Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

A DEEP REINFORCEMENT LEARNING APPROACH FOR OPTIMAL REPLENISHMENT POLICY IN A VENDOR MANAGED INVENTORY SETTING FOR SEMICONDUCTORS

Muhammad Tariq Afridi
Santiago Nieto-Isaza

Technical University of Munich
Arcisstraße 21
Munich, 80333, GERMANY

Hans Ehm
Thomas Ponsignon
Abdelgafar Hamed

Infineon Technologies AG
Am Campeon 1 – 15
Neubiberg, 85579, GERMANY

ABSTRACT

Vendor Managed Inventory (VMI) is a mainstream supply chain collaboration model. Measurement approaches defining minimum and maximum inventory levels for avoiding product shortages and over-stocking are rampant. No approach undertakes the responsibility aspect concerning inventory level status, especially in semiconductor industry which is confronted with short product life cycles, long process times, and volatile demand patterns. In this work, a root-cause enabling VMI performance measurement approach to assign responsibilities for poor performance is undertaken. Additionally, a solution methodology based on reinforcement learning is proposed for determining optimal replenishment policy in a VMI setting. Using a simulation model, different demand scenarios are generated based on real data from Infineon Technologies AG and compared on the basis of key performance indicators. Results obtained by the proposed method show improved performance than the current replenishment decisions of the company.

1 INTRODUCTION

Due to intense competition and speedy innovation of technology, firms today are facing extremely volatile markets with short product life cycles (Aytac 2013). Dependence of modern technology products on semiconductor components and the exponential increase in the number of transistors used in a dense integrated circuit, requires efficiency in the semiconductor industry. Mastering end-to-end supply chains (SC) in semiconductor industry is therefore inevitable for achieving a competitive edge in a globalized economy (Ehm et al. 2011). However the semiconductor SC management, architecture and optimization face unique challenges due to specific attributes like short innovation cycles, long production lead times, and high demand uncertainties. Considering Infineon Technology AG as an example where frontend and backend production lead times can span a period of six months and even more, small changes in demands from end customers can result in significant fluctuations while moving further down the SC (Ehm and Ponsignon 2012). Due to the upstream position in the SC, semiconductor manufacturers are exposed to these amplifications of demand fluctuations, commonly known as the bullwhip effect, to a much greater extent. Lee et al. (1997) describes the bullwhip effect as a phenomenon where variance of orders tends to be larger than the variance of sales, and the distortion propagates in an amplified manner as one moves upstream in a SC.

While the bullwhip effect on the one hand results in “boom or bust” production cycles, it also necessitates a greater need for maintaining excessive safety stock if stock outs have to be avoided (Mackelprang and Malhotra 2015). In situations where demand depends on many exogenous variables, one idea would be to set safety stock levels (Beutel and Minner 2012). However in the semiconductor industry safety stock levels needed are very high and also comes with a high risk of scrap. Thus the fluctuations down the SC can result in tremendous inefficiencies like erroneous capacity plans, missed production

schedules, excessive inventory investment, and inventory stock outs. Various initiatives are adopted by companies for tackling the bullwhip, which are categorized on the basis of the underlying coordination mechanism, namely, information sharing, operational efficiency, and channel alignment. One particular initiative-known as Vendor Managed Inventory (VMI) which comes under the ambit of channel alignment, is the focus of our study (Lee et al. 1997).

VMI is a collaboration strategy adopted by various industries for managing complex SC, in which the supplier takes over the full responsibility over customer inventory replenishment and its related decisions (Kamalapur et al. 2013). Every VMI system has predefined conditions which commonly includes minimum and maximum inventory levels (Simchi-Levi and Kaminsky 2008). Due to the long production lead times in the semiconductor industry, it is a common approach that the customers sends demand forecasts on a rolling horizon basis to the supplier. VMI allows the supplier to independently satisfy the demand, provided that demands from the customer are within a certain range. Effective VMI implementation has produced benefits for both customers and suppliers, such as improved supplier service levels and production plans optimization, capacity utilization rates and reduction in transportation costs (Marquès et al. 2010).

In order to enable the collaborating partners to track and optimize their performance, seamless evaluation is required. In the semiconductor industry, measurement approaches are developed and applied by customers for evaluating performance of their suppliers (Continental AG 2010). Moreover, suppliers serving multiple customers, measure their own performance by making use of their individual approach (Ehm et al. 2018). Given that the replenishment decisions are transferred to the supplier, there is a common view that the responsibility for the performance of the VMI system lies entirely on the supplier (Odette International 2006). Nevertheless, due to long production lead time, the supplier may not necessarily be able to instantly adapt the replenishments in situations when the forecasts provided by the customer are inaccurate e.g. unforeseen pull of all available stocks. In such a situation, the underperformance shall be attributed to the behavior of the customer. This feature of shared responsibility for the success of VMI has been addressed by Ehm et al. (2018), and further extended by incorporating two exceptions in our simulation study. On top of this root-cause enabling VMI performance measurement simulation, a Deep Reinforcement Learning (DRL) approach is proposed for determining optimal replenishment quantities in a VMI setting. This allows for significant reduction in stock violations, resulting in fewer instances of responsibility assignment and ultimately better VMI performance. In Section 2 an overview of literature is provided. Next, the root-cause enabling VMI performance measurement case study is presented in Section 3. A theoretical basis of the DRL model is discussed in Section 4. Experiment and results are presented in Section 5. Finally, the conclusion and future research paths are discussed in Section 6.

2 LITERATURE REVIEW AND RESEARCH BACKGROUND

2.1 Supply Chain Collaboration and Typical Features of Vendor Managed Inventory

The application of SC collaboration models is an extensively debated and recognized solution to mitigate the bullwhip effect. Within SC collaboration models, independent companies uphold relationships marked by openness and trust where risks, rewards and costs are shared between parties (Li et al. 2010). Although the majority of publications mention how clear and accurate measurements can improve the collaboration with a higher level of trust, Ehm et al. (2018) applied measurements with split responsibility to VMI setting.

The parties entering a VMI contract have to agree on the details of some necessary features, namely, inventory location, the point of ownership transfer, and content of information sharing. Concerning the inventory location in a VMI setting, there is a consensus in literature on the customer's site (Disney and Towill 2003). This implies that replenishment decision is made by the supplier and the customer can directly pull its demand when needed. Hines et al. (2000) suggested that the inventory can also be held at a central warehouse, production line of the customer, or a third party logistic provider. Depending on the SC structure and product type, the parties entering a VMI agreement have to agree upon the location of inventory.

Next, the parties entering a VMI contract also need to settle on the point where the ownership is transferred from the supplier to the customer. In a standard arrangement, it could be assumed that the

ownership is deemed to have transferred once the product reaches the customer site. However, in essence, the concept of consignment inventory can also be observed, which indicates that the inventory at the customer's premises is owned and replenished by the supplier, and the ownership transfer only takes place when the product is actually consumed or pulled by the customer (Hieber & Schönsleben 2002).

The content of information that is expected to be shared among the collaborating parties however varies across various industrial publications and academic literature. In majority cases, sharing of point-of-sales data is suggested. This allows for the speedy transfer of sales data upstream, thereby allowing the suppliers to rapidly react to changes in the demands. Information appertaining to inventory level of the VMI stock is another commonly mentioned content in literature (Disney and Towill 2003). This aids the supplier in making appropriate decisions related to replenishments in order to keep the inventory within the predefined minimum and maximum levels. In order to foster efficient planning, a variety of other information can also be included. Angulo et al. (2004) in their paper stress upon the need of making customer forecast data as part of the VMI arrangement. For this to happen, information about upcoming promotions and new product introduction should be shared with the supplier (DeToni & Zamolo 2005). Sharing of production schedules is also one suggested component of information content. Vigtil (2007) suggests that by providing production schedules, valuable information on future stock withdrawals could be revealed, thus allowing suppliers to better plan their replenishments.

2.2 Inventory Replenishment in Vendor Managed Inventory Systems

Coelho and Laporte (2015) proposed optimized target level (OTL) inventory replenishment policy, under which the replenishment always keep the final inventory at the same customer-dependent optimized target level. They perform computational experiments to evaluate the OTL policy against maximum level (ML) and order-up-to (OU) policies. Their results show that OTL yields lower costs and inventory levels than the OU policy, and is slightly more expensive than the ML policy, while being easier to implement. Cetinkaya and Lee (2000) presented an analytical model for the simultaneous computation of a time-based consolidation policy and the optimal replenishment quantity in a VMI setting. Govindan (2015) proposed an Adjusted Silver–Meal heuristic for a time-varying stochastic demand in two-echelon SCs with an aim to minimize the total SC cost by comparing various performance measures between traditional and VMI systems. Escuin et al. (2017) aimed at developing a mathematical model for computing the optimal inventory composition in order to deal with random demand at minimum cost in a two-tier SC under capacity and service level constraints. They used a simulation model to minimize the inventory costs and improve the level of customer service by incorporating adequate production planning and appropriate replenishment schedule in VMI setting.

2.3 Reinforcement Learning for Optimal Inventory Replenishment in Vendor Managed Inventory Systems

Gijsbrechts et al. (2019) applied DRL for solving classical intractable dual sourcing inventory replenishment problem. They compared their results to well established heuristics and approximate dynamic programming methods and found that with extensive tuning of hyper-parameters, matching performance could be achieved. Oroojlooyjadid et al. (2017) proposed a reinforcement learning (RL) algorithm based on deep Q-networks for optimizing replenishment decisions at a given stage in a multi-agent, decentralized, cooperative SC problem. They showed that near-optimal solutions could be achieved when base-stock policy is followed by the agents. Also, it outperforms the base-stock policy when the other agents utilize a more realistic ordering behavior model. Sui et al. (2010) proposed an approach based on RL for determining optimal replenishment policy in a VMI system with consignment inventory. They compared their results to the newsvendor solution and showed that their approach outperformed the newsvendor model. No literature was found on simulating and finding optimal replenishment policy on top of a measurement including split responsibility to the concept of VMI.

3 CASE STUDY: ROOT-CAUSE ENABLING VENDOR MANAGED INVENTORY PERFORMANCE MEASUREMENT

3.1 Generic Vendor Managed Inventory Measurement Approach

It is important to define clear responsibility assignment apropos of VMI application and whenever the defined Min/Max inventory limits are violated. Consequently, a metric is outlined and further developed to monitor stock violations and assign responsibilities. It is expected that such a metric fosters collaboration, eventually resulting in mitigating the bullwhip effect.

The process for developing a metric like that begins with the analysis of the underlying VMI configuration as shown in Figure 1. The collaboration begins with the customer providing the demand forecasts to the supplier. Considering the current stock information, the supplier plans and deliver replenishments, which may be pulled by the customer from the stock at any point in time. It is pertinent to mention that the supplier does not receive any information with regard to generation of the demand forecasts. Therefore, there could be instances when the pull from the customer rises abruptly without being forecasted, resulting in a stock-out situation. As per the current setup, the supplier will be held responsible for a failed delivery. This call for the use of root-cause enabling VMI performance measurement approach, which could adequately assign responsibilities for any kind of stock violation.

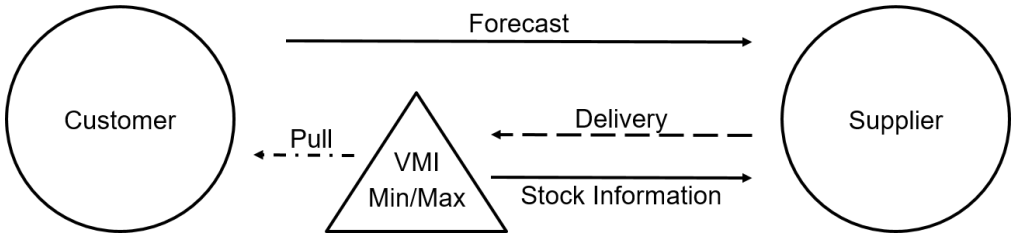

Figure 1: Typical VMI configuration.

The VMI performance measurement approach used for responsibility assignment comprises of four main steps as shown in Figure 2. Foremost, the overall weekly performance WP of the VMI system is calculated using formula (1) adopted from Odette International (2006), and compared to target weekly performance $WP_{Target} = 75\%$. Here, NV, OS, US, and SO represents no-violation, over-stock, under-stock, and stock-out, respectively, whereas $Weight_V$ are the weights associated to the different inventory states reflecting the severity of the respective stock violation type.

$$WP_w = \frac{Days_{NV,w} \times Weight_{NV}}{\sum_{V=\{NV,OS,US,SO\}} Days_{V,w} \times Weight_V} \times 100 \quad (1)$$

Being in a certain inventory state requires the calculation of minimum z and maximum Z target stocks levels. These are calculated using formula (2):

$$z_w = c \times \frac{\sum_{i=p}^{q} FC_{w,i}}{q-p+1} \quad \text{and} \quad Z_w = b \times \frac{\sum_{i=p}^{q} FC_{w,i}}{q-p+1} \quad (2)$$

where c, b, p, and q are the parameters values decided by the VMI partners. While the values of c and b defines the range of no-violation NV inventory state, the value of q is positively correlated to the production and delivery time.

If WP is found to be less than WP_{Target}, then further steps are performed in order to establish the responsibility for not achieving the target weekly performance. This requires first the calculation of forecast

accuracy FA for that particular week, which is compared to the target forecast accuracy FA_{Target}. If the FA is found to be greater than or equal to the FA_{Target}, the responsibility for not achieving WP_{Target} value is assigned to the supplier. The calculation of FA is performed using Symmetric Mean Absolute Percentage Error (SMAPE) technique (Ott et al. 2013) as given in formula (3). Here, D_w is the demand and AFC_w is the average forecast for that particular week. The calculation of AFC is done using formula (4). Here, u and z are the parameters values decided by the VMI partners, and depends on the production and delivery strategy of the supplier.

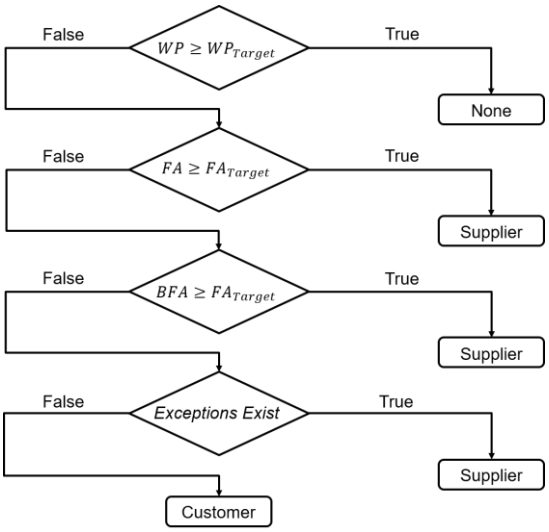

Figure 2: Scheme for Root-cause Enabling VMI Performance Measurement.

$$FA_w = \left(1 - \frac{|AFC_w - D_w|}{AFC_w + D_w}\right) \times 100 \tag{3}$$

$$AFC_w = \frac{\sum_{i=w-z}^{w-u} FC_{i,w}}{z-u+1} \times 100 \tag{4}$$

If the FA is found to be less than the FA_{Target}, the responsibility for not achieving WP_{Target} value is not directly assigned to the customer, and further investigation into the forecast history is performed to give leverage to the customer for unbiased forecasts. Since bias in forecasts can lead to stock violations, consideration of bias in the forecast and calculation of bias-adjusted forecast accuracy is required. Therefore, forecast bias FB is first calculated using formula (5) below as proposed by Trigg (1964):

$$FB_w = \frac{\sum_{i=w-j}^{w-1} AFC_i - D_i}{\sum_{i=w-j}^{w-1} |AFC_i - D_i|} \tag{5}$$

The forecast bias considered for the evaluation of week *w* is taking into account a certain historical period *j* decided mutually by the VMI partners. Formula (5) results in values between -1 and +1. If for all considered weeks the forecasts are higher than the demand (*over-forecasting*) the value of forecast bias would be $FB = +1$. Alternatively, if all considered weeks the forecasts are lower than the demand (*under-forecasting*) the value of forecast bias would be $FB = -1$. Entire unbiased forecasts would generate a value of $FB = 0$. It should be noted that this measure does not provide any information about the extent of the forecast error. Hereupon the forecast bias is incorporated into the forecast accuracy measurement calculate the bias-adjusted forecast accuracy as shown in formula (6) below (Trigg 1964):

$$BFA_w = FA_w + BF \times (1 - |FB_w|) \times (1 - FA_w) \quad (6)$$

where $BF = [0; 1]$ is the bias factor value which is used to steer how much it is desired to consider the forecast bias, and therefore reward the customer. No general suggestion for the bias factor value was found in literature. However, it depends on the agreed minimum stock, production specifications, and the supplier's capacity and inventory flexibility. Finally the BFA is compared to the FA_{Target}. The supplier is held responsible if BFA is found to be greater than or equal to FA_{Target}.

3.2 Extension of the Existing Algorithm

Two exceptions are checked in the final step of the performance measurement algorithm when the BFA is found to be less than FA_{Target}: (1) Over-stock despite *under-forecasting* ($AFC_w < D_w$) and (2) Under-stock (or stock-out) despite *over-forecasting* ($AFC_w > D_w$). The reasons for such exceptions to occur could either be due to the inefficiencies in the delivery strategy of the supplier, or the delays in supplier's production system. Both exceptions, if occur, could potentially attribute the poor weekly performance to the customer when the FA is found to be less than the FA_{Target}. For either of the exception to occur, the supplier is held responsible for the poor performance in that respective week.

It should be noted that for exception 1 to hold true, the weekly sum of over-stock OS shall be greater than the weekly sum of under-stock US and stock-out SO, when $AFC_w < D_w$. Also for exception 2 to hold true, the weekly sum of under-stock US and stock-out SO shall be greater than weekly sum of over-stock OS, when $AFC_w > D_w$.

4 DEEP REINFORCEMENT LEARNING FOR SELECTING OPTIMAL REPLENISHMENT QUANTITIES

4.1 Reinforcement Learning and Markov Decision Process

RL, also known as approximate dynamic programming, is an area of machine learning that has been effectively applied in recent years for solving complex sequential decision problems. While classical dynamic programming (DP) became ineffective in solving large-scale Markov decision problems (MDP) due to the curse of modeling and dimensionality, strong mathematical roots of RL in the principles of function approximation and DP allowed for solving such problems (Gosavi 2009). Curse of dimensionality refers to the dramatic increase in time and space required for finding an approximate solution to a MDP, when its state space and control variables become intractably large (Gosavi 2009). RL deals with the question on what action an agent must take to maximize (minimize) the cumulative reward (penalty).

Figure 3 shows a MDP in which an agent interacts with its environment. The agent observes system's current state $s_t \in \mathbb{S}$ ($\mathbb{S}$ being the set of all possible states) in a given time t, and takes an action $a_t \in \mathbb{A}(s_t)$ ($\mathbb{A}(s_t)$ being the set of all possible actions given that the system is in state s_t) to receive a reward $r_t \in \mathbb{R}$. Afterwards, the system randomly makes a transition into state $s_{t+1} \in S$ (Sutton and Barto 1998). In order to find solution to such a problem, RL can be utilized.

The probability for transitioning from state s to a state s' upon taking an action a is provided by the transition probability matrix $P_a(s, s')$ as shown in formula (7). Also, the reward matrix is defined by $R_a(s, s')$. Every period t, the agent takes and action $a_t = \pi_t(s)$ as per a given policy π_t. The objective of the RL is to determine a policy $\pi: \mathbb{S} \rightarrow \mathbb{A}$ which maximize the expected discounted sum of the rewards r_t when the system operate for an infinite time horizon, as given in formula (8). Here $a_t = \pi_t(s_t)$, and $0 \leq \gamma < 1$ is the discount factor.

$$P_a(s, s') = P_r(s_{t+1} = s' \mid s_t = s, a_t = a) \quad (7)$$

$$Max \sum_{t=0}^{\infty} \gamma^t E[R_{at}(s_t, s_{t+1})] \quad (8)$$

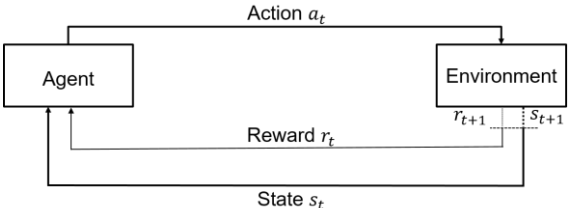

Figure 3:Generic Reinforcement Learning Process.

According to Sutton and Barto (1998), for a given $P_a(s,\ s')$ and $R_a(s,\ s')$, the optimal policy can be calculated using linear or dynamic programming. Another method to solve this problem is the Q-learning approach, which captures the Q-value for any $a = \pi(s)$ and $s \in S$, i.e., $Q(s,a) = \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \gamma^3 r_{t+3} + \cdots \mid s_t = s, a_t = a; \pi\ ]$. This begins with a preliminary assumption for $Q(s,a)\ \forall\ s,a$, and then advances to update the values on the basis of an iteration as shown in formula (9) below:

$$Q(s_t,a_t) = (1-\alpha_t)Q(s_t,a_t) + \alpha_t\left(r_{t+1} + \gamma \max_a Q(s_{t+1},a_t)\right), \forall t = 1,2,3,\ldots, \quad (9)$$

where α_t is the learning rate in a given time t. The agent decides on taking an action in each observed state via an ϵ-greedy algorithm. This implies that a random action is chosen with a probability ϵ_t in a given time t, and an action with the highest cumulative action value ($a_{t+1} = arg\max_a Q(s_{t+1},a_t)$) is chosen with a probability $1-\epsilon_t$. This enables the algorithm to explore the solution space and provide an assurance of optimality if $\epsilon_t \to 0$, when $t \to \infty$. Once the optimal Q^* is found, the optimal policy could be retrieved in the form of $\pi^*(s) = arg\max_a Q^*(s,a)$.

While both dynamic programming and Q-learning algorithms provide assurance of optimality, they are confronted with the curse of dimensionality while solving MDPs having large state and action spaces. To address this, Mnih et al. (2015) developed a deep Q-Network (DQN) that combines RL with artificial neural networks known as deep neural networks, in order to achieve an approximation of the Q-function, and which is trained via the Q-learning algorithm iterations while updating another target network (Oroojlooyjadid et al. 2019). The work carried out here is based on this approach.

4.2 Simulation and Deep Reinforcement Learning for Training Replenishment Policy

The performance measurement AnyLogic model is further extended with reward function, state (observation) space, and action space, in order to prepare it as a training environment in an external integrated development environment (IDE) called IntelliJ IDEA. The model is exported as a Java standalone application and imported into IntelliJ. A RL for Java (RL4J) library is utilized in order to make the agent learn a policy. The trained model is imported back into AnyLogic model as a testbed, where the extended model is used as an environment in order to teach the learning agent on taking appropriate actions in order to achieve a desired state. Below we briefly describe the action space, state space and reward function:

- **Action**: A discrete space is used to define actions. We define them as having 9 possible values from $0 \to 8$. Considering m as a magnitude of an action, the replenishment policy is a function of action a given by $f(a) = m \times a$. For example if $a = 4$ and $m = 2500$, the value of action would be 7500, i.e., a replenishment amount of 7500 units in our case ($a = 1$ corresponds to 0).

- **State**: We have one unique state (observation) which is related to the anticipated stock position based on the predictions of the current week, the actual demand, and the anticipated minimum z and maximum Z target stock levels. These are calculated by using same formulas as in formula (2), but slightly modified to incorporate order lead time OLT, as under:

$$z_wF = c \times \frac{\sum_{i=p+OLT}^{q+OLT} FC_{w,i}}{q-p+1} \quad \text{and} \quad Z_wF = b \times \frac{\sum_{i=p+OLT}^{q+OLT} FC_{w,i}}{q-p+1} \tag{10}$$

The anticipated stock position FSP is given by the formula (11), where CSP is the current stock position, FR is the anticipated replenishment based on the forecast at the time of decision, FC is the forecast, and D is the actual demand. Given an OLT we know the amount replenished in future. This value (FR) is calculated on a daily basis, which means that whenever there is a new replenishment, it is summed up for a given OLT. It can be observed from formula (11) that the demand is comprised of two elements: D is the actual demand, whereas FC is the anticipated demand. This is equal to the sum of all forecasts that could exist during the total time span of the OLT. Afterwards, the mean value of anticipated minimum z_wF and maximum Z_wF target stock levels is calculated using $MF = (z_wF + Z_wF)/2$. Since we are interested in knowing how much the FSP deviates from the MF (as ideally we would want our inventory to stay within the z_wF and Z_wF), the anticipated distance to the mean is calculated using $DTMF = FSP - MF$. A negative value of $DTMF$ would correspond to the FSP below the MF, and vice-versa. Finally the state (observation) is a normalized value between -1 and $+1$ as shown in formula (12).

$$FSP = CSP + \sum_{i=1}^{OLT} FR - \left(\sum_{i=1}^{OLT} FC + D\right) \tag{11}$$

$$State = \begin{cases} +1, for\ FSP > 2 \times Z_wF \\ \frac{DTMF}{(2\times Z_wF)-MF}, for\ \ MF \le FSP \le 2 \times Z_wF \\ \frac{DTMF}{MF}, for\ \ 0 \le FSP \le MF \\ -1, for\ \ FSP < 0 \end{cases} \tag{12}$$

- **Reward Function**: The reward function is different from the state function is a sense that it uses the current stock position CSP instead of anticipated stock position FSP, and naturally so, since we are interested in the immediate reward. First the mean value of minimum z and maximum Z target stock levels is calculated using $M = (z_w + Z_w)/2$, which is followed by calculating distance to the mean using $DTM = CSP - M$. It is pertinent to mentioned that we are assigning penalties if the current stock position CSP deviates from mean M. Finally the reward (penalty) is normalized value between -1 and $+1$ as shown in formula (13).

$$Reward = \begin{cases} 1 - \frac{2\times DTM}{(2\times Z_w)-M}, for\ \ M \le CSP \le 2 \times Z_w \\ 1 + \left(\frac{2\times DTM}{M}\right), for\ \ 0 \le CSP \le M \\ -1, \qquad elsewhere \end{cases} \tag{13}$$

4.3 Data Preparation and Performance Evaluation

Three demand scenarios are used to evaluate performance of our approach. The first scenario concerns the actual data (853 days). For this purpose demand forecasts, actual demand data, and replenishment data is used for a randomly selected product type and customer in a VMI partnership. Two other random demand scenarios are generated, representative of the demand (and forecasts) in scenario one, to be used as a training set and tested on the real data for the purpose of further validating our approach. The second scenario (random demand and forecast) is generated using a comprehensive method known as Martingale Method of Forecast Evolution (MMFE). For this, readers are advised to follow the work carried out by Heath and Jackson (1994). Finally the third scenario (random demand and forecast with sporadic rise and fall) is generated by slightly modifying the second scenario through random introduction of random multipliers.

5 EXPERIMENT AND RESULTS

5.1 Discrete Event Simulation based Validation

The VMI performance measurement approach is validated in AnyLogic discrete event simulation environment. The sensitivity of the developed approach is tested using different parameters which includes forecast information, daily replenishments, and actual demand (pull) of 853 days. This provides the basis to calculate the minimum z target stock level, maximum Z target stock level, and daily status of the stock level, as shown in Figure 4(a). The weekly performance WP is calculated afterwards which is compared to the target weekly performance WP_{Target}. In case of poor performance, the assignment of responsibility is carried out with the calculation of forecast accuracy FA, bias-adjusted forecast accuracy BFA, and the exceptions check. The simulation provides a clear insight into the inventory states and the assigned responsibilities as shown in Figure 4(b). Values of the different parameters mutually decided by the VMI partners are provided in Table 1.

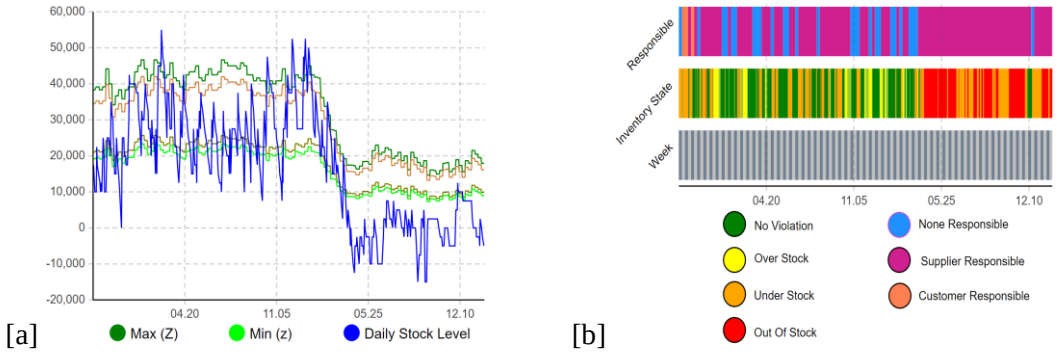

Figure 4: [a] Max Z, Min z, and Daily Stock Level and, [b] Responsibility and Inventory States.

Table 1: Parameters used in VMI Performance Measurement.

Parameter	c	b	p	q	u	z	j	WP_{Target}	BF	FA_{Target}
Value	2	4	1	12	1	12	12	75%	0.5	90%

We focus on three Key Performance Indicators (KPIs): (1) Alpha (α) service level, which is the ratio of days without shortages to the total number of days. (2) Beta (β) service level, which is the fraction of demand immediately satisfied from stock. (3) Percentage no-violation, which is the ratio of number of days with no-violation inventory states to the total number of days. In addition to these, we also make a comparison of the total number of shipments done for the different demand scenarios and order lead times.

5.2 Demand Scenario 1

For demand scenario 1, the initial 70% of data is used for training purpose, and the testing is performed on the last 30%. ϵ-greedy is used as the selection method for learning. Double DQN is used, and the Stochastic Gradient Method used for updating the parameter is Adam. Three hidden layers with hundred neurons each are used. A learning rate $\alpha = 0.00001$, Q target frequency of 10000, experience replay size 50000, batch size 256, and gamma , $\gamma = 1$ is used. Figure 5(a) visualize the daily stock level with $OLT = 0$ after the trained policy is used as a testbed. It should be noted that the testing is performed on the last 30% of the data which is highlighted in the red box. Figure 5(b) shows the corresponding results with $OLT = 28$ days.

5.3 Demand Scenario 2 and 3

Random demand and demand forecast are generated based on MMFE. This method provides realistic representation of how demand forecasts evolve over time and is useful in our problem. Also, since the training is performed on the randomly generated data, testing the trained policy on the complete set of actual data (853 days) provides better insights. Note that for scenario 1, the initial inventory state is given, having 17500 units. For scenario 2 and 3, a random initial inventory is taken within a range of -30000 and 60000 units. Also the analysis for $OLT = 28$ is performed considering the time needed for backend production. Figure 5(c) and Figure 5(d) visualize the daily stock level for scenario 3, using $OLT = 0$ and 28. Figures for scenario 2 are not included due to space limitation, however the results are presented in Table 2.

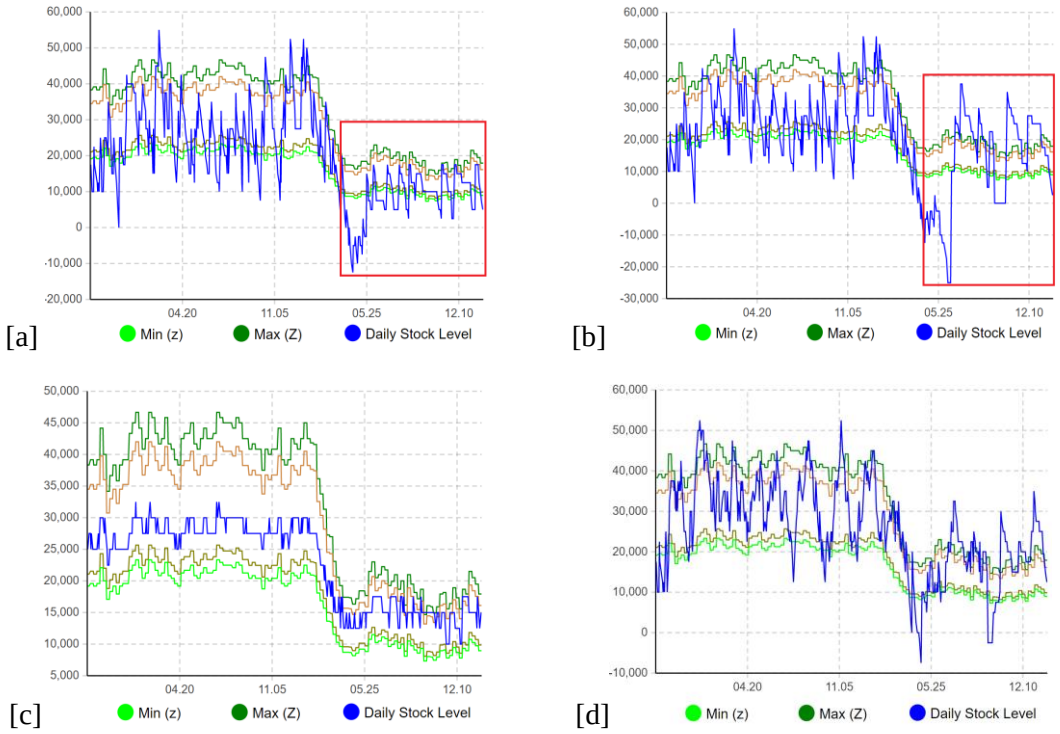

Figure 5: Daily Stock Level using RL [a] Scenario 1 using OLT=0; [b] Scenario 1 using OLT=28; [c] Scenario 3 using OLT=0; and [d] Scenario 3 using OLT=28.

It can be observed, that for $OLT = 0$, the replenishment policy is such that the daily stock level stays within the the no-violation state and hardly have any stock violations occuring. For $OLT = 28$, there are substantially fewer instances of stock violations even when compared to the daily stock levels of Figure 4(a), which is a simulated inventory without using RL. Table 2 provides an overview of the achieved performance and compare the KPIs for all three scenarios (both for $OLT = 0$ and $OLT = 28$), to the KPIs when no policy is used. From the results presented in Table 2, it can be seen that while significant improvements have been made in %-age no-violations, this comes at the cost of increased shipments. Naturally so, the RL algorithm is set with an objective to improve the inventory, and there exist no constraint on the number of shipments. Having said that, the number of shipments could be reduced by consolidating consecutive shipments that are suggested by the RL algorithm, while maintaining the same service levels and inventory performance.

Table 2: Comparison of KPIs.

	No Policy		* Scenario 1		** Scenario 2		** Scenario 3	
	*	**	*OLT* 0	*OLT* 28	*OLT* 0	*OLT* 28	*OLT* 0	*OLT* 28
% α Service Level	63.14%	85.13%	100%	88.58%	100%	96.72%	100%	98.01%
% β Service Level	45.90%	88.45%	100%	82.25%	100%	97.54%	100%	99.02%
% No-Violation	5%	43%	66%	21%	95%	62%	99%	63%
Total Shipments	20	80	10	9	227	159	246	253

** Using Last* 30% *Data ** Using Complete Data*

6 CONCLUSION AND NEXT STEPS

The term VMI is commonly considered as a strategy in which the replenishment decision is shifted to the supplier. In this paper, the root-cause enabling VMI performance measurement approach was extended to measure the responsibility for poor performance by taking account of the forecast accuracy for the demand mutually agreed between the collaborating partners. The approach was tested and validated via simulation on a set of company data. Considering room for reduction in stock violations from a supplier's perspective, optimization in the replenishment policy was studied and implemented using DRL algorithm in a simulation environment. Result show that the percentage no-violation inventory status improved from 43% to 95% and 99% for both scenarios respectively, while maintaining higher α and β service levels. This comes with increased transportation costs, could be reduced if consecutive shipments are consolidated outside the RL. The challenge ahead would be to investigate additional demand scenarios and conduct a full design of experiment taking into account the volatility in demand and demand levels. It would also be very useful to extend the scope to multi-stage SC for bullwhip analysis.

REFERENCES

Angulo, A., H. Nachtmann, and M. A.Waller. 2004. "Supply Chain Information Sharing in a Vendor Managed Inventory Partnership". *Journal of Business Logistics* 25(1):101-120.

Aytac, B., and S. D. Wu. 2013. "Characterization of Demand for Short Life-Cycle Technology Products". *Annals of Operations Research* 203:255-277.

Beutel, A. L., and S. Minner. 2012. "Safety Stock Planning under Causal Demand Forecasting". *International Journal of Production Economics* 140(2):637-645.

Cetinkaya, S., and C-Y. Lee. 2000. "Stock Replenishment and Shipment Scheduling for Vendor Managed Inventory Systems". *Management Science* 46(2):217-232.

Coelho, L. C., and G. Laporte. 2015. "An Optimized Target-Level Inventory Replenishment Policy for Vendor-Managed Inventory Systems". *International Journal of Production Research* 53(12):3651-3660.

Continental AG. 2010. Global Logistics Standards and Processes of Continental Automotive: Supplier Manual Logistics. https://www.continental-automotive.com/en-gl/Organization/Company/Supplier-Information, accessed 3rd August 2020.

De Toni, A., and E. Zamolo. 2005. "From A Traditional Replenishment System to Vendor-Managed Inventory: A Case Study From the Household Electrical Appliances Sector". *International Journal of Production Economics* 96(1):63-79.

Disney, S., and D. R. Towill. 2003. "Vendor-Managed Inventory and Bullwhip Reduction in a Two-Level Supply Chain". *International Journal of Operations & Production Management* 23(6):625-651.

Ehm, H., F. Jankowiak, V. Filser, T. Lauer, and A. Nguyen. 2018. "A Generic VMI Measurement and Application in the Semiconductor Industry". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 3449–3460. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ehm, H. and T. Ponsignon. 2012. "Future Research Directions for Mastering End-to-End Semiconductor Supply Chains". In *Proceedings of the 2012 International Conference on Automation Science and Engineering*, 641–645. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ehm, H., T. Ponsignon, and T. Kaufmann. 2011. "The Global Supply Chain is Our New Fab: Integration and Automation Challenges". In *Proceedings of the 2011 SEMI Advanced Semiconductor Manufacturing Conference*, 1–6. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Escuín, D., L. Polo, D. Ciprés. 2017. "On the Comparison of Inventory Replenishment Policies With Time-Varying Stochastic Demand for the Paper Industry". *Journal of Computational and Applied Mathematics* 309:424-434.

Forrester, J. W. 1958. "Industrial Dynamics: A Major Breakthrough for Decision Makers". *Harvard Business Review* 36:37-66.

Gijsbrechts, J., R. Boute, D. Zhang, and J. Van Mieghem. 2019. "Can Deep Reinforcement Learning Improve Inventory Management? Performance on Dual Sourcing, Lost Sales and Multi-Echelon Problems". *SSRN Electronic Journal*.

Gosavi, A.. 2009 "Reinforcement Learning: A Tutorial Survey and Recent Advances," *INFORMS Journal on Computing*, 21(2):178-192.

Govindan, K.. 2015. "The Optimal Replenishment Policy for Time-Varying Stochastic Demand Under Vendor Managed Inventory". *European Journal of Operational Research* 242(2):402-423.

Heath, D. C., and P. L. Jackson. 1994. "Modeling the Evolution of Demand Forecasts with Application to Safety Stock Analysis in Production/Distribution Systems". *IIE Transactions* 26(3):17-30.

Kamalapur, R., D. Lyth, and A. Houshyar. 2013. "Benefits of CPFR and VMI Collaboration Strategies: A Simulation Study". *Journal of Operations and Supply Chain Management* 6(2):59-73.

Lee, H. L., V. Padmanabhan, and S. Whang. 1997. "The Bullwhip Effect in Supply Chains". *MIT Sloan Management Review* 38(3):93-102.

Mackelprang, A. W., and M. K. Malhotra. 2015. "The Impact of Bullwhip on Supply Chains: Performance Pathways, Control Mechanisms, and Managerial Levers". *Journal of Operations Management* 36:15-32.

Marquès, G., J. Lamothe, C. Thierry, and D. Gourc. 2010. "Vendor Managed Inventory, from Concepts to Processes, for an Unified View". *Production Planning and Control* 21(6):547-561.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. 2015. "Human-level Control through Deep Reinforcement Learning". *Nature*, 518(7540):529-533.

Odette International. 2006. Key Performance Indicators (KPI) for Global Materials Management and Logistics: https://www.odette.org/publications/file/key-performance-indicators-kpi-for-global-materials-management-and-logistic, accessed 13th April 2020.

Oroojlooyjadid, A., M. R. Nazari, L. Snyder, and M. Takáč. 2017. "A Deep Q-Network for the Beer Game: A Reinforcement Learning algorithm to Solve Inventory Optimization Problems". *arXiv preprint arXiv:1708.05924*.

Ott, H. C., S. Heilmayer, and C. Sng. 2013. "Granularity Dependency of Forecast Accuracy in Semiconductor Industry". *Research in Logistics & Production* 3(1):49-58.

Schönsleben, P., and R. Hieber. 2002."Gestaltung von effizienten Wertschöpfungspartnerschaften im Supply Chain Management". In *Integriertes Supply Chain Management*, edited by A. Busch and W. Dangelmaier, 47-64. Wiesbaden: Gabler Verlag.

Simchi-Levi, D. and P. Kaminsky, and E. Simchi-Levi. 2008. *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies.* 3rd ed. Boston: McGraw-Hill Irwin.

Sui, Z., A. Gosavi, and L. Lin. 2010. "A Reinforcement Learning Approach for Inventory Replenishment in Vendor-Managed Inventory Systems With Consignment Inventory". *Engineering Management Journal* 22(4):44-53.

Sutton, R. S., and A. G. Barto. 1998. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.

Trigg, D. W. 1964 "Monitoring a Forecasting System" *Journal of the Operational Research Society* 15(3):271-274

Vigtil, A. 2007. "Information Exchange in Vendor Managed Inventory". *International Journal of Physical Distribution & Logistics Management* 37(2):131-147.

AUTHOR BIOGRAPHIES

MUHAMMAD TARIQ AFRIDI is a Master student in Management at the Technical University of Munich. His major is Logistics and SC Management. He holds a Masters in C-Tech from University of Ulm, Germany. His email address is tariq.afridi@tum.de.

SANTIAGO NIETO-ISAZA is a reasearch associate and doctoral student at the Technical University of Munich.. His email address is santiago.nieto-isaza@tum.de.

HANS EHM is Lead Principal heading the Supply Chain Innovation department at Infineon Technologies. His email address is hans.ehm@infineon.com.

THOMAS PONSIGNON is a Senior Staff Engineer in the Supply Chain Innovation department at Infineon Technologies. His email address is thomas.ponsignon@infineon.com.

ABDELGAFAR ISMAIL MOHAMMED HAMED is a Supply Chain Expert in the Supply Chain Innovation department at Infineon Technologies. His email address is abdelgafar.ismail@infineon.com.