delta EWMAF_{t,t+i} = \frac{EWMAF_{t,t+i+1} - EWMAF_{t,t+i}}{EWMAD_t},\ i = 1\ to\ LT\text{-}1 \tag{7}$$

7. The sum of change in ratio of EWMA value of forecasted demand ($sum\ of\ \Delta EWMAF_{t,LT}$) is obtained by Eq. 8.

$$sum\ of\ \Delta EWMAF_{t,LT} = \sum_{i=1}^{LT} \Delta EWMAF_{t,i} \tag{8}$$

8. The weight of real demand (ω): ω (0≤ω≤1) represents the relative importance of the trend of EWMA of actual demand. It can be set by decision makers. When the rolling forecasting from the market is reliable, ω can be set as small as possible; this means that the information of market forecasts weighs more. When ω is set to 0.5, it means that the actual demand shares the same degree of importance as the forecasted market information.
9. Trend indicator of the t^{th} period (k_t) summarizes the demand trend and is calculated by Eq. 9.

$$k_t = \omega \times sum\ of\ \Delta EWMAD_t + (1-\omega) \times sum\ of\ \Delta EWMAF_{t,LT} \tag{9}$$

10. Trend threshold of the t^{th} period (R_t) is used to be compared with the value of k_t in order to determine the demand trend. If the value of k_t is greater than R_t, this means that the demand trend is increasing; if the value of k_t is smaller than - $R_{t,}$, this means that the demand trend is decreasing. Otherwise, demand trend is not apparent. The value of R_t is calculated by Eq.10.

$$\frac{Z_{1-\alpha} \times \sqrt{LT} \times \sigma_t}{EWMAD_{mean,t}} \tag{10}$$

In Eq. 10, $EWMAD_{mean,t}$ and σ_t are the mean and standard deviation of EWMA from the period of t-LT+1 to period t. $Z_{1-\alpha}$ is the inverse value of standard normal distribution given the probability of 1-α. If a 90% of service level is preferred, then we can set α as 0.1.

11. Expected inventory: Let $FOH_{t,t+i}$ represent the expected inventory of period $t+i$ calculated at the t^{th} period, which can be calculated by Eq. 11.

$$FOH_{t,t+i} = \begin{cases} OH_t + POR_{t+i} - F_{t,t+i}, & if\ i = 1 \\ FOH_{t-1,t+i} + POR_{t+i} - F_{t,t+i}, & if\ i > 1 \end{cases} \tag{11}$$

2.2.2 Decision of replenishing quantity and buffer management

The purpose of this study is to apply the EWMA method to integrate the actual demand along with market demand forecasts such that the efficiency of classical DPBM approach can be further improved. This research proposed a set of decision rules to determine the replenishment and the size of target buffer. Based on the type of demand trend, buffer zone of the inventory level located, expected inventory, the rules for replenishment decisions and buffer management are organized in Table 1.

ISBN: 978-605-64453-0-9

2.3 Key Performance indicators

Average inventory (*AI*), calculated by Eq. 12, and service level (*SL*), calculated by Eq. 13, are used as the key performance indicators to evaluate the effectiveness of the inventory replenishment.

$$AI = \frac{\sum_{t=1}^{T} OH_t}{T} \tag{12}$$

$$SL = 1 - \frac{\sum_{t=1}^{T} OS_t}{\sum_{t=1}^{T} D_t} \tag{13}$$

In Eq. 12, T represents the total number of periods in the planning horizon. In Eq. 13, OS_t stands for the Out of Stock Quantity (i.e., the amount of shortage) for the t^{th} period.

Table 1. Rules for replenishment and buffer management.

Demand trend	Zone of Inventory level	Expected inventory	Rules for replenishment decision and buffer management
Decrease	red	Expected inventory > 1/2 of target buffer	Lower target buffer. Do not replenish.
		Expected inventory <1/2 target inventory; no shortage	Replenish the amount of current demand.
		Shortage in expected inventory	Replenish the sum of current demand and the amount of shortage.
	yellow		Replenish the amount of current demand.
	green		Lower target buffer. Do not replenish.
Not apparent		Expected inventory > 1/2 of target buffer	Do not replenish.
		Expected inventory <1/2 target inventory; no shortage	Replenish the amount of current demand.
		Shortage in expected inventory	Replenish the sum of current demand and the amount of shortage.
Increase	red		Increase target buffer. Do not replenish.
	Yellow		Replenish the amount of current demand.
	green	Expected inventory > 1/2 of target buffer	Do not replenish.
		Expected inventory <1/2 target inventory; no shortage	Replenish the amount of current demand.
		Shortage in expected inventory.	Replenish the sum of current demand and the amount of shortage.

3. Case study

In this section, a real case scenario is used to compare the effectiveness between the proposed method, the method of combining EWMA with past and forecast information, and the classical DPBM method. The data of Product X is provided by a famous wafer foundry in Taiwan. The product is of: (1) lead time is 9 weeks; (2) warehouse inventory is examined on a weekly basis, and request for replenishment is sent out every week; (3) initial target buffer is based on the sum of actual demand during the period of previous lead time and multiplies by 1.5; (4) if the target inventory requires adjusting, the ratio of adjustment is 1/3. Fig. 2 depicts the actual 104 weeks of demand of Product X.

ISBN: 978-605-64453-0-9

The demand of this product The average and standard deviation of the demand is 227.8 and 104.3, respectively. The coefficient of variance (CV) of the demand is 0.46.

Figure 3 and Figure 4 show the inventory on-hand by using the classical DPBM and the proposed approach, respectively.

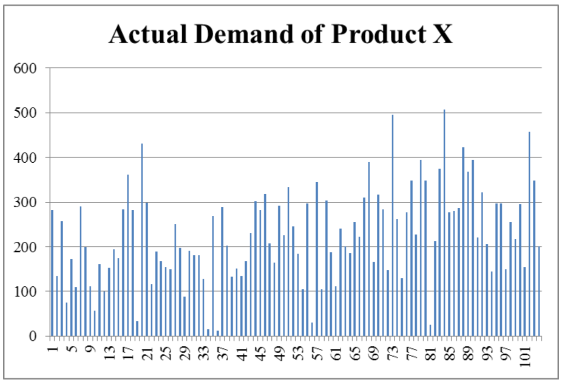

Fig. 2. Actual demand of Product X.

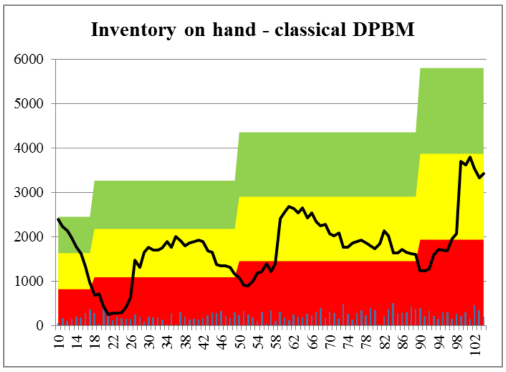

Fig. 3 Inventory on hand by the classical DPBM along with buffer adjustment.

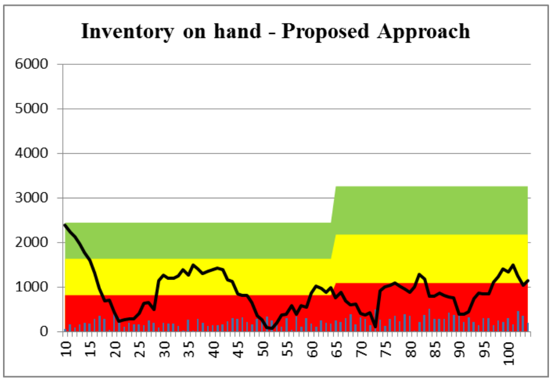

Fig. 4 Inventory on hand by the proposed approach along with buffer adjustment.

From Fig. 3 and Fig. 4, it is found that several adjusts of target buffer are necessary in the classical DPBM method while only one adjustment was made by using the proposed method. The comparison of inventory on hand between the two methods is shown in Fig. 5.

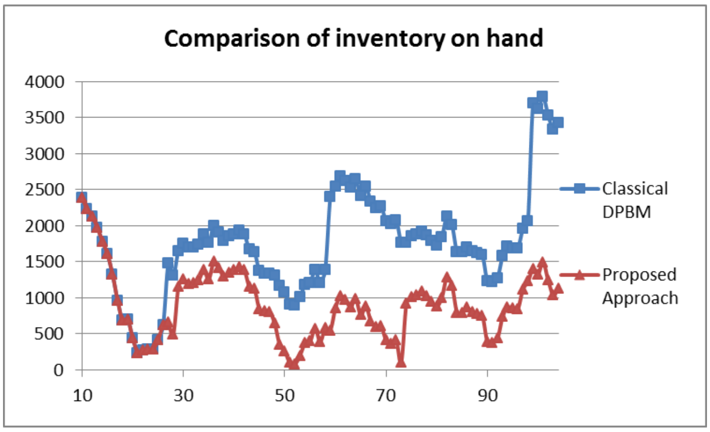

Fig. 5. Comparison of inventory on hand between the classical DPBM and the proposed approach

We can see that, from Fig. 5, the proposed approach provided 100% service by lower level of inventory than what the classical DPBM did. Table 2 lists the performance indicators resulted from usng these two approaches. From Table 2, it is found that the average inventory generated by the proposed method is about 48% less than the classical DPBM did.

ISBN: 978-605-64453-0-9

Table 2. Performance indicators comparison between the classical DPBM and the proposed approach.

	The proposed approach	Classical DPBM	% of Improvement made on classical DPBM by the proposed approach
Average Inventory(*AI*)	910.23	1750.80	48.01%
Service Level (*SL*)	100.00%	100.00%	0%

4. Conclusion

This research combines EWMA (Exponentially Weighted Moving Average) with rolling forecast information with respect to the market demand and EWMA of the past demand to improve the effectiveness of the classical DPBM approach when being applied to manage the inventory of semi-conductor products. The decisions on the timing and the amount of replenishment are based on the inventory level as well as the trend shown. Using a real case as an example, this paper demonstrated that the proposed approach can hold a lower inventory level and at the same time provide a higher service level in contrast to the classical DPBM approach for a product with fluctuating demand. More tests are required to validate the proposed approach on other types of demand pattern.

5. Acknowledgement

This research was financially supported by the National Science Council of Taiwan under Grant NSC NSC 101-2410-H-009-005-MY2.

References

[1] E. M. Goldratt, and J. Cox, *The Goal: Excellence in Manufacturing*, 1984.

[2] C. Y. Huang, C. P. Chen, R. K. Li *et al.*, “Applying Theory of Constraint on logistic management in large scale construction sites - A case study of steel bar in TFT-LCD factory build-up,” *Asian Journal on Quality*, vol. 9, no. 1, pp. 68-93, 2008.

[3] H. H. Wu, C. P. Chen, C. H. Tsai *et al.*, “A study of an enhanced simulation model for TOC supply chain replenishment system under capacity constraint,” *Expert Systems with Applications*, vol. 37, no. 9, pp. 6435-6440, 2010.

[4] K. W. Zheng, C. H. Tsai, R. K. Li *et al.*, “The Development of The Distribution/VMI Game Based on Theory of Constraints,” *Asian Journal on Quality*, vol. 10, no. 1, pp. 53-76, 2009.

[5] Y.-C. Chang, K.-H. Chang, and Y.-C. Lei, “Probe of the replenishment strategy and grouping rule in the semiconductor industry. ,” *Journal of Testing and Evaluation*, vol. in press, 2013.

[6] K.-T. Hung, E. D. Liou, C.-P. Wen *et al.*, "Decision support system for inventory management by TOC demand-pull approach." pp. 23-26, 2010.

[7] Y.-C. Chang, K.-H. Chang, and C.-W. Huang, “Integrate market demand forecast and demand-pull replenishment to improve the inventory management effectiveness of wafer fabrication. ,” *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. in press, 2013.

[8] Y.-C. Chang, and M.-C. L. Lee, “Buffer adjustment mechanism in TOC demand-pull inventory replenishment approach by EWMA,” in 2012 Journal of Arts and Sciences' (IJAS) international conference for academic disciplines, Munich, Germany, 2012.

[9] F. Chen, J. K. Ryan, and D. Simchi-Levi, “The impact of exponential smoothing forecasts on the bullwhip effect,” *Naval Research Logistics*, vol. 47, no. 4, pp. 269-286, Jun, 2000.