Journal of Industrial and Management Optimization

JIMO, 22(10): 5308–5350.
DOI: 10.3934/jimo.2026183
Received: 04 May 2026
Revised: 09 September 2026
Accepted: 11 September 2026
Published: 22 September 2026

https://www.aimspress.com/journal/jimo

Research article

Rolling-horizon supply chain capacity planning for IC design house considering forecast evolution and capacity regret

James C. Chen[1,*], Tzu-Li Chen[1], Yin-Yann Chen[2], Chen-Yu Wang[1] and Dewanti Anggrahini[1,3]

[1] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 30013, Taiwan
[2] Department of Industrial Management, National Formosa University, Yunlin 632301, Taiwan
[3] Department of Industrial and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

* **Correspondence:** Email: james@ie.nthu.edu.tw.

Abstract: Demand uncertainty and long capacity lead times make supply chain capacity planning particularly challenging for integrated circuit (IC) design houses that manage multi-stage and multi-site outsourcing networks. In this study, we developed a long-term rolling-horizon supply chain capacity planning (RHSCCP) framework to support production planning decisions under evolving demand forecasts. For the analysis, we used simulation-generated scenarios based on additive and multiplicative martingale models of forecast evolution (MMFE) processes, incorporating different demand scenarios, covariance levels, and capacity regret strategies. Four RHSCCP models were developed in this study: minimax regret (RHSCCP-MMR), upper bound and lower bound (RHSCCP-ULB), expected value (RHSCCP-EV), and stochastic programming (RHSCCP-SP). Their performance was evaluated in a rolling-horizon simulation environment using total cost, capacity utilization, and order fulfillment rate. The results showed distinct performance characteristics across the four strategies. RHSCCP-ULB adopted a more conservative capacity-planning policy, resulting in relatively high fulfillment rates but lower capacity utilization as demand variability increased. In contrast, RHSCCP-EV maintains relatively stable capacity utilization but shows lower fulfillment performance. RHSCCP-MMR and RHSCCP-SP provide a better balance between fulfillment and capacity utilization. Under additive MMFE, particularly at higher demand fluctuations, RHSCCP-MMR is the preferred strategy because of its stronger fulfillment performance and robustness to

demand uncertainty. Under multiplicative MMFE, RHSCCP-SP is preferred because it achieves minimum total cost while maintaining relatively high fulfillment and capacity utilization rates.

Keywords: additive and multiplicative martingale model of forecast evolution; rolling-horizon planning, supply chain capacity planning; capacity regret; IC design house
Mathematics Subject Classification: 90B06, 37M05, 90C15

1. Introduction

In the competitive and capital-intensive electronics market, demand for integrated circuits (ICs) is highly dynamic and difficult to predict. Therefore, demand and capacity management are critical in semiconductor supply chain planning. Fabless IC design houses mainly focus on IC research and development and downstream sales to customers, while outsourcing wafer manufacturing to specialized foundries, increasing the significance and complexity of supply chain capacity planning (SCCP).

Due to the costly investments in outsourcers' capacity equipment, IC design houses must set capacity levels based on long-term wafer production demand forecasts. Given the long production cycle time, the IC design house employs a hybrid production planning mode that combines make-to-stock (MTS) and make-to-order (MTO) strategies. Wafer fabrication (wafer fab) production follows MTS, while back-end processes follow MTO. The wafer bank maintains wafer-type inventory for MTS, while the finished goods bank holds die-type inventory to satisfy actual orders within the MTO strategy. Figure 1 shows the process flow of the IC design house supply chain planning (SCP).

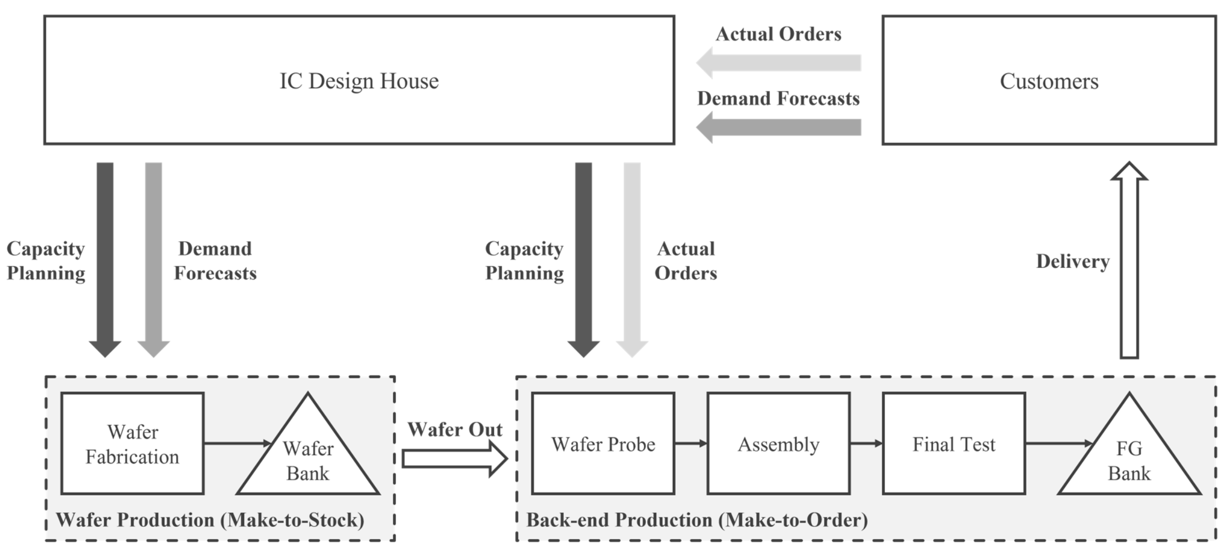


Figure 1. The process flow of the IC design house SCP.

Due to the capacity reservation mechanism, IC design houses must secure capacity with outsourcers at least six months in advance, driven by long-term demand forecasts before receiving actual orders. Since capacity levels cannot be revised once set, accurate planning is notably important. After completing capacity planning, the IC design house initiates wafer fab based on short-term demand information, and then follows with back-end production upon order release. The constrained capacity levels for both phases are derived from long-term demand forecasts. Figure 2 delineates the phases of each planning period leading up to the delivery date of the actual order.

Generally, wafer fabrication has a lead time of eight to twelve weeks, while back-end production takes four to six weeks. The main objective of capacity planning is to maximize the profit from capacity investment while maintaining or improving the demand fulfillment rate. Short-term SCP aims to accurately allocate demand to outsourcers based on preserved capacity levels set during the capacity planning period. Because the preserved capacity level becomes the capacity constraint and can lead to surpluses or shortages, accurate demand information and capacity planning are very critical in the SCCP model.

As market trends evolve rapidly, demand forecasts change over time. The researchers in [1] defined demand forecast error as the portion of the forecast not converted to actual orders and described forecast bias in the semiconductor industry. The researcher in [2] emphasized demand uncertainty as a significant risk factor impacting supply chain performance. This research models realistic demand uncertainty in SCCP using forecast evolution within a rolling-horizon planning framework. Figure 3 describes the operational concept of forecast evolution in rolling-horizon planning.

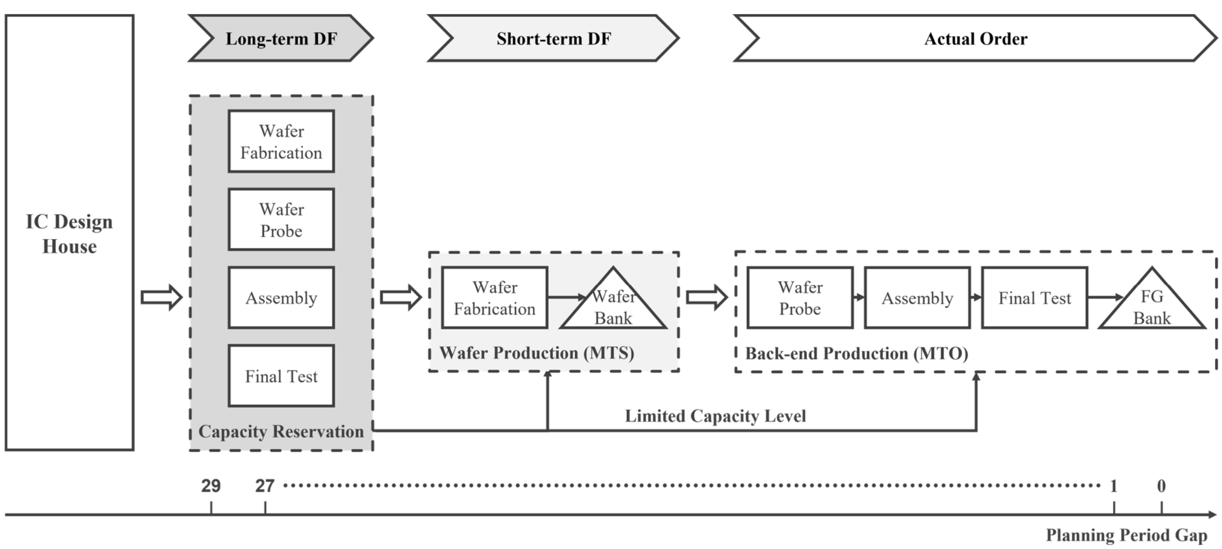


Figure 2. The phases of each planning period.

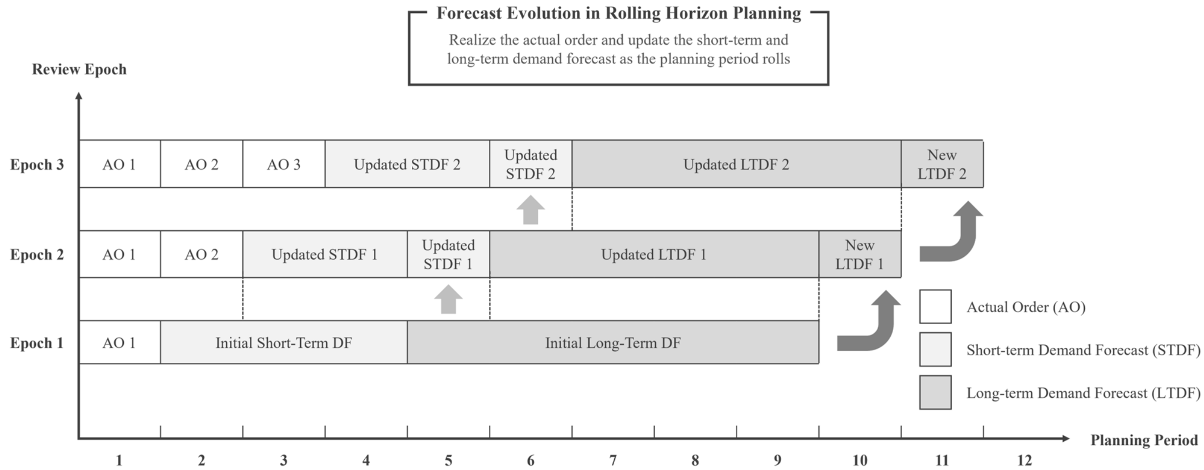


Figure 3. Forecast evolution in rolling-horizon planning.

As time rolls on, customers update their demand forecasts based on market trends at each planning epoch. The IC design house has a 13-week frozen period before the delivery due date. Demand during this period cannot be revised and is considered actual orders. These orders remain fixed, while the other remaining forecasts are updated. To represent the dynamic nature of demand information, the researchers in [3] developed the martingale model of forecast evolution (MMFE) for dynamic production environments, while the researchers in [4] subsequently applied MMFE within a rolling-horizon planning framework. In semiconductor supply chain capacity planning (SCCP), the researchers in [5] addressed capacity decisions using a two-stage stochastic programming model.

Despite these advances, researchers have largely examined demand uncertainty, forecast evolution, and capacity planning as separate issues. Although MMFE has been used to represent dynamically evolving forecasts and stochastic programming has been applied to semiconductor capacity planning, limited attention has been given to integrating forecast evolution directly into rolling-horizon SCCP decisions. In particular, the comparative incorporation of additive and multiplicative MMFE under multiple demand scenarios within a regret-based capacity planning framework remains underexplored. This gap motivates this study. To address this gap, we integrate additive and multiplicative MMFE models within a rolling-horizon SCCP. We employ the minimax capacity regret strategy extended in [6], considering lead time, transportation time, and yield rate for each site and machine, as shown in Figure 4.

In this study, we emphasize long-term capacity planning, aligning capacity reservation planning with short-term production planning that would be realized after six months, and omit inventory considerations from wafer fab to the finished goods bank. Our objective is to minimize the maximum capacity regret under several demand scenarios. The contributions of this study are as follows:

(1) Propose SCCP model for the IC design house that accounts for capacity reservation rules and network selection flexibility, including multi-stage, multi-site, and technical constraints in the final test stage.
(2) Incorporate the concept of capacity regret into the SCCP model to determine the preserved capacity level in each stage, site, and machine to outsourcers, enabling robust capacity planning.

(3) Apply additive and multiplicative MMFE models to generate several demand scenarios in a rolling-horizon planning framework to simulate the demand uncertainty and its evolution over time.
(4) Conduct the simulation experiments with four SCCP models, two types of MMFE models, and three covariance levels as factors.
(5) Evaluate performance across factor combinations to provide the numerical experiment results and a robust capacity planning strategy for the IC design house.

The remainder of this paper consists of Sections 2–6: In Section 2, we review studies related to capacity planning, supply chain planning, rolling-horizon planning, and forecast evolution models. In Section 3, we define the SCCP problem and present the mathematical model. In Section 4, we discuss the simulation setup, the rolling-horizon framework, and the experimental results. In Section 5, we explain the managerial implications, while in Section 6, we conclude with insights and recommendations for future research.

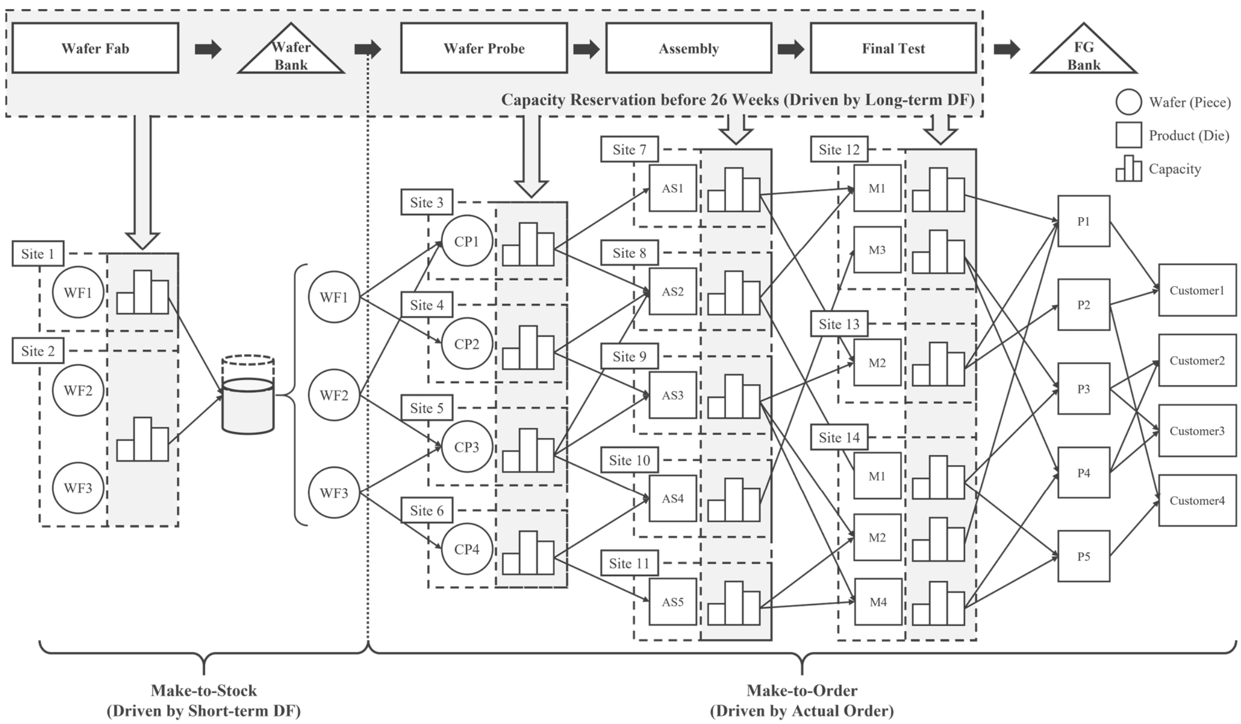


Figure 4. The multi-stage and multi-site network of SCCP.

2. Literature review

Effective decision-making regarding capacity investment is crucial in the semiconductor industry due to the high cost of equipment and technical constraints. The researcher in [7] developed a stochastic model that incorporates demand uncertainty to solve the single-period tool procurement problem for a semiconductor wafer fab, comprising two stages: determining the required number of tools for procurement and allocating production quantity for each tool. Then, the researcher in [8] modified the model to a multi-period stochastic model. A scenario-based stochastic programming model for long-term capacity planning in the semiconductor industry, considering multiple stages, was demonstrated in [9]. Building on this research, the researchers in [10] introduced a two-stage stochastic model that accounts for demand uncertainty and technical constraints for different types of wafers and tools, as well as the capacity changes arising from tool uncertainty downtime, or preventive maintenance.

The capacity planning challenge in the thin-film transistor liquid crystal display (TFT-LCD) industry is similar to the semiconductor industry, with both networks characterized by multi-stage, multi-site, and tool constraints. The researchers in [11] developed a stochastic dynamic model for the TFT-LCD industry, introducing a multi-site capacity expansion policy with demand generated from a one-step demand state transition probability and updated in each period. The researchers in [12] considered the workforce and work shift, formulating a nonlinear mixed-integer model to determine the capacity expansion strategy. The researchers in [6] proposed a two-stage capacity planning model with a capacity regret strategy that aims to minimize the possible maximum regret across scenarios in a time-rolling environment. This model generates fluctuating demand across several scenario sets. Furthermore, this research identified three types of capacity decisions: Capacity expansion, which may cause a capacity surplus; capacity contraction, which may result in a capacity shortage; and doing nothing to maintain the current capacity status, which may incur two types of regret. Based on this research, The researchers in [13] extended the model and incorporated the concept of capacity regret into the MMFE model in the rolling-horizon environment, aiming to support robust capacity planning for wafer fabs without accounting for technical constraints.

Given that the semiconductor industry is capital-intensive and highly competitive, effective resource allocation is important. The semiconductor supply chain involves numerous process stages and sites, making it complex and the subject of several studies. In response to the increasing diversity of downstream customer demands, enterprises expand production capacity by adding new facilities or outsourcing to meet the demand, which, in turn, shifts manufacturing environments from traditional single-site to multi-site. This evolution transforms the supply chain into a multi-stage and multi-site network characterized by sequential and complementary features, similar to what is observed in the semiconductor industry [14]. The researchers in [15] developed a mixed-integer programming (MIP) model to aid resource allocation for long-term planning in the semiconductor industry, including strategic decisions on whether to add new facilities, equipment sets, or outsourcing. The researchers in [5] extended the model and presented a multi-period, two-stage stochastic integer-programming model with recourse to analyze the effects of product correlation on strategic capacity decisions across demand scenarios.

The researchers in [16] proposed a MIP model for IBM to minimize total production cost, incorporating supporting heuristics for the semiconductor supply chain to address complexities in lot sizing, material flow, and capacity level constraints. The researchers in [17] considered the order fulfillment rate and limited capacity level when determining optimal wafer production quantities for several products, facilities, and periods. They proposed heuristic approaches to address master planning problems in semiconductor manufacturing networks. Subsequently, the researchers in [18] extended the previous model by introducing a simulation-based framework that models market demand behavior and the production system, accounting for demand and execution uncertainty throughout the rolling horizon. However, because they focused only on wafer fab planning, they did not incorporate the flexibility of supply chain network selection or material flow between stages. The researchers in [19] provided a deterministic MIP model to meet the forecast demand over a six-month planning horizon, considering the capacity constraints, facility qualification criteria, and inventory requirements to schedule production quantities. The researchers in [20] presented an inventory management strategy for the semiconductor supply chain by introducing a die bank. The researchers in [21] developed a hybrid MTS-MTO supply chain network for an IC design house with the multi-objective order promising problem. In subsequent research, the researchers in [22] proposed a MIP model to address the problem, using a product-based decomposition algorithm with profit-based criteria and a Lagrangian decomposition algorithm with the product-based decomposition method. While these studies demonstrate the extensive use of MIP and decomposition-based approaches in semiconductor planning, recent research has broadened the methodological landscape by incorporating machine learning, metaheuristics, and robust or scenario-based optimization to address more complex decision-making problems under uncertainty.

The researcher in [23] combined machine learning and metaheuristics for robust project scheduling, while the researchers in [24] integrated machine learning with optimization techniques for healthcare operations. The researcher in [25] also developed a hybrid artificial immune system (AIS) and artificial fish swarm (AFS) algorithms for optimizing renewable energy project portfolio selection. In supply chain applications, the researchers in [26] proposed a robust agricultural supply chain network design under uncertainty, while the researchers in [27] applied scenario-based robust optimization to a sustainable biodiesel supply chain. The researchers in [28] further highlighted the growing need for uncertainty-aware optimization in supply chain network design. Despite these advances, limited attention has been given to planning problems involving sequential forecast updates and repeated capacity adjustments within a rolling-horizon framework, particularly in the semiconductor supply chain.

Building on these uncertainty-aware optimization approaches, the literature also provides broader mathematical perspectives for modeling dynamic and complex decision environments. Robust optimization and fuzzy goal programming have been extensively applied to configure sustainable supply chain networks under demand and environmental uncertainties [29,30]. Moreover, continuous and dynamic modeling techniques, originally developed for complex regulatory networks or optimal control under uncertainty, provide profound insights into dynamic system behavior [31]. These mathematical structures and dynamic modeling approaches have proven highly applicable across fields, particularly in the optimal control of unsteady dynamic systems [32]. By exploring these mathematical perspectives, we adopt a rolling-horizon framework to continuously update stochastic demand, providing a highly dynamic modeling approach for semiconductor capacity planning.

In the rolling-horizon environment, demand is continuously revised as new information becomes available. The researchers in [33] pointed out that there is a lack of research on rolling-horizon supply chain planning systems. The researchers in [3,4] developed the MMFE model to simulate forecast evolution in a dynamic production environment, enabling demand updates at each epoch, accommodating the evolving nature of information. The researchers in [4] incorporated the mean of the conditional demand into their analysis and integrated the MMFE model into the safety stock management policy to estimate the optimal safety stock level within a rolling-horizon planning framework. The researcher in [34] considered a stochastic multi-item Capacitated Lot-Sizing Problem

(CLSP) using a scenario tree for simulating demand uncertainty and proposed a plant-location-based model and heuristic solution in a rolling-horizon planning environment.

The researchers in [35] extended the MMFE model by incorporating conditional and unconditional covariance and applied it to aid inventory management. The experimental results demonstrated that the optimal inventory policy depends on the conditional covariance. Based on the previous model, the researchers in [36] extended it to a chance-constrained model to address the preserved capacity allocation problem. The researchers used actual semiconductor industry data, along with the additive MMFE model, in rolling-horizon planning to evaluate performance. Subsequently, the researchers in [37] extended the chance-constrained multiplicative MMFE model to solve the time-varying demand distribution problem in the rolling-horizon. The researchers in [38] studied the methodologies for demand-driven and cycle-time-driven safety stock policies under both additive and multiplicative MMFE models. Derived from the additive MMFE model presented by [4,39], discussed the qualification management problem in the wafer fab. The researchers in [40] investigated additive and multiplicative MMFE models and applied stochastic lot-sizing in a rolling-horizon environment. The researchers in [13] integrated additive and multiplicative MMFE models with the concept of capacity regret from [6]. They presented a semiconductor production model for a wafer fab operating in a rolling manner.

Researchers have extensively advanced supply chain and resource planning research through robust optimization, stochastic modeling, machine learning-optimization integration, and hybrid metaheuristics under uncertainty. However, these researchers mainly focus on project scheduling, resource allocation, portfolio selection, and strategic supply chain network design, with limited attention to how demand forecasts evolve sequentially and how planning decisions should be repeatedly adjusted within a rolling-horizon framework.

This gap is particularly relevant to the semiconductor industry, where capacity commitments are often made well before actual demand is realized and must be revised as forecast information evolves. Moreover, limited research has entailed long-term capacity planning for IC design houses by continuously considering rolling-horizon decision making, capacity reservation mechanisms, multi-stage and multi-site networks, technical constraints, and forecast evolution. Therefore, a more integrated framework is required to evaluate how alternative capacity-planning strategies respond to evolving demand uncertainty in this setting.

3. Problem definition

3.1. Problem statement and assumption

The rolling-horizon supply chain capacity planning (RHSCCP) network for the IC design house comprises the wafer fab, wafer bank, wafer probe, assembly, final test, and finished goods bank. In the rolling-horizon planning, the demand forecasts undergo continuous updates at each epoch r. Based on long-term demand forecasts at each epoch from customers, the IC design house makes reservations of the capacity level 26 weeks forward of the delivery date to the outsourcers in the planning network. In the RHSCCP model, determining reserved capacity levels for multiple outsourcers involves a trade-off between capacity regrets under different demand scenarios. Capacity regret is defined as the gap between the demand forecast and the reserved capacity level. It can be classified into two types: capacity surplus regret and capacity shortage regret. The first type, capacity surplus regret, occurs when the reserved capacity level exceeds the demand forecast under a given scenario. This condition indicates that part of the reserved outsourcing capacity remains unused, leading to capacity waste and unnecessary outsourcing capacity reservation cost. In contrast, capacity shortage regret occurs when the demand forecast exceeds the reserved capacity level. This condition indicates that the reserved capacity is insufficient to fulfill the forecasted demand and may result in unfulfilled demand or shortage-related costs. Figure 5 shows the concept of capacity regret.

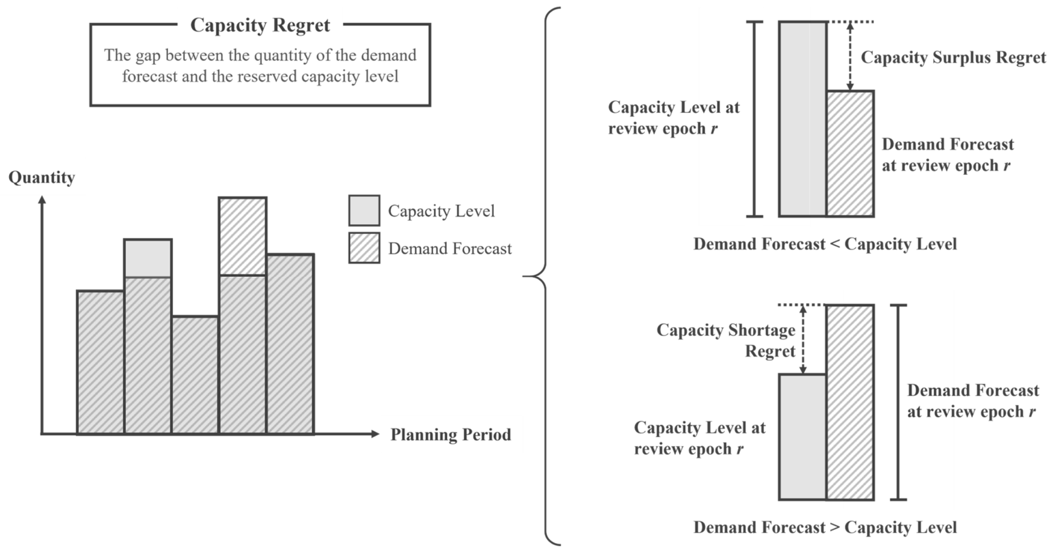


Figure 5. The concept of capacity regret.

To simulate a rolling-horizon environment, we use the forecast evolution method. We establish T planning period within each review epoch r as the total planning horizon, which consists of actual orders within Δ planning periods referred to as the frozen periods. The demand forecasts within δ planning periods for the wafer fab production under MTS mode, and the demand forecasts within $T-\Delta$ planning periods are referred to as the flexible periods. For long-term capacity planning, capacity levels are planned from $r+T-\Delta-\delta$ to $r+T-1$, corresponding to the short-term production planning anticipated to be realized in the future, where the capacity reservation period T is the demand forecast that will be translated into the actual order after 26 review epochs.

We assume that a set of p products is in the planning horizon, and defines $df_{p,t}^{r}$ as the demand forecast of product p required in period t at review epoch r, $ao_{p,t}^{r}$ as the actual order of product p required in period t at review epoch r. The initialization of the first planning horizon demand vector is $[ao_{p,1}^{1}, \dots, ao_{p,\Delta}^{1}, df_{p,\Delta+1}^{1}, \dots, df_{p,T}^{1}]$. During subsequent review epochs, the planning horizon rolls by one planning period in each iteration, and the demand forecast vector $df^{r+1} = [df_{p,r+\Delta}^{r+1}, df_{p,\Delta+1}^{r+1}, \dots, df_{p,r+T-2}^{r+1}, df_{p,r+T-1}^{r+1}]$ is updated from the last demand forecast vector $df^{r} = [df_{p,r+\Delta}^{r}, df_{p,\Delta+1}^{r}, \dots, df_{p,r+T-2}^{r}, df_{p,r+T-1}^{r}]$ by the forecast update vector $\varepsilon^{r+1} = [\varepsilon_{p,r+\Delta}^{r+1}, \varepsilon_{p,\Delta+1}^{r+1}, \dots, \varepsilon_{p,r+T-2}^{r+1}, \varepsilon_{p,r+T-1}^{r+1}]$ at the beginning of the review epoch $r+1$. The update process of the actual is the same as the demand forecast. At the beginning of the review epoch $r+1$, the actual order $ao_{p,r+\Delta}^{r+1}$ is realized from the demand forecast $df_{p,r+\Delta}^{r}$ by the forecast update vector $\varepsilon_{p,r+\Delta}^{r+1}$. Once the demand is realized as an actual order, it will no longer be revised in subsequent review epochs. For instance, a set of p products in the rolling-horizon planning with Δ frozen periods and $T-\Delta$ flexible periods, where the capacity levels before $r+\Delta+25$ planning period are decided at the previous review epochs, and the capacity levels after the planning period $r+\Delta+25$ will be determined based on the corresponding demand forecasts. Figure 6 illustrates the forecast evolution process.

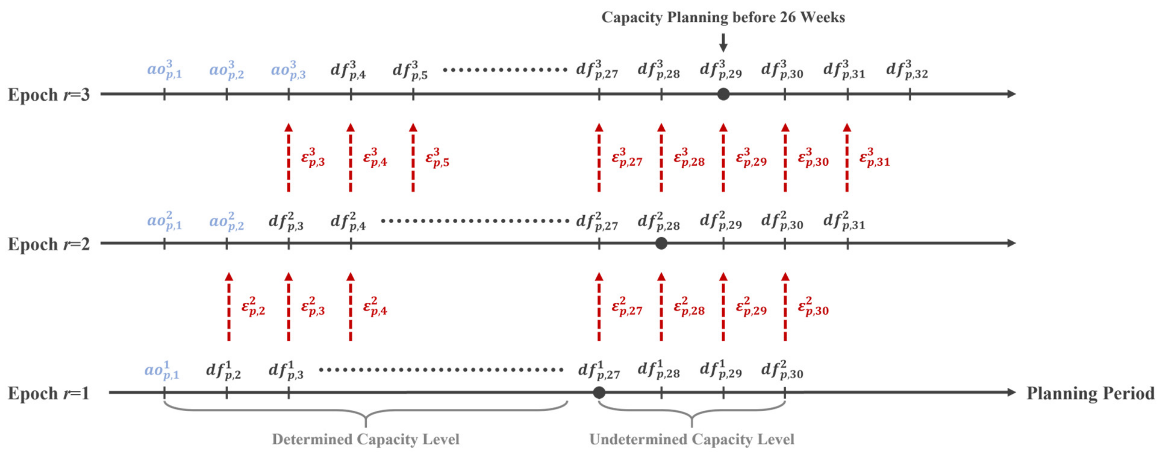


Figure 6. The illustration of the forecast evolution process.

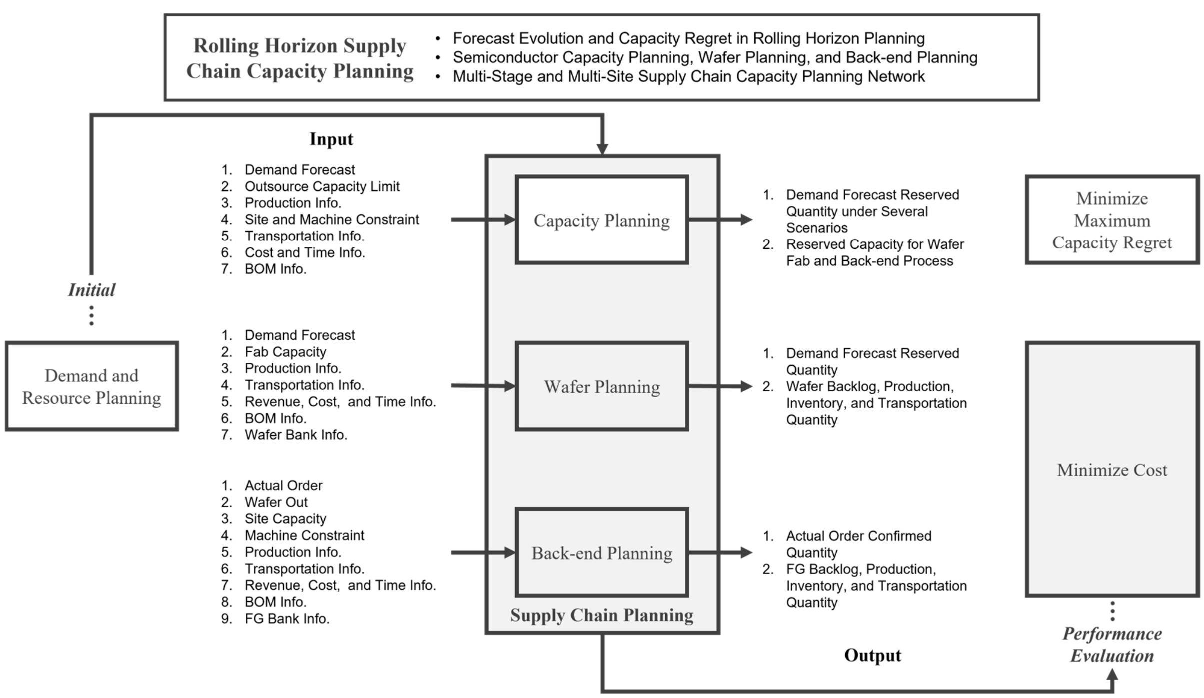


Figure 7. The role and structure of the RHSCCP model in the planning horizon.

In this research, we develop the RHSCCP model to simulate the decision-making process for capacity level and the dynamic evolution of demand, generated by the additive and multiplicative MMFE models. The RHSCCP model considers production, transportation, and regret costs to determine the optimal capacity level under a trade-off across several demand scenarios. Furthermroe, this model accommodates multi-stage, multi-site, and technical constraints. The primary objective of this model is to minimize overall cost in the planning horizon. Figure 7 shows the role and structure of the RHSCCP model within the planning horizon, with a focus on capacity planning.

The model assumptions are as follows:

(1) We focus on long-term capacity planning, considering the costs of fixed, production, transportation, capacity surplus, and shortage, but do not account for setup costs. All costs are fixed in the model.

(2) Upon completion of production processes in the current stage, WIP items are ready for immediate transportation to the next stage; therefore, we do not account for inventory at each stage or site.

(3) Each wafer can be produced into multiple final products, enabling the use of alternative wafer types as substitutes to fulfill orders.

(4) Turnkey mode is not considered, implying that each assembly and final test facility is treated as a one-function station.

(5) The final test stage is subject to capacity limitations and machine technical constraints; each product can undergo testing using only particular types of machines that meet technical constraints.

(6) In long-term capacity planning, we perform the wafer fab capacity planning from $r+T-\Delta-\delta$ to $r+T-\Delta$ and back-end capacity planning from $r+T-\Delta$ to $r+T-1$, which correspond to the short-term production planning that will be realized after half a year.

(7) Each order comprises only one product type with a specific due date. In other words, if the customer necessitates multiple product types or distinct due dates, it should be separated into multiple orders.

3.2. Minimax regret model formulation

In this study, we develop a mathematical model with an objective function that minimizes total costs. The formulation is as follows:

$$\textbf{Minimize} \quad Z = \theta \tag{1}$$

Subject to:

$$\theta \geq C^{s}_{production} + C^{s}_{transportation} + C^{s}_{regret}\,, \ \ \forall s \in S \tag{2}$$

$$C^{s}_{production} = \sum_{l\in L_F}\sum_{w}\sum_{p}(pc_{lwp} \times \sum_{t=r+T-\Delta-\delta}^{r+T-\Delta} Q^{s1}_{lwpt})$$
$$+ \sum_{l\in L_O}\sum_{w}\sum_{p}(pc_{lwp} \times \sum_{t=r+T-\Delta}^{r+T-1} Q^{s1}_{lwpt})$$
$$+ \sum_{l\in L_A}\sum_{p}(pc_{lp} \times \sum_{w}\sum_{t=r+T-\Delta}^{r+T-1} Q^{s1}_{lwpt} \times gd_{wp}) \quad (3)$$
$$+ \sum_{l\in L_T}\sum_{m}\sum_{p}(pc_{lmp} \times \sum_{t=r+T-\Delta}^{r+T-1} Q^{s1}_{lmpt}) + \sum_{l\in L_F}\sum_{p}(fc_{lp} \times \sum_{t=r+T-\Delta-\delta}^{r+T-\Delta} X^{s}_{lpt})$$
$$+ \sum_{l\in L_O, L_A, L_T}\sum_{p}(fc_{lp} \times \sum_{t=r+T-\Delta}^{r+T-1} X^{s}_{lpt}) \ , \ \forall s \in S$$

$$C^{s}_{transportation} = \sum_{l\in L_F}\sum_{l'\in B_W}(tc_{ll'} \times \sum_{w}\sum_{p}\sum_{t=r+T-\Delta-\delta}^{r+T-\Delta} T^{s1}_{ll'wpt})$$
$$+ \sum_{l\in B_W}\sum_{l'\in L_O}(tc_{ll'} \times \sum_{w}\sum_{p}\sum_{t=r+T-\Delta}^{r+T-1} T^{s1}_{ll'wpt})$$
$$+ \sum_{l\in L_O}\sum_{l'\in L_A}(tc_{ll'} \times \sum_{w}\sum_{p}\sum_{t=r+T-\Delta}^{r+T-1} T^{s1}_{ll'wpt}) \quad (4)$$
$$+ \sum_{l\in L_A}\sum_{l'\in L_T}(tc_{ll'} \times \sum_{p}\sum_{t=r+T-\Delta}^{r+T-1} T^{s1}_{ll'pt})$$
$$+ \sum_{l\in L_T}\sum_{l'\in B_P}(tc_{ll'} \times \sum_{p}\sum_{t=r+T-\Delta}^{r+T-1} T^{s1}_{ll'pt}) \ , \ \forall s \in S$$

$$C^{s}_{regret} = \sum_{l\in L_F}\sum_{t=r+T-\Delta-\delta}^{r+T-\Delta}(OR^{s}_{lt} + SR^{s}_{lt}) + \sum_{l\in L_O, L_A}\sum_{t=r+T-\Delta}^{r+T-1}(OR^{s}_{lt} + SR^{s}_{lt})$$
$$+ \sum_{l\in L_T}\sum_{m}\sum_{t=r+T-\Delta}^{r+T-1}(OR^{s}_{lmt} + SR^{s}_{lmt}) + \sum_{p}(OR^{s}_{p(t=r+T-\Delta)} + SR^{s}_{p(t=r+T-\Delta)}) \quad (5)$$
$$+ \sum_{p}(OR^{s}_{p(t=r+T-1)} + SR^{s}_{p(t=r+T-1)}) \ , \ \forall s \in S$$

The objective function of this model is to minimize total costs across several demand scenarios. It is reflected by equation (1). Equation (2) describes the total cost, containing production costs, transportation costs, and regret costs. Equation (3) is the total production costs incurred from the wafer fab process to the back-end process. Equation (4) reflects the total transportation costs from the wafer fab to the finished goods bank, while equation (5) defines the total regret costs. The total regret cost is divided into two parts: capacity regret and demand regret. The following subsections detail the technical constraints.

3.2.1. Production limit

Equations (6) to (9) constrain the production quantities given to the maximum reservable capacity levels. Equations (6) and (7) ensure that the consumed capacity for wafer production does not exceed the upper bound of the maximum reservable capacity of the respective site. In addition, equations (8) and (9) restrict the consumed capacity at each site and by each machine within the site to remain within the upper bound of the corresponding machine's maximum reservable capacity. The respective constraints are:

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw} \leq UBC_l\,,\ \ \forall s \in S, l \in L_F,\ t = r + T - \Delta - \delta,\ \ \ldots, r + T - \Delta \tag{6}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw} \leq UBC_l\,,\ \ \forall s \in S, l \in L_O, t = r + T - \Delta, \ldots, r + T - 1 \tag{7}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lwp} \leq UBC_l\,,\ \ \forall s \in S, l \in L_A,\ t = r + T - \Delta, \ldots, r + T - 1 \tag{8}$$

$$\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lmpt'}^{s1} \times cc_{lmp} \leq UBC_{lm}\,,\ \ \forall s \in S, l \in L_T, m \in M, t = r + T - \Delta, \ldots, r + T - 1 \tag{9}$$

3.2.2. Production action

Equations (10) to (12) represent the production action constraints that identify whether a product is produced at each site in each time period. These constraints ensure that the binary variable takes a value of 1 when a production action is required for a given site, product, and time. Otherwise, when no production action is required, the binary variable is set to 0. The production action restrictions are:

$$\sum_{w} Q_{lwpt}^{s1} \leq UBP \times X_{lpt}^{s}\,,\ \ \forall s \in S, l \in L_F,\ p \in P, t = r + T - \Delta - \delta,\ \ \ldots, r + T - \Delta \tag{10}$$

$$\sum_{w} Q_{lwpt}^{s1} \leq UBP \times X_{lpt}^{s}\ ,\ \ \forall s \in S, l \in L_O, L_A, p \in P, t = r + T - \Delta,\ \ \ldots,\ r + T - 1 \tag{11}$$

$$\sum_{m} Q_{lmpt}^{s1} \leq UBP \times X_{lpt}^{s}\ ,\ \ \forall s \in S, l \in L_T,\ p \in P, t = r + T - \Delta,\ \ \ldots,\ r + T - 1 \tag{12}$$

3.2.3. Capacity regret

Equations (13) to (28) reflect the restriction of the capacity regret. Equations (13) to (16) calculate the capacity oversupply regret costs, and equations (17) to (20) calculate the capacity shortage regret costs. Equations (21) to (28) serve as constraints to ensure capacity regret is either oversupply or shortage. The constraints of capacity regret are as follows:

$$OR_{lt}^{s} = (1 - Y_{lt}^{s}) \times oc_l \times (CE_{lt} - \sum_{w} \sum_{p} \sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw})\ , \quad \forall s \in S, l \in L_F,\ t = r + T - \Delta - \delta,\ \ \ldots, r + T - \Delta \tag{13}$$

$$OR_{lt}^{s} = (1 - Y_{lt}^{s}) \times oc_l \times (CE_{lt} - \sum_{w} \sum_{p} \sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw})\ ,\ \ \forall s \in S, l \in L_O, t = r + T - \Delta, \ldots, r + T - 1 \tag{14}$$

$$OR_{lt}^{s} = (1 - Y_{lt}^{s}) \times oc_l \times (CE_{lt} - \sum_{w} \sum_{p} \sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lwp})\ ,\ \ \forall s \in S, l \in L_A,\ t = r + T - \Delta, \ldots, r + T - 1 \tag{15}$$

$$OR_{lmt}^{s} = (1 - Y_{lmt}^{s}\ ) \times oc_{lm} \times (CE_{lmt} - \sum_{p} \sum_{t'=t-lt_l}^{t} Q_{lmpt'}^{s1} \times cc_{lmp})\,, \quad \forall s \in S, l \in L_T, m \in M, t = r + T - \Delta, \ldots, r + T - 1 \tag{16}$$

$$SR_{lt}^{s} = Y_{lt}^{s} \times sc_l \times (\sum_{w} \sum_{p} \sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw} - CE_{lt})\,,\ \ \forall s \in S, l \in L_F,\ t = r + T - \Delta - \delta,\ \ \ldots, r + T - \Delta \tag{17}$$

$$SR_{lt}^{s} = Y_{lt}^{s} \times sc_{l} \times (\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw} - CE_{lt}),\ \ \forall s \in S, l \in L_O, t = r+T-\Delta, \dots, r+T-1 \tag{18}$$

$$SR_{lt}^{s} = Y_{lt}^{s} \times sc_{l} \times (\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lwp} - CE_{lt}),\ \ \forall s \in S, l \in L_A,\ t = r+T-\Delta, \dots, r+T-1 \tag{19}$$

$$SR_{lmt}^{s} = Y_{lmt}^{s} \times sc_{lm} \times (\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lmpt'}^{s1} \times cc_{lmp} - CE_{lmt}),\ \ \forall s \in S, l \in L_T, m \in M, t = r+T-\Delta, \dots, r+T-1 \tag{20}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw} \geq CE_{lt} - UBC_{l} \times (1 - Y_{lt}^{s}),\ \ \forall s \in S, l \in L_F,\ t = r+T-\Delta-\delta,\ \dots, r+T-\Delta \tag{21}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw} \geq CE_{lt} - UBC_{l} \times (1 - Y_{lt}^{s}),\ \ \forall s \in S, l \in L_O, t = r+T-\Delta, \dots, r+T-1 \tag{22}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lwp} \geq CE_{lt} - UBC_{l} \times (1 - Y_{lt}^{s}),\ \ \forall s \in S, l \in L_A,\ t = r+T-\Delta, \dots, r+T-1 \tag{23}$$

$$\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lmpt'}^{s1} \times cc_{lmp} \geq CE_{lmt} - UBC_{lm} \times (1 - Y_{lmt}^{s}),\ \ \forall s \in S, l \in L_T, m \in M, t = r+T-\Delta, \dots, r+T-1 \tag{24}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q_{lwpt'}^{s1} \times cc_{lw} < CE_{lt} + UBC_{l} \times Y_{lt}^{s},\ \ \forall s \in S, l \in L_F,\ t = r+T-\Delta-\delta,\ \dots, r+T-\Delta \tag{25}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q^{s1}_{lwpt'} \times cc_{lw} < CE_{lt} + UBC_l \times Y^s_{lt},\ \ \forall s \in S, l \in L_O, t = r + T - \Delta, \dots, r + T - 1 \tag{26}$$

$$\sum_{w}\sum_{p}\sum_{t'=t-lt_l}^{t} Q^{s1}_{lwpt'} \times cc_{lwp} < CE_{lt} + UBC_l \times Y^s_{lt},\ \ \forall s \in S, l \in L_A,\ t = r + T - \Delta, \dots, r + T - 1 \tag{27}$$

$$\sum_{p}\sum_{t'=t-lt_l}^{t} Q^{s1}_{lmpt'} \times cc_{lmp} < CE_{lmt} + UBC_{lm} \times Y^s_{lmt},\ \ \forall s \in S, l \in L_T, m \in M, t = r + T - \Delta, \dots, r + T - 1 \tag{28}$$

3.2.4. Demand regret

Equations (29) to (36) represent the demand regret constraints. Equations (29) to (32) calculate the demand oversupply and shortage regret costs, respectively. Equations (33) to (36) judge the type of demand regret in the wafer bank and the finished goods bank. The demand regret restrictions are:

$$OR^s_{pt} = \left(1 - Y^s_{pt}\right) \times oc_p \times \left(\sum_{l\in B_W}\sum_{l'\in L_O}\sum_{w} T^{s1}_{ll'wpt} \times yro_{wp} \times yra_p \times yrt_p \times gd_{wp} - df^s_{p(t'=r+T-1)}\right),\ \ \forall s \in S, p \in P,\ t = r + T - \Delta \tag{29}$$

$$OR^s_{pt} = \left(1 - Y^s_{pt}\right) \times oc_p \times \Big(\sum_{l\in L_T}\sum_{l'\in B_P}\sum_{t'=r+T-\Delta}^{r+T-1} T^{s2}_{ll'pt'} - df^s_{pt}\Big),\ \ \forall s \in S, p \in P, t = r + T - 1 \tag{30}$$

$$SR^s_{pt} = Y^s_{pt} \times sc_p \times \left(df^s_{p(t'=r+T-1)} - \sum_{l\in B_W}\sum_{l'\in L_O}\sum_{w} T^{s1}_{ll'wpt} \times yro_{wp} \times yra_p \times yrt_p \times gd_{wp}\right),\ \ \forall s \in S, p \in P,\ t = r + T - \Delta \tag{31}$$

$$SR_{pt}^{s} = Y_{pt}^{s} \times sc_{p} \times (df_{pt}^{s} - \sum_{l \in L_T} \sum_{l' \in B_P} \sum_{t'=r+T-\Delta}^{r+T-1} T_{ll'pt'}^{s2}), \ \forall s \in S, p \in P, t = r + T - 1 \quad (32)$$

$$\begin{aligned} df_{p(t'=r+T-1)}^{s} \geq & \sum_{l \in B_W} \sum_{l' \in L_O} \sum_{w} T_{ll'wpt}^{s1} \times yro_{wp} \times yra_{p} \times yrt_{p} \times gd_{wp} \\ & - UBP \times \left(1 - Y_{pt}^{s}\right), \\ & \forall s \in S, p \in P, \ t = r + T - \Delta \end{aligned} \quad (33)$$

$$df_{pt}^{s} \geq \sum_{l' \in L_T} \sum_{l \in B_P} \sum_{t'=r+T-\Delta}^{r+T-1} T_{l'lpt'}^{s2} - UBP \times \left(1 - Y_{pt}^{s}\right), \ \forall s \in S, p \in P, t = r + T - 1 \quad (34)$$

$$\begin{aligned} df_{p(t'=r+T-1)}^{s} < & \sum_{l \in B_W} \sum_{l' \in L_O} \sum_{w} T_{ll'wpt}^{s1} \times yro_{wp} \times yra_{p} \times yrt_{p} \times gd_{wp} + UBP \times Y_{pt}^{s}, \\ & \forall s \in S, p \in P, \ t = r + T - \Delta \end{aligned} \quad (35)$$

$$df_{pt}^{s} < \sum_{l' \in L_T} \sum_{l \in B_P} \sum_{t'=r+T-\Delta}^{r+T-1} T_{l'lpt'}^{s2} + UBP \times Y_{pt}^{s}, \ \forall s \in S, p \in P, \ t = r + T - 1 \quad (36)$$

3.2.5. Production flow

Equations (37) to (40) are the production flow constraints considering the yield rate and gross die quantities. These constraints ensure that the output quantity in each period is consistent with the input quantity, accounting for the lead time at each site and production stage. The corresponding yield rate adjusts it. In the next stage, in which wafers serve as the input and dies as the output, the gross die quantity is considered. The production flow restrictions are:

$$\begin{aligned} Q_{lwpt}^{s2} = & Q_{lwp(t-lt_l)}^{s1} \times yrf_{wp}, \ \forall s \in S, l \in L_F, \ w \in W, p \in P, t \\ & = r + T - \Delta - \delta, \ \ldots, r + T - \Delta, t > lt_l \end{aligned} \quad (37)$$

$$\begin{aligned} Q_{lwpt}^{s2} = & Q_{lwp(t-lt_l)}^{s1} \times yro_{wp}, \ \forall s \in S, l \in L_O, \ w \in W, p \in P, t \\ & = r + T - \Delta, \ \ldots, \ r + T - 1, t > lt_l \end{aligned} \quad (38)$$

$$\begin{aligned} Q_{lpt}^{s2} = & \sum_{w} Q_{lwp(t-lt_l)}^{s1} \times yra_{p} \times gd_{wp}, \ \forall s \in S, l \in L_A, \ p \in P, t \\ & = r + T - \Delta, \ \ldots, \ r + T - 1, t > lt_l \end{aligned} \quad (39)$$

$$\begin{aligned} Q_{lmpt}^{s2} = & Q_{lmp(t-lt_l)}^{s1} \times yrt_{p}, \ \forall s \in S, l \in L_T, \ m \in M, p \in P, t \\ & = r + T - \Delta, \ \ldots, \ r + T - 1, t > lt_l \end{aligned} \quad (40)$$

3.2.6. Transportation flow

Equations (41) to (43) compute the transportation flow constraints between each site, taking transportation lead time into account. These constraints ensure that the quantity of transportation dispatched from the upstream site equals the quantity received at the downstream site. The constraints related to transportation flow are as follows:

$$T^{s2}_{ll'wpt} = T^{s1}_{ll'wp(t-tt_{ll'})}, \ \forall s \in S, l \in L_F, l' \in B_W, w \in W, p \in P, t = r + T - \Delta - \delta, \ \dots, r + T - \Delta, t > tt_{ll'} \tag{41}$$

$$T^{s2}_{ll'wpt} = T^{s1}_{ll'wp(t-tt_{ll'})}, \ \forall s \in S, l \in B_W, L_O, l' \in L_O, L_A, w \in W, p \in P, t = r + T - \Delta, \ \dots, \ r + T - 1, t > tt_{ll'} \tag{42}$$

$$T^{s2}_{ll'pt} = T^{s1}_{ll'p(t-tt_{ll'})}, \ \forall s \in S, l \in L_A, L_T, l' \in L_T, B_P, p \in P, t = r + T - \Delta, \ \dots, \ r + T - 1, t > tt_{ll'} \tag{43}$$

3.2.7. Transportation limit

Equations (44) to (47) are the transportation limit. These constraints ensure that the quantity does not exceed the capacity. Moreover, the constraints require that the transportation quantities must equal the finished manufactured product at each site, since we do not account for inventory. The following equations are the transportation limit constraints:

$$\sum_{l' \in B_W} T^{s1}_{ll'wpt} = Q^{s2}_{lwpt}, \ \forall s \in S, l \in L_F, \ w \in W, p \in P, t = r + T - \Delta - \delta, \ \dots, r + T - \Delta \tag{44}$$

$$\sum_{l' \in L_A} T^{s1}_{ll'wpt} = Q^{s2}_{lwpt}, \ \forall s \in S, l \in L_O, \ w \in W, p \in P, t = r + T - \Delta, \ \dots, \ r + T - 1 \tag{45}$$

$$\sum_{l' \in L_T} T^{s1}_{ll'pt} = Q^{s2}_{lpt}, \ \forall s \in S, l \in L_A, \ p \in P, t = r + T - \Delta, \ \dots, \ r + T - 1 \tag{46}$$

$$\sum_{l' \in B_P} T^{s1}_{ll'pt} = \sum_{m} Q^{s2}_{lmpt}, \ \forall s \in S, l \in L_T, \ p \in P, t = r + T - \Delta, \ \dots, \ r + T - 1 \tag{47}$$

3.2.8. Production and transportation balance

Equations (48) to (51) ensure that, once the product arrives at the previous site, it is produced immediately at the current site. These constraints reflect the production and transportation balance and are presented as follows:

$$\sum_{l' \in L_F} T^{s2}_{l'lwpt} = \sum_{l' \in L_O} T^{s1}_{ll'wpt}\,, \quad \forall s \in S, l \in B_W, w \in W, p \in P, t = r + T - \Delta - \delta, \; \dots, r + T - \Delta \tag{48}$$

$$\sum_{l' \in B_W} T^{s2}_{l'lwpt} = Q^{s1}_{lwpt}, \quad \forall s \in S, l \in L_O, w \in W, p \in P, t = r + T - \Delta, \dots, r + T - 1 \tag{49}$$

$$\sum_{l' \in L_O} \sum_{w} T^{s2}_{l'lwpt} = \sum_{w} Q^{s1}_{lwpt}\,, \quad \forall s \in S, l \in L_A, p \in P, t = r + T - \Delta, \dots, r + T - 1 \tag{50}$$

$$\sum_{l' \in L_A} T^{s2}_{l'lpt} = \sum_{m} Q^{s1}_{lmpt}\,, \quad \forall s \in S, l \in L_T, p \in P, t = r + T - \Delta, \dots, r + T - 1 \tag{51}$$

3.2.9. Upper bound, lower bound, and binary variables

Equations (52) to (67) represent the upper-bound constraints, including the production quantities, transportation quantities, and capacity levels. Equations (68) to (78) are the lower-bound constraints, which ensure the related decision variables are non-negative. Equations (79) to (82) are the binary constraints to determine whether the product is produced and the type of regret.

$$Q^{s1}_{lwpt} \le UBP \times a_{wp}\,, \quad \forall s \in S, l \in L_F, w \in W, p \in P, t = r + T - \Delta - \delta, \; \dots, r + T - \Delta \tag{52}$$

$$Q^{s1}_{lwpt} \le UBP \times a_{wp}\,, \quad \forall s \in S, l \in L_O, L_A, w \in W, p \in P, t = r + T - \Delta, \dots, r + T - 1 \tag{53}$$

$$Q^{s1}_{lmpt} \le UBP \times t_{lmp}\,, \quad \forall s \in S, l \in L_T, m \in M, p \in P, t = r + T - \Delta, \dots, r + T - 1 \tag{54}$$

$$Q^{s2}_{lwpt} \le UBP \times a_{wp}\,, \quad \forall s \in S, l \in L_F, w \in W, p \in P, t = r + T - \Delta - \delta, \; \dots, r + T - \Delta \tag{55}$$

$$Q^{s2}_{lwpt} \le UBP \times a_{wp}\,, \quad \forall s \in S, l \in L_O, w \in W, p \in P, t = r + T - \Delta, \dots, r + T - 1 \tag{56}$$

$$Q^{s2}_{lpt} \le UBP\,, \quad \forall s \in S, l \in L_A, p \in P, t = r + T - \Delta, \dots, r + T - 1 \tag{57}$$

$$Q^{s2}_{lmpt} \leq UBP \times t_{lmp}\ ,\ \ \forall s \in S, l \in L_T, m \in M, p \in P, t = r+T-\Delta, \dots, r+T-1 \tag{58}$$

$$T^{s1}_{ll'wpt} \leq UBT \times a_{wp}\ ,\ \ \forall s \in S, l \in L_F,\ l' \in B_W, w \in W, p \in P, t = r+T-\Delta-\delta,\ \dots, r+T-\Delta \tag{59}$$

$$T^{s1}_{ll'wpt} \leq UBT \times a_{wp}\ ,\ \ \forall s \in S, l \in B_W, L_O, l' \in L_O, L_A, w \in W, p \in P, t = r+T-\Delta, \dots, r+T-1 \tag{60}$$

$$T^{s1}_{ll'pt} \leq UBT\ ,\ \ \forall s \in S, l \in L_A, L_T, l' \in L_T, B_P, p \in P, t = r+T-\Delta, \dots, r+T-1 \tag{61}$$

$$T^{s2}_{ll'wpt} \leq UBT \times a_{wp}\ ,\ \ \forall s \in S, l \in L_F, l' \in B_W, w \in W, p \in P, t = r+T-\Delta-\delta,\ \dots, r+T-\Delta \tag{62}$$

$$T^{s2}_{ll'wpt} \leq UBT \times a_{wp}\ ,\ \ \forall s \in S, l \in B_W, L_O, l' \in L_O, L_A, w \in W, p \in P, t = r+T-\Delta, \dots, r+T-1 \tag{63}$$

$$T^{s2}_{ll'pt} \leq UBT\ ,\ \ \forall s \in S, l \in L_A, L_T, l' \in L_T, B_P, p \in P, t = r+T-\Delta, \dots, r+T-1 \tag{64}$$

$$CE_{lt} \leq UBC_l\ ,\ \ \forall l \in L_F,\ t = r+T-\Delta-\delta,\ \dots, r+T-\Delta \tag{65}$$

$$CE_{lt} \leq UBC_l\ ,\ \ \forall l \in L_O, L_A,\ t = r+T-\Delta,\ \dots, r+T-1 \tag{66}$$

$$CE_{lmt} \leq UBC_{lm}\ ,\ \ \forall s \in S, l \in L_T, m \in M, t = r+T-\Delta, \dots, r+T-1 \tag{67}$$

$$Q^{s1}_{lwpt} \geq 0, Q^{s2}_{lwpt} \geq 0\ ,\ \ \forall l \in L_F, w \in W, p \in P, t = r+T-\Delta-\delta,\ \dots, r+T-\Delta \tag{68}$$

$$Q^{s1}_{lwpt} \geq 0, Q^{s2}_{lwpt} \geq 0, Q^{s2}_{lpt} \geq 0, Q^{s1}_{lmpt} \geq 0, Q^{s2}_{lmpt} \geq 0\ , \quad \forall s \in S, l \in L_O, L_A, L_T, w \in W, m \in M, p \in P, t = r+T-\Delta, \dots, r+T-1 \tag{69}$$

$$T^{s1}_{ll'wpt} \geq 0, T^{s2}_{ll'wpt} \geq 0,\ \ \forall s \in S, l \in L_F,\ l' \in B_W,\ w \in W, p \in P, t = r+T-\Delta-\delta,\ \dots, r+T-\Delta \tag{70}$$

$$T_{ll'wpt}^{s1} \geq 0, T_{ll'wpt}^{s2} \geq 0, T_{ll'pt}^{s1} \geq 0, T_{ll'pt}^{s2} \geq 0, \tag{71}$$
$$\forall s \in S, l \in B_W, L_O, L_A, L_T,\ l' \in L_O, L_A, L_T, B_P, w \in W, p \in P, t = r + T - \Delta, \dots, r + T - 1$$

$$CE_{lt} \geq 0,\ \ \forall l \in L_F,\ t = r + T - \Delta - \delta,\ \ \dots, r + T - \Delta \tag{72}$$

$$CE_{lt} \geq 0,\ \ \forall l \in L_O, L_A,\ t = r + T - \Delta, \dots, r + T - 1 \tag{73}$$

$$CE_{lmt} \geq 0,\ \ \forall l \in L_T,\ m \in M, t = r + T - \Delta, \dots, r + T - 1 \tag{74}$$

$$OR_{lt}^{s} \geq 0, SR_{lt}^{s} \geq 0,\ \ \forall s \in S, l \in L_F, t = r + T - \Delta - \delta,\ \ \dots, r + T - \Delta \tag{75}$$

$$OR_{lt}^{s} \geq 0, SR_{lt}^{s} \geq 0,\ \ \forall s \in S, l \in L_O, L_A, t = r + T - \Delta, \dots, r + T - 1 \tag{76}$$

$$OR_{lmt}^{s} \geq 0, SR_{lmt}^{s} \geq 0,\ \ \forall s \in S, l \in L_T,\ m \in M, t = r + T - \Delta, \dots, r + T - 1 \tag{77}$$

$$OR_{pt}^{s} \geq 0,\ SR_{pt}^{s} \geq 0,\ \ \forall s \in S,\ p \in P,\ t = r + T - \Delta, r + T - 1 \tag{78}$$

$$X_{lpt}^{s} \in \{0,1\},\ \ \forall s \in S, l \in L_F,\ L_O,\ L_A,\ L_T,\ p \in P,\ t = r + T - \Delta - \delta,\ \ \dots, r + T - 1 \tag{79}$$

$$Y_{lt}^{s} \in \{0,1\},\ \ \forall s \in S, l \in L_F,\ L_O,\ L_A,\ t = r + T - \Delta - \delta,\ \ \dots, r + T - 1 \tag{80}$$

$$Y_{lmt}^{s} \in \{0,1\},\ \ \forall s \in S, l \in L_T,\ m \in M,\ t = r + T - \Delta, \dots, r + T - 1 \tag{81}$$

$$Y_{pt}^{s} \in \{0,1\},\ \ \forall s \in S,\ p \in P,\ t = r + T - \Delta, r + T - 1 \tag{82}$$

The research objective is to identify optimal capacity planning strategies that effectively meet demand while minimizing total costs. Figure 8 depicts all the stages and sites being considered in the planning network, which comprises four production stages and two inventory stages: wafer fabrication ($l \in L_F$), wafer probing ($l \in L_O$), assembly ($l \in L_A$), and final testing ($l \in L_T$), along with two inventory stages: wafer bank ($b \in B_W$) and finished goods bank ($b \in B_P$). The capacity planning level in each stage and site is marked as capacity (CE_{lt}, CE_{lmt}). The production lead time in each stage is marked as time (lt_l). The input and output production quantities are marked as quantity ($Q_{lwpt}^{s1}, Q_{lmpt}^{s1}, Q_{lwpt}^{s2}, Q_{lpt}^{s2}, Q_{lmpt}^{s2}$), respectively. The transportation lead time is marked as time ($tt_{ll'}$).

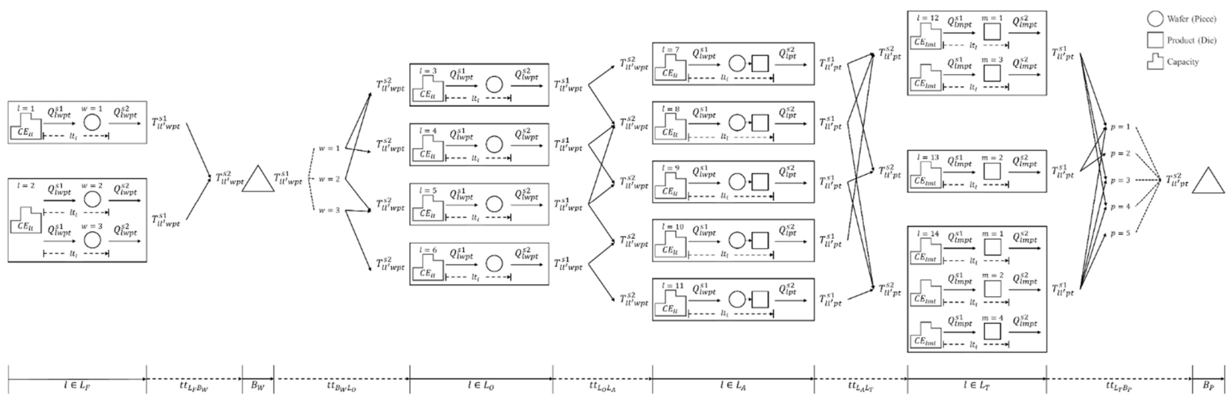


Figure 8. The supply chain planning network for the IC design house.

3.3. *Capacity regret methodology*

Since demand forecasts change in the rolling-horizon environment, while capacity levels must be determined in advance, we employ stochastic capacity planning strategies to fulfill actual orders. To effectively meet the dynamic actual orders in the future, we apply and compare the results of four capacity planning models under different demand covariance (CV) levels, which are the minimax regret (RHSCCP-MMR) model, the upper bound and lower bound (RHSCCP-ULB) model, the expected value (RHSCCP-EV) model, and the stochastic programming (RHSCCP-SP) model.

3.3.1. MMR model

The objective of the RHSCCP-MMR model is to minimize the maximum total cost under all demand scenarios of the sum of production, transportation, and regret costs, which are presented as equations (1) to (5). In this model, the capacity planning strategy considers all demand forecast scenarios, so the input size of the scenario set to the model is $|S|$.

3.3.2. ULB model

To plan capacity under significant demand fluctuations, this model identifies upper and lower bounds for demand scenarios to determine the capacity level under extreme demand conditions. The upper bound in the scenario set is defined as $df_{pt}^{U} = \text{Maximum}\{df_{pt}^{s}, \forall s \in S\}$, and the lower bound is $df_{pt}^{L} = \text{Minimum}\{df_{pt}^{s}, \forall s \in S\}$, which indicates the Maximum and Minimum of demand scenarios, respectively. Consequently, the input scenario size for the capacity planning model is two.

3.3.3. EV model

In this capacity planning model, we assume that all demand scenarios have equal probability of being realized in the future and use the expected value of the demand forecast as the input to determine the appropriate capacity level in the supply chain. The expected value is defined as $df_{pt} =$

$\frac{1}{|S|}\sum_{s\in S} df_{pt}^{s}$. Since we perform capacity planning by considering the expected demand scenario, the input scenario size for the capacity planning model is one.

3.3.4. SP model

The stochastic programming model evaluates the uncertain demand scenarios across probabilistic events and identifies the optimal solution by optimizing the objective value across all possible demand scenarios. Consequently, in this model, the objective function $\text{Minimize } Z = \theta$ is replaced by $\text{Minimize } Z = \frac{1}{|S|}\sum_{s\in S}\theta^{s}$. Additionally, equation (2) is adjusted from $\theta \geq C_{production}^{s} + C_{transportation}^{s} + C_{regret}^{s}\ \forall s \in S$ to $\theta^{s} \geq C_{production}^{s} + C_{transportation}^{s} + C_{regret}^{s}\ \forall s \in S$. Since we take all demand scenarios into account, the scenario size for this model is $|S|$, which is the same as the MMR model. Table 1 shows the difference between the four capacity planning models.

Table 1. The comparison between four capacity planning models.

	RHSCCP-MMR		RHSCCP-ULB	
Objective Function	$\text{Minimize } Z = \theta$			(1)
Related Constraint	$\theta \geq C_{production}^{s} + C_{transportation}^{s} + C_{regret}^{s}\ \forall s \in S$			(2)
Demand Forecast	$df_{pt}^{s}\ \forall s \in S$		$df_{pt}^{U} = \text{Maximum}\{df_{pt}^{s}\ \forall s \in S\}$ $df_{pt}^{L} = \text{Minimum}\{df_{pt}^{s}\ \forall s \in S\}$	
Scenario Sets	$\lvert S\rvert$		2	
	RHSCCP-EV		**RHSCCP-SP**	
Objective Function	$\text{Minimize } Z = \theta$	(1)	$\text{Minimize } Z = \frac{1}{\lvert S\rvert}\sum_{s\in S}\theta^{s}$	(1a)
Related Constraint	$\theta \geq C_{production} + C_{transportation} + C_{regret}$	(2a)	$\theta^{s} \geq C_{production}^{s} + C_{transportation}^{s} + C_{regret}^{s}\ \forall s \in S$	(2b)
Demand Forecast	$df_{pt} = \frac{1}{\lvert S\rvert}\sum_{s\in S} df_{pt}^{s}$		$df_{pt}^{s}\ \forall s \in S$	
Scenario Sets	1		$\lvert S\rvert$	

3.4. Forecast evolution methodology

3.4.1. Additive MMFE

The additive MMFE model employs the forecast evolution method as equations (83) and (84). In the new review epoch $r+1$, the forecast update quantity $\varepsilon_{p,t}^{r+1}$ of product p belongs to the update vector $\varepsilon_{p}^{r+1} = [\varepsilon_{p,r+\Delta}^{r+1}, \varepsilon_{p,\Delta+1}^{r+1}, \dots, \varepsilon_{p,r+T-2}^{r+1}, \varepsilon_{p,r+T-1}^{r+1}]$. The vector describes the planning horizon periods $[ao_{p,r}^{r+1}, \dots, ao_{p,r+\Delta-1}^{r+1}, df_{p,+\Delta}^{r+1}, \dots, df_{p,+T-1}^{r+1}]$, which includes two parts: The frozen received actual orders $[ao_{p,r}^{r+1}, \dots, ao_{p,r+\Delta-1}^{r+1}]$, and the non-fixed demand forecasts $[df_{p,r+\Delta}^{r+1}, \dots, df_{p,r+T-1}^{r+1}]$. The formula for the additive MMFE is as follows:

$$df_{p,t}^{r+1} = df_{p,t}^{r} + \varepsilon_{p,t}^{r+1}, \forall p \in P, t = r + \Delta + 1, \ \dots, r + T - 1 \tag{83}$$

$$ao_{p,t}^{r+t-1} = df_{p,t}^{r} + \sum_{\tau=1}^{t} \varepsilon_{p,t-\tau+1}^{r+\tau}, \forall p \in P, t = r + \Delta, \ \dots, r + T - 1 \tag{84}$$

The forecast update vector in the additive MMFE model follows a multivariate normal distribution $\varepsilon_p^r \sim MN(\mu, \Sigma)$ with the mean value $\mu = 0$ and covariance matrix $\Sigma = \begin{bmatrix} (\sigma_{r+\Delta}^p)^2 & \cdots & \sigma_{r+\Delta,r+T-1}^p \sigma_{r+\Delta}^p \sigma_{r+T-1}^p \\ \vdots & \sigma_{t_1,t_2}^p \sigma_{t_1}^p \sigma_{t_2}^p & \vdots \\ \sigma_{r+T-1,r+\Delta}^p \sigma_{r+T-1}^p \sigma_{r+\Delta}^p & \cdots & (\sigma_{r+T-1}^p)^2 \end{bmatrix}$. The σ_t^p represents the standard deviation of product p in period t, and σ_{t_1,t_2}^p denotes the correlation of product p between different planning periods t_1 and t_2. The forecast update vectors are observed independently in each review epoch, with the covariance between the demands of updated products in different planning periods provided by equation (85):

$$\gamma_{t1,t2}^p = Cov\left(df_{p,t1}^r, df_{p,t2}^r\right) = \sum_{\tau=1}^{min\,(t1,t2)} \sigma_{t_1-\tau+1,t_2-\tau+1}^p \sigma_{t_1-\tau+1}^p \sigma_{t_2-\tau+1}^p \tag{85}$$

3.4.2. Multiplicative MMFE

The multiplicative MMFE model uses the forecast evolution method as described by equations (86) and (87), where the forecast update process is similar to the additive MMFE model. In each new review epoch $r + 1$, the planning horizon rolls forward by one planning period, and the actual orders and demand forecasts are updated by multiplying the forecast update quantity by an exponential factor. The forecast evolution method for multiplicative MMFE is as follows:

$$df_{p,t}^{r+1} = df_{p,t}^{r} \times exp(\varepsilon_{p,t}^{r+1}), \forall p \in P, t = r + \Delta + 1, \ \dots, r + T - 1 \tag{86}$$

$$ao_{p,t}^{r+t-1} = df_{p,t}^{r} \times exp\,(\sum_{\tau=1}^{t} \varepsilon_{p,t-\tau+1}^{r+\tau}), \forall p \in P, t = r + \Delta, \ \dots, r + T - 1 \tag{87}$$

The forecast update vector in the multiplicative MMFE model follows a multivariate normal distribution $\varepsilon_p^r \sim MN(\mu, \Sigma)$, where each marginal distribution is given by $\varepsilon_{p,t}^r \sim N\left(-\frac{(\sigma_t^p)^2}{2}, (\sigma_t^p)^2\right)$, with the mean $\mu = -\frac{(\sigma_t^p)^2}{2}$ and covariance matrix $\Sigma = (\sigma_t^p)^2$.

The demand covariance can be analyzed by the covariance of the forecast evolution process in the log domain, as shown by equation (88):

$$\gamma_{t1,t2}^{p} = Cov\left(log\ (df_{p,t1}^{r}), log\ (df_{p,t2}^{r})\right) = \sum_{\tau=1}^{min\ (t1,t2)} \sigma_{t_1-\tau+1,t_2-\tau+1}^{p}\sigma_{t_1-\tau+1}^{p}\sigma_{t_2-\tau+1}^{p} \tag{88}$$

The forecast update process in additive and multiplicative MMFE models is unbiased, and there exists a fundamental distinction between the two MMFE models. In the additive MMFE model, the demand uncertainty of the forecast value is determined by the variance of the forecast update distribution, which is normally distributed. In contrast, in the multiplicative MMFE model, the demand uncertainty of forecast updates is directly related to the forecast value and follows a lognormal distribution. Consequently, the multiplicative MMFE model exhibits a heavier tail compared to the additive MMFE model. Due to these characteristics, the multiplicative MMFE model is considered a more suitable approach for simulating the demand update process in practical scenarios, as it evaluates the forecast value in a relative manner.

4. Simulation experiment

4.1. Simulation environment

The simulation environment is written in the Julia programming language, and the supply chain capacity planning models are solved using Gurobi 11.0. The simulation experiments are conducted on a computer equipped with a 2.9 GHz Intel Core i5-10400 CPU and 16 GB of RAM.

4.2. Demand generation scheme

Given that demand generation occurs in a rolling-horizon environment, we utilize demand forecasts to plan the capacity level for each planning period. As time passes, these demand forecasts will evolve into actual orders, which are then used to evaluate the performance of the capacity planning models. Figure 9 shows the evolution of the forecast. Demand data can be categorized into two types: Actual orders during the frozen period, which cannot be adjusted, and demand forecasts, which can be continuously revised at each review epoch. The demand forecast at the new epoch is updated by the forecast update vector from the previous epoch until it becomes an actual order. In this paper, we utilize demand forecast data as in-sample data to determine the capacity level. The IC design house conducts production planning based on this capacity level and the actual orders received. The actual orders derived from the demand forecast data are considered out-of-sample and used to evaluate the model's performance. The rolling-horizon framework with demand generation involves several steps, as illustrated in Figure 10.

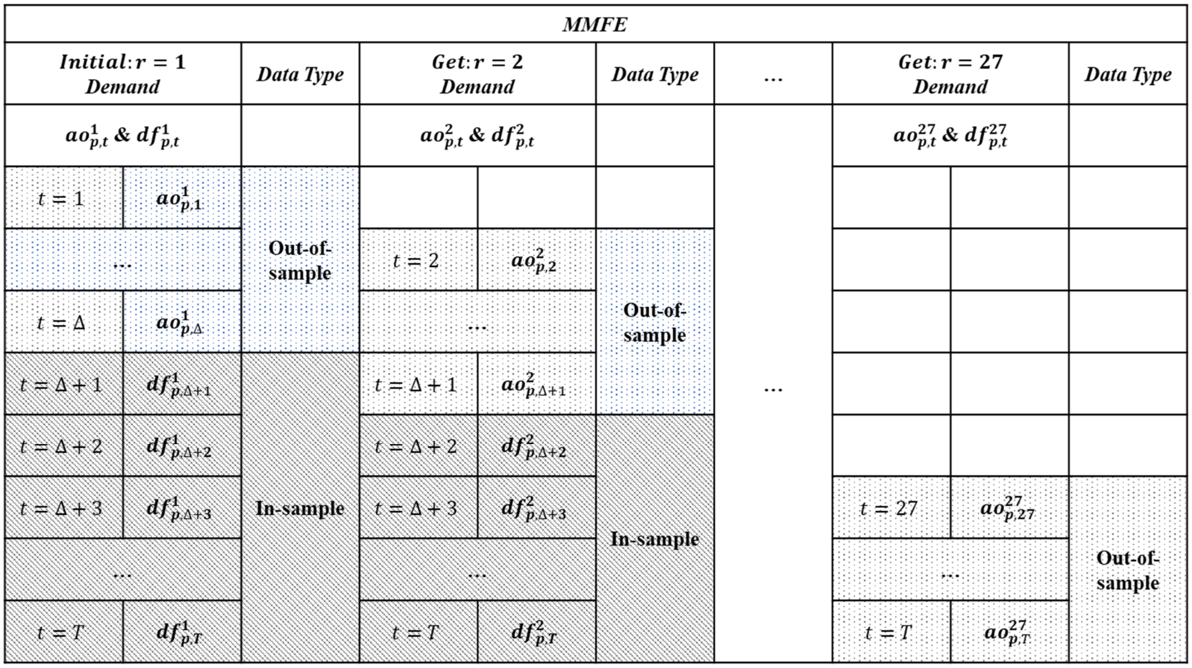


Figure 9. The example of in-sample and out-of-sample data process.

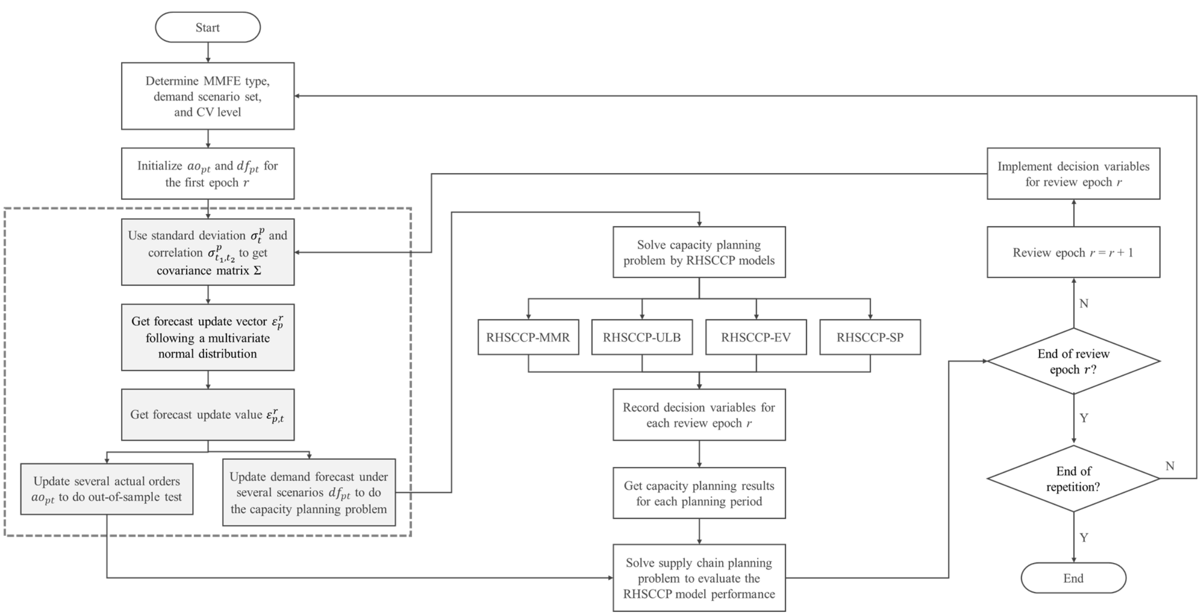


Figure 10. The process of the rolling-horizon framework.

4.3. Factor and level

For the simulation experiment, we aim to evaluate the performance and effectiveness of capacity planning models under various demand scenarios. Table 2 lists the factors and levels used in the experiments. Factors are divided into demand- and capacity-planning model factors. The MMFE model has two levels: additive and multiplicative. The CV level factor has three levels: low, middle, and high. The capacity planning models have four levels: RHSCCP-MMR, RHSCCP-ULB, RHSCCP-EV, and RHSCCP-SP, which determine the capacity planning strategies used in the experiments. These factors and levels are carefully designed to examine the performance and significance of the simulation experiments across demand scenarios and planning models.

Table 2. List of factors and levels.

Factor		Level				Count
Demand	*MMFE Type	Additive		Multiplicative		2
	*CV Level	Low	Middle		High	3
	Repetition					30
	Total Demand = 2 × 3 × 30					**180**
Supply Chain Capacity Planning Model	*Model Type	RHSCCP-MMR	RHSCCP-ULB	RHSCCP-EV	RHSCCP-SP	4
	Total Capacity Planning Model = 4					**4**
Total Simulation Experiment = 180 × 4						**720**

To address evolving market demand in the semiconductor industry, the IC design house needs to plan capacity for outsourcers across several demand forecast scenarios. Our objective of this research is to evaluate the performance of supply chain capacity planning models under different demand scenarios and capacity strategies. First, we employ the forecast evolution method with three CV levels to simulate the real-world demand variability over time. Second, we compare the performance of four capacity models to assess their suitability in different situations.

Each rolling-horizon simulation experiment includes 24 instances, each repeated 30 times to ensure reliability. Capacity planning models are evaluated using seven performance indices: average capacity level (in-sample), average total cost (in-sample), average total capacity action counts (in-sample), average total capacity adjustment (in-sample), average fulfillment rate (out-of-sample), average capacity utilization rate (out-of-sample), and average total cost (out-of-sample).

4.4. In-sample experiment result analysis

Capacity planning models are evaluated based on the planned capacity level to assess differences in capacity planning behavior across various demand scenarios, using four performance indices: average capacity level, average total cost, average total capacity action counts, and average total capacity adjustments.

4.4.1. Average capacity level

Figure 11 shows the average capacity level under different demand scenarios and RHSCCP models. The capacity levels during the planning periods vary based on the strategies applied in the four RHSCCP models. In the additive and multiplicative MMFE models, the average capacity levels, from highest to lowest, are: RHSCCP-ULB, RHSCCP-MMR, RHSCCP-SP, and RHSCCP-EV.

These findings yield three insights. First, the RHSCCP-EV model offers an optimistic solution by considering only a single expected demand scenario, resulting in the lowest capacity level. Second, the RHSCCP-ULB model takes the most conservative approach by considering maximum and minimum demand scenarios to mitigate future shortage risks, leading to the highest capacity level. Last, the RHSCCP-MMR and RHSCCP-SP models plan similar capacity levels. However, the RHSCCP-MMR model, which assumes the maximum total cost, yields a higher average capacity than the RHSCCP-SP model, which accounts for total cost across all scenarios.

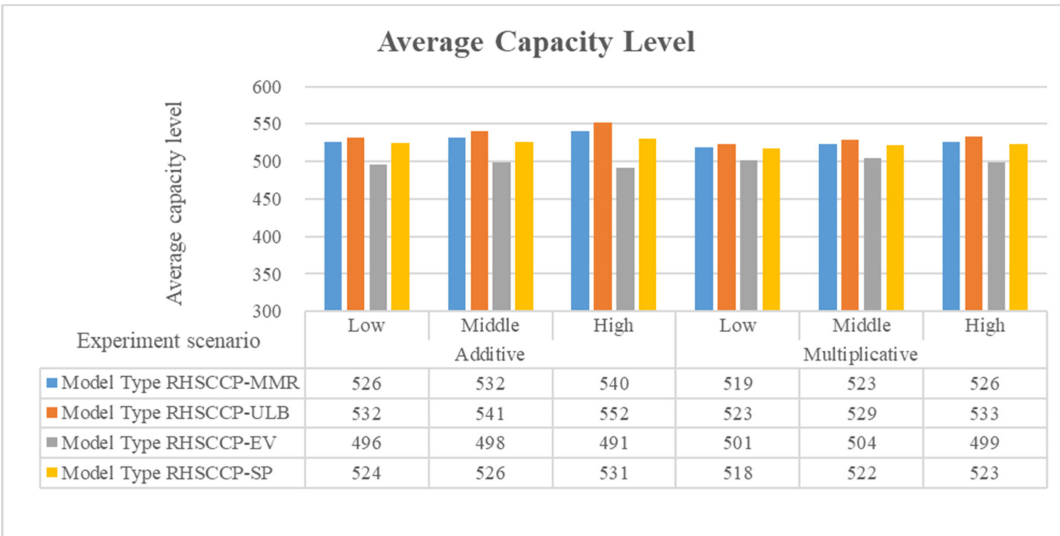


Figure 11. Summary of average capacity level for the in-sample experiment.

4.4.2. Average total cost

Figure 12 shows the average total cost based on in-sample demand data, serving as input to the RHSCCP model. The total cost includes production, transportation, and regret costs, as defined in the objective function. Results indicate that the total cost rises with the CV level. In the additive and multiplicative MMFE models, the average total cost from high to low is: RHSCCP-ULB, RHSCCP-MMR, RHSCCP-SP, and RHSCCP-EV. This order matches the observed average capacity levels. Since capacity level dictates production and transportation quantities at each site and stage, it directly influences production and transportation costs, which in turn affect the total cost. Thus, the average total cost distribution pattern mirrors the average capacity levels.

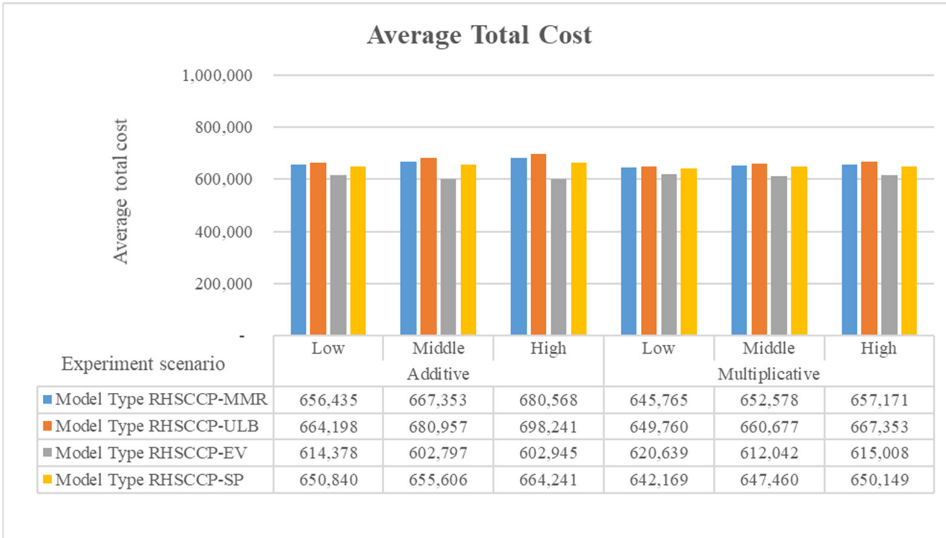


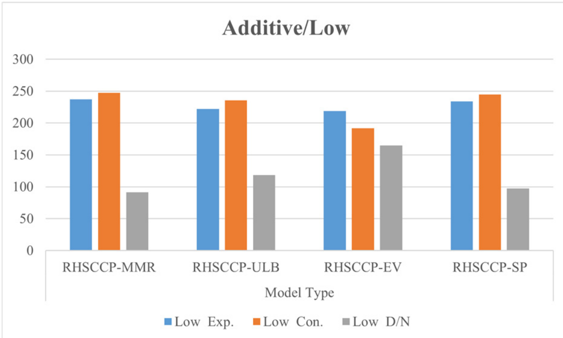
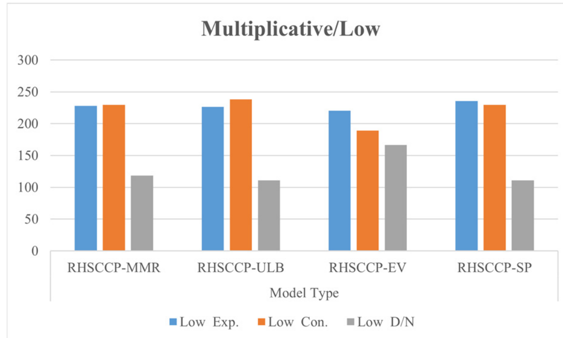
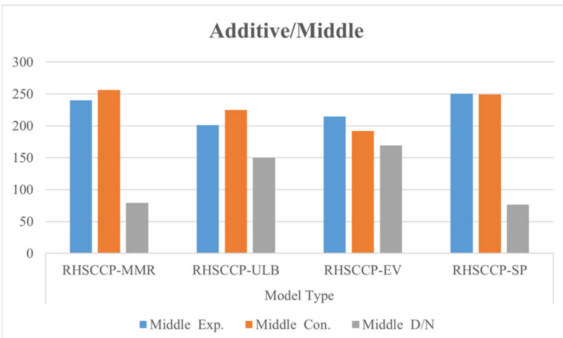
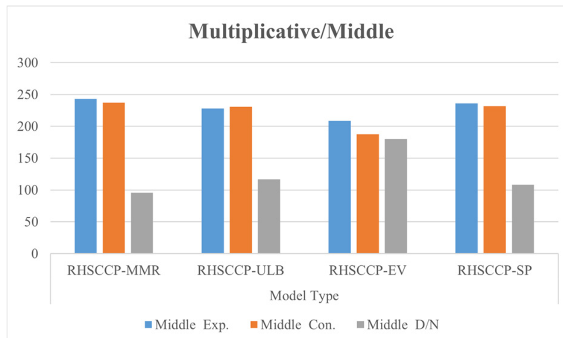
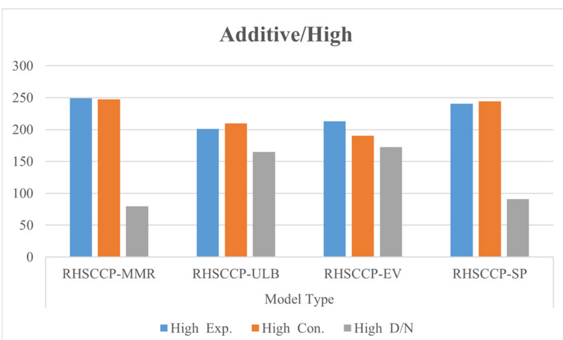
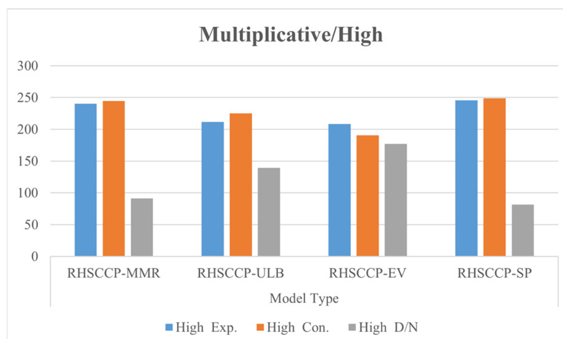
Figure 12. Summary of average total cost for the in-sample experiment.

4.4.3. Average total capacity action counts

To investigate the impact of demand fluctuations on capacity planning behavior, we define three capacity actions: Capacity expansion (Exp.) decision, where the capacity level in the current planning period is higher than the last period; capacity contraction (Con.) decision, where the capacity level in the current planning period is lower than the last period; and do-nothing (D/N) decision, where the capacity level in the current planning period remains unchanged from the last period.

Table 3 presents the average total capacity action counts across the four RHSCCP models. Three insights can be drawn from the results. First, compared with the other three RHSCCP models, the RHSCCP-EV model takes more do-nothing actions because it dilutes the worst-case demand scenario by considering only one expected demand scenario, making it less sensitive to extreme scenarios. Second, the RHSCCP-ULB model shows an increase in do-nothing actions as the CV level rises. Based on the average planned capacity level discussed in the previous section, it is evident that the RHSCCP-ULB model tends to maintain a high-capacity level, leading to an increased number of do-nothing actions. Last, the RHSCCP-MMR and RHSCCP-SP models individually adopt similar capacity action counts under different CV levels. This consistency is due to the model's sensitivity, which dynamically adjusts capacity levels based on current demand scenarios at each epoch. More significant differences are observed in the adjustment amounts for capacity levels, which will be discussed in the following section.

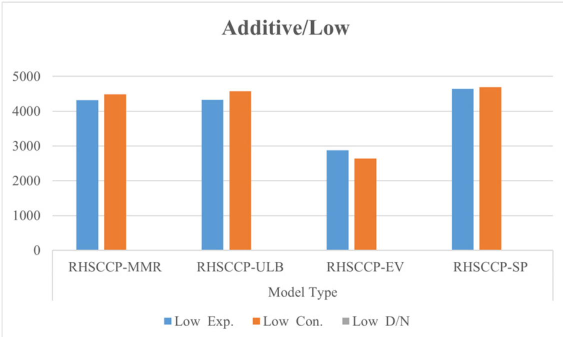
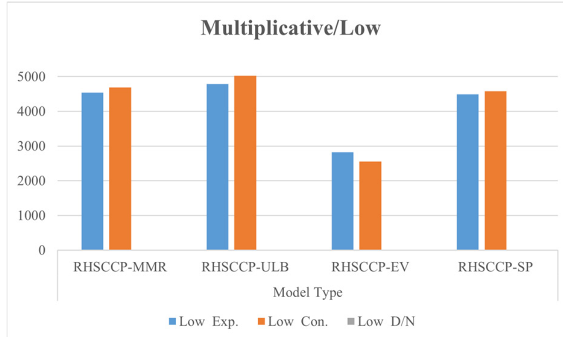
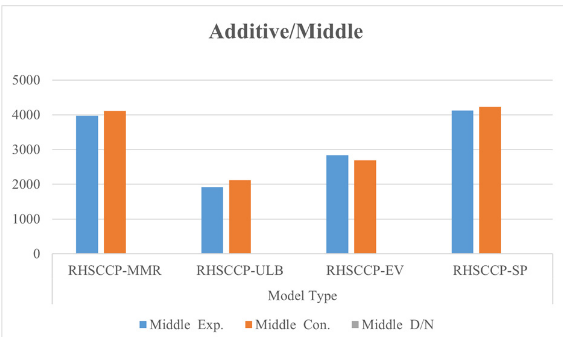
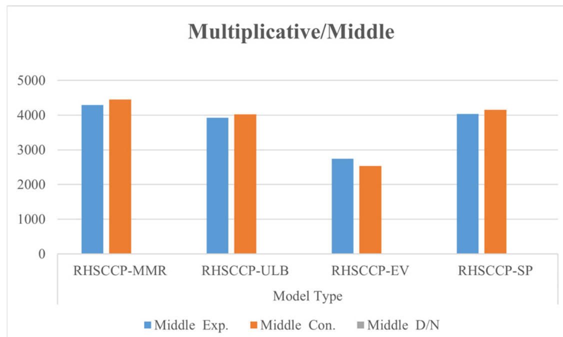
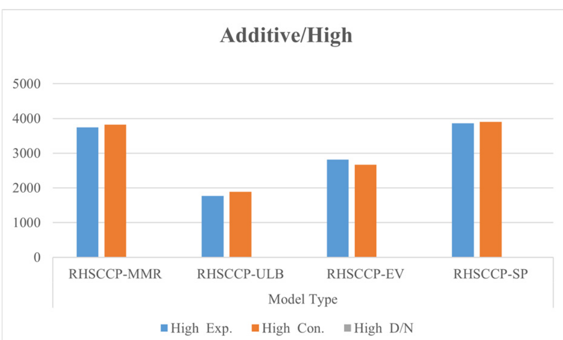
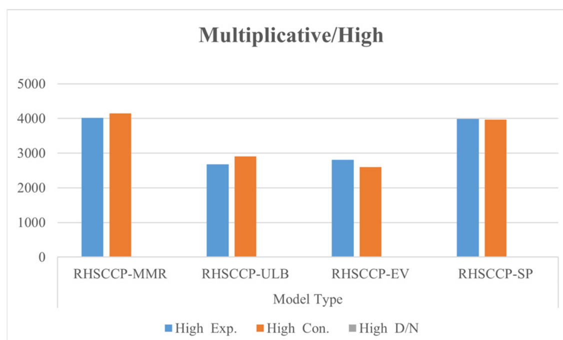
Table 3. Summary of average total capacity action counts.

	Additive MMFE	Multiplicative MMFE
CV Level (Low)		
CV Level (Middle)		
CV Level (High)		

4.4.4. Average total capacity adjustment

Table 4 shows the average total capacity adjustment for each RHSCCP model. All models reduce their total capacity adjustment amounts as the CV level increases, mirroring the pattern of average capacity actions. The RHSCCP-ULB model exhibits a significant decline in capacity adjustments with rising CV levels, indicating that the model maintains a high-capacity level, leading to more do-nothing actions and fewer adjustments. We suggest a stable approach to capacity adjustments. Conversely, the RHSCCP-MMR and RHSCCP-SP models demonstrate relatively higher capacity adjustment amounts, indicating a more fluctuating approach to capacity adjustments compared to the other models.

Table 4. Summary of average total capacity adjustment.

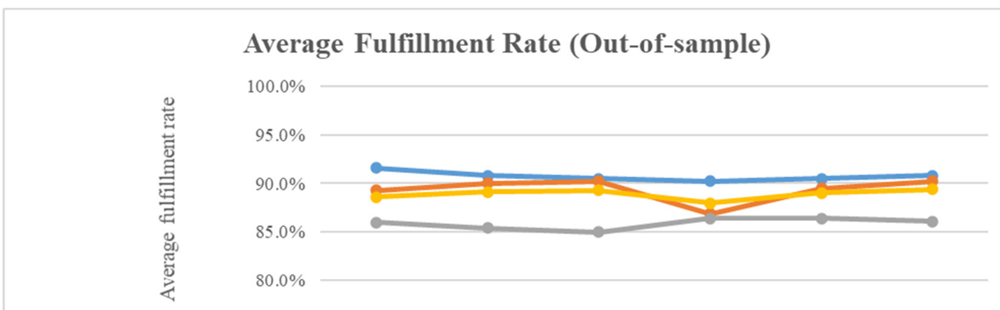
	Additive MMFE	Multiplicative MMFE
CV Level (Low)		
CV Level (Middle)		
CV Level (High)		

4.5. Out-of-sample experiment result analysis

Given that capacity level directly influences the production plan, capacity planning models are evaluated using actual orders to assess whether the planned capacity can meet final demand. This evaluation is based on three performance indices: average fulfillment rate, average capacity utilization rate, and average total cost.

4.5.1. Average fulfillment rate

Since demand data will be continuously updated and converted into actual orders after 26 weeks, we use out-of-sample demand data to represent these actual orders and evaluate the performance of the capacity planning models. The fulfillment rate represents the proportion of final demand that the planned capacity level can satisfy. Figure 13 illustrates the average fulfillment rate. The results show that the RHSCCP-MMR model achieves the highest fulfillment rate across the MMFE demand scenarios, as indicated by the numerical values reported in Figure 13. This result suggests that the minimax regret-based approach provides a more robust capacity planning strategy for maintaining demand fulfillment under simulated demand uncertainty. The RHSCCP-ULB model achieves relatively high fulfillment performance by reserving a higher capacity level, but it does not outperform the RHSCCP-MMR model. The RHSCCP-SP model shows comparable but slightly lower fulfillment rates than the RHSCCP-MMR model in most scenarios. In contrast, the RHSCCP-EV model produces the lowest fulfillment rate among the compared models. This result reflects its optimistic planning policy, which tends to reserve a lower capacity level across demand scenarios. As a consequence, the RHSCCP-EV model generates less capacity waste and maintains a more stable utilization rate. However, by diluting the effect of the worst demand scenario, the model becomes less responsive to extreme demand realizations. It increases the risk of lower fulfillment performance.



Experiment scenario	Additive Low	Additive Middle	Additive High	Multiplicative Low	Multiplicative Middle	Multiplicative High
Model Type RHSCCP-MMR	91.6%	90.8%	90.5%	90.2%	90.5%	90.8%
Model Type RHSCCP-ULB	89.3%	90.0%	90.2%	86.9%	89.5%	90.2%
Model Type RHSCCP-EV	86.0%	85.4%	85.0%	86.4%	86.4%	86.1%
Model Type RHSCCP-SP	88.6%	89.1%	89.3%	88.0%	89.0%	89.4%

Figure 13. Summary of the average fulfillment rate for the out-of-sample experiment.

4.5.2. Average capacity utilization

An effective capacity planning strategy should balance the risks of shortages and oversupply. The fulfillment rate assesses the capacity shortage level, and the capacity utilization rate evaluates the extent of capacity waste. Therefore, the fulfillment rate can directly correspond to the capacity utilization rate. The capacity utilization rate is the proportion of the planned capacity level used to fulfill actual orders. Figure 14 shows that, in the additive and multiplicative models, the capacity utilization rates of the RHSCCP-ULB and RHSCCP-SP models decline as the CV level increases. This trend corresponds to the progressively increasing planned capacity level and fulfillment rate observed under the higher CV levels. The RHSCCP-EV has a more stable, higher capacity utilization rate across demand scenarios. The RHSCCP-MMR and RHSCCP-ULB models exhibit lower capacity utilization rates, leading to greater capacity waste.

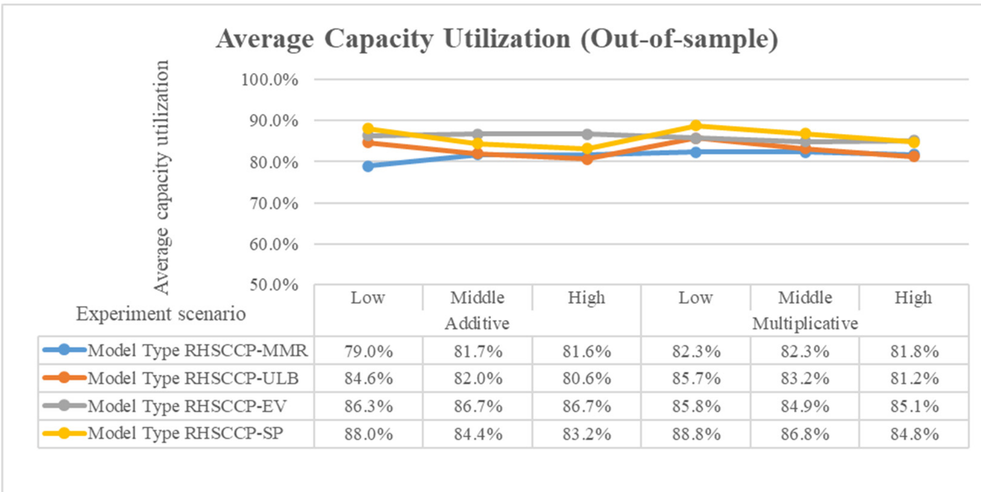


Figure 14. Summary of the average capacity utilization rate for the out-of-sample experiment.

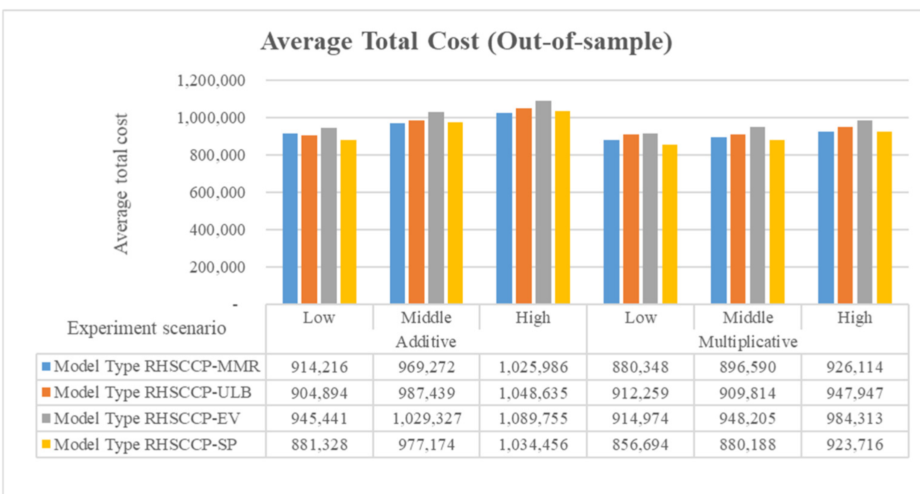


Figure 15. Summary of the average total cost for the out-of-sample experiment.

4.5.3. Average total cost

To assess the performance of the RHSCCP models, the total cost is calculated using out-of-sample data and includes production, transportation, backlog, and capacity waste costs. The backlog cost refers to the cost of unfulfilled orders, while the capacity waste cost pertains to the cost of unused capacity.

Figure 15 shows the average total cost using out-of-sample data. It shows that the average total cost across all RHSCCP models increases with the CV level. The RHSCCP-EV model exhibits the highest total costs among all models. This outcome contrasts with the average total cost calculated from in-sample data and is primarily due to higher backlog costs. In the additive MMFE model, the RHSCCP-SP model incurs slightly higher average total costs compared to the RHSCCP-MMR model. Conversely, in the multiplicative MMFE model, the RHSCCP-SP model demonstrates lower total costs than the RHSCCP-MMR model. In most cases, the RHSCCP-ULB has the second-highest average total cost, due to capacity waste costs.

4.6. Summary of the simulation experiment

The objective of capacity planning is to maximize the fulfillment rate of actual orders and the capacity utilization rate while minimizing capacity waste, thereby ensuring customer needs are met as efficiently as possible. The simulation experiments show that, in the additive and multiplicative MMFE methods, the RHSCCP-ULB models achieve relatively high fulfillment rates; however, their conservative over-planning policy also leads to greater capacity waste. The RHSCCP-EV model adopts an optimistic planning policy. It shows poor fulfillment performance but also less capacity waste. The RHSCCP-MMR model presents the lowest average total cost in the additive MMFE method under the middle and high CV levels. In contrast, the RHSCCP-SP model performs better in the additive MMFE method at low CV levels and in the multiplicative MMFE model across all CV levels.

4.7. Statistical analysis

In this study, we conduct a statistical test with the 0.05 significance level to examine whether MMFE type, CV level, and model type significantly affect total cost. In addition, this test determines whether the effects of these factors depend on one another through two-way and three-way interactions. Table 5 shows the test results.

Table 5. ANOVA test result.

Effect	SS	df	F	p-value	η^2	η_p^2	Conclusion
CV Level	3.83253E+11	2	118.9343935	5.01439E-52	0.009978808	0.013597266	Significant
Model Type	8.62551E+12	3	1784.496624	0	0.224583669	0.236780657	Significant
MMFE Type × CV Level	41117259645	2	12.75987539	2.90102E-06	0.001070576	0.001476708	Significant
MMFE Type × Model Type	5.92411E+11	3	122.561542	1.50771E-78	0.015424698	0.020863096	Significant
CV Level × Model Type	4.92197E+11	6	50.91430185	2.00641E-62	0.012815402	0.017395214	Significant
MMFE Type × CV Level × Model Type	53200038453	6	5.503170424	1.05264E-05	0.001385177	0.001909826	Significant
Error	2.78028E+13	17256					

The results show that all three main effects are statistically significant, as presented by their F-statistics, p-values, η^2, and η_p^2. These findings indicate that total cost differs significantly across MMFE types, CV levels, and model types. Among these three main factors, model type exhibits the highest F-statistic, indicating particularly strong evidence of differences in total cost across model types relative to residual variability. In addition, all two-way interactions are statistically significant: MMFE Type × CV Level, MMFE Type × Model Type, and CV Level × Model Type. Therefore, the effect of each factor on total cost cannot be considered entirely independent of the other factors.

Furthermore, the MMFE Type × CV Level × Model Type interaction is also statistically significant. This result indicates that the effect of model type on total cost varies jointly with MMFE type and CV level; thus, the model's relative performance is condition-dependent rather than uniform across all experimental settings. Consequently, although the significant main effects confirm that MMFE type, CV level, and model type individually contribute to differences in total cost, these main effects should be interpreted cautiously because of the significant higher-order interaction. Overall, total cost is significantly influenced not only by the individual experimental factors but also by their combined effects.

4.8. Comparison of findings with other studies

Semiconductor and production planning studies have established the importance of integrated network, capacity, and dynamic decision-making. The researchers in [14] developed an integrated multi-site planning framework for the TFT-LCD industry, involving production routing, capacity consumption, material supply, and inventory constraints. In the IC design house context, the researchers in [21] extended supply network planning to multiobjective order-promising decisions involving multi-stage and multi-site outsourcing networks, capacity restrictions, product flexibility, and fulfillment considerations, while in their subsequent study [22], they addressed back-end production networks with Turnkey operations. The researchers in [39] further demonstrated the relevance of hierarchical and repeatedly updated decision-making in wafer-fab qualification management through a rolling-horizon simulation setting. Several researchers have also shown that explicitly incorporating forecast evolution can improve planning under demand uncertainty. The researchers in [37] employed multiplicative MMFE into rolling-horizon wafer-fab production planning and found that the chance-constrained model with forecast updates outperformed its counterpart without forecast updates in terms of expected service level and profit. The researchers in [40] integrated MMFE into stochastic lot-sizing within a rolling horizon and reported that forecast-evolution models reduced realized costs compared with deterministic planning. Their results further showed that the performance of forecast-evolution approaches depends on multiple factors. In semiconductor capacity expansion, the researchers in [13] combined forecast evolution with a minimax regret strategy and demonstrated its applicability for developing robust capacity plans under different demand scenarios.

Our findings of this study extend work by providing evidence that the relative effectiveness of capacity-planning strategies depends on the underlying forecast-evolution pattern. Under additive MMFE, particularly at higher demand fluctuations, RHSCCP-MMR provides stronger fulfillment performance and greater robustness to demand uncertainty. In contrast, under multiplicative MMFE, RHSCCP-SP achieves the minimum total cost while maintaining relatively high fulfillment and capacity utilization rates. The results reveal a trade-off between capacity protection and utilization efficiency. RHSCCP-ULB responds conservatively to worst-case demand and protects fulfillment through additional capacity reservation, but this behavior reduces capacity utilization as demand variability increases. In contrast, RHSCCP-EV maintains more stable utilization by reducing the influence of extreme scenarios, but consequently exhibits relatively weaker fulfillment performance. Therefore, the results indicate that demand uncertainty alone is insufficient to determine the appropriate planning strategy; how the demand forecast evolves is also an important determinant of strategy performance.

5. Managerial implication

In this study, we successfully model capacity planning across additive and multiplicative MMFE in a rolling-horizon planning. It is applicable to the IC design house. The models consider several important factors and technical constraints. This section is provided to discuss critical points related to the applicability and managerial implications of this research.

Several assumptions are adopted to maintain the tractability of the proposed long-term capacity planning model. First, the model does not consider setup costs because we focus on strategic capacity-related decisions, including fixed, production, transportation, capacity surplus, and shortage costs. Therefore, setup costs are implicitly included in other cost components. However, this assumption may restrict the model's applicability in real production systems with frequent or significant changeovers and preparation costs.

Second, the model assumes that WIP items are immediately transported to the next stage upon completion of the current production process. As a result, inventory at each stage is not explicitly considered. This assumption simplifies the material-flow representation, but it is suitable only for a continuous flow system with limited intermediate storage. However, it may be less applicable to systems that use buffer inventory or intermediate storage to absorb production variability, transportation delays, or demand uncertainty.

Third, each order is assumed to contain only one product type, which simplifies the demand-allocation process. This assumption is suitable for cases where orders are relatively homogeneous or can be decomposed into single-product demand units. In real settings involving multi-product orders, additional consolidation decisions may be required. Researchers may relax these assumptions by incorporating setup or changeover costs, intermediate inventory, and multi-product order fulfillment decisions.

Moreover, in the RHSCCP model, determining reserved capacity levels for multiple outsourcers requires balancing capacity regret under different demand scenarios. These regrets fall into two types. The first is capacity surplus regret, which occurs when the reserved capacity exceeds the required capacity under the given scenarios. This condition leads to unused reserved capacity and is represented in the model through the capacity oversupply cost. The second is capacity shortage regret, which occurs when the required demand exceeds the reserved capacity. In this type of case, part of the demand requirement cannot be fully covered by reserved capacity and is represented in the model through the capacity shortage cost.

From a managerial perspective, shortages may also lead to practical consequences, such as delayed fulfillment, reduced service reliability, or customer dissatisfaction. However, these consequences are not modeled as separate cost components in this study. Similarly, surplus capacity is represented through a capacity oversupply cost rather than through additional opportunity-cost variables. Therefore, the regret parameters provide a mechanism for translating managerial preferences into the capacity planning model. A higher surplus-regret indicates that the decision maker places greater emphasis on avoiding excessive or unused reserved capacity. Conversely, a higher shortage-regret indicates a stronger preference for avoiding insufficient reserved capacity. Thus, the proposed regret-based formulation enables decision makers to balance the trade-off between cost efficiency and demand fulfillment. Table 6 provides a summary of managerial interpretation.

The proposed models can support capacity-planning decisions under demand uncertainty. However, future research can further enhance their practical applicability. Incorporating scenario probabilities based on historical demand data would enable the models to better reflect the actual likelihood of different demand conditions. Validating the models with actual operational data from an IC design house would also strengthen their empirical relevance and demonstrate their performance under real data structures. In addition, conducting systematic sensitivity analysis of key parameters would help managers assess the robustness of the selected capacity strategies under different operational conditions.

Table 6. Summary of managerial interpretation.

Regret type	Managerial interpretation	Business implication
Capacity surplus regret	Represents the cost consequence of unused or excessive reserved capacity.	Managers should avoid reserving excessive capacity when cost efficiency and capacity utilization are the main priorities.
Capacity shortage cost	Represents the cost consequence of insufficient capacity to satisfy demand requirements.	Managers should reserve sufficient capacity when avoiding a capacity shortage is the main priority.
Production cost	Represents the operational cost of fulfilling production requirements.	Managers should consider how capacity decisions affect total production cost across scenarios.
Transportation cost	Represents the logistics cost associated with capacity allocation and product movement.	Managers should select capacity and outsourcing decisions that also maintain efficient transportation flows.

In practical terms, the proposed approach provides a decision-support framework for tactical and operational planning under demand uncertainty. By adopting a rolling-horizon structure, planning decisions can be periodically updated as a new information becomes available, which reflects the dynamic nature of real-world production and supply chain environments. The framework can therefore support the coordination of production, inventory, capacity, and shortage-related decisions using periodically updated data from enterprise planning systems. However, its practical implementation depends on the availability of operations data and the frequency of replanning.

Besides, this study reports computational performance for the considered instances. The full industrial-scale evaluation may involve hundreds of products and a number of sites. Such large-scale cases would substantially increase the number of decision variables and constraints, potentially affecting computational tractability. Therefore, researchers may test and validate the proposed models using larger instances and/or actual data, and report detailed computational performance statistics, including solution time, optimality gap, number of variables, number of constraints, and memory usage. An inferential statistical test would also strengthen the significance comparison between models once replicate-level observations and larger, real-world data are available. In addition, decomposition methods and scenario-reduction techniques may be explored to improve model scalability and applicability.

Finally, researchers may extend the proposed RHSCCP framework toward strategic interactions among supply chain actors in a rolling-horizon environment or coordinated decision-making among autonomous supply chain actors within a rolling-horizon framework. Game-theoretical approaches could therefore be incorporated to capture strategic interactions as demand information is progressively updated. Several researchers [41–43] have explored cooperative coordination and allocation, strategic supply chain with multiple actors, and dynamic decisions under stochastic uncertainty. Integrating these perspectives with rolling-horizon planning could provide more adaptive, decentralized, and strategically interacting lot-sizing decisions.

6. Conclusion

In conclusion, we develop and compare four RHSCCP models, RHSCCP-MMR, RHSCCP-ULB, RHSCCP-EV, and RHSCCP-SP, under additive and multiplicative MMFE models for an IC design house. The proposed RHSCCP MILP framework integrates network selection, capacity, forecast evolution, and technical constraints to support capacity decisions under continuously updated demand information.

The primary scientific contribution of this study is demonstrating that no single capacity-planning strategy is generally superior; model performance depends on the pattern of demand evolution and the decision-maker's priorities. RHSCCP-ULB provides relatively high fulfillment by reserving additional capacity, but results in lower capacity utilization as CV level increases. RHSCCP-EV maintains relatively stable capacity utilization but may generate lower fulfillment because of its optimistic treatment of uncertainty. Moreover, RHSCCP-MMR provides stronger protection against adverse demand scenarios, whereas RHSCCP-SP balances expected cost, fulfillment, and capacity utilization without excessive capacity reservation.

Furthermore, the experimental results show that the preferred strategy depends on the demand-update procedure. Under additive MMFE, particularly when demand variability is high, RHSCCP-MMR is preferable because of its stronger fulfillment performance and robustness. Under multiplicative MMFE, RHSCCP-SP is preferable because it achieves the lowest cost while maintaining relatively high fulfillment and capacity utilization. From a managerial perspective, these findings suggest that IC design houses should select capacity-planning strategies according to the observed demand-evolution pattern and their priorities regarding cost, service fulfillment, utilization, and protection against uncertainty.

Several limitations should be acknowledged. This study is based on simulated additive and multiplicative MMFE demand patterns, while actual demand may follow more complex dynamics. Scenario probabilities are not estimated directly from historical demand data, and the experiments are limited to the product and operational settings considered in this study. These limitations provide opportunities for future studies.

The following are some research directions for future studies:

(1) Integrate advanced demand-forecasting techniques, including artificial intelligence models, into the capacity planning with rolling-horizon framework.
(2) Develop and compare more dynamic and adaptive capacity-planning model in the rolling-horizon framework by taking into account the unique characteristics of the semiconductor supply chain.
(3) Extend the framework to additional semiconductor products, such as multi-chip-module (MCM) and wafer-level chip-scale packaging (WLCSP).
(4) Derive scenario probabilities from historical demand data and evaluate the models using larger industrial-scale settings.
(5) Validate the proposed models using actual operational data and conduct statistical significance tests and systematic sensitivity analysis of key parameters, including the capacity shortage-to-surplus cost ratio, yield rates, and lead times.
(6) Incorporate game-theoretical approaches to enable more adaptive, decentralized lot-sizing decisions that account for strategic interactions among supply chain actors within the RHSCCP framework.

Overall, this study provides a methodological framework and practical guidance for capacity planning under evolving demand uncertainty. Its key implication is that effective rolling-horizon capacity planning requires matching the planning strategy to the underlying demand-evolution process rather than relying on a single strategy across all operational conditions.

Author contributions

James C. Chen: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – review & editing; **Tzu-Li Chen:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing; **Yin-Yann Chen:** Methodology, Validation, Writing – review & editing; **Chen-Yu Wang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft; **Dewanti Anggrahini:** Validation, Writing – review & editing.

Use of generative-AI tools declaration

The authors declare that they did not utilize any AI tools in the creation of this article.

Acknowledgement

The authors would like to thank the Editorial board, and the anonymous referees for their constructive comments and careful readings that remarkably improved this paper.

Conflict of interest

The authors declare that there is no conflict of interest.

Reference

1. A. Seitz, M. Grunow, R. Akkerman, Data driven supply allocation to individual customers considering forecast bias, *Int. J. Prod. Econ.*, **227** (2020), 107683. https://doi.org/10.1016/j.ijpe.2020.107683
2. Y. Boulaksil, Safety stock placement in supply chains with demand forecast updates, *Oper. Res. Perspect.*, **3** (2016), 27–31. https://doi.org/10.1016/j.orp.2016.07.001
3. S. C. Graves, H. C. Meal, S. Dasu, Y. Qui, Two-stage production planning in a dynamic environment, in *Multi-Stage Production Planning and Inventory Control,* S. Axsäter, C. Schneeweiss, E. Silver (Eds.), Lecture Notes in Economics and Mathematical Systems, **266** (1986), 9–43. https://doi.org/10.1007/978-3-642-51693-1_2
4. D. C. Heath, P. L. Jackson, Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems, *IIE Trans.*, **26** (1994), 17–30. https://doi.org/10.1080/07408179408966604
5. A. P. Rastogi, J. W. Fowler, W. M. Carlyle, O. M. Araz, A. Maltz, B. Buke, Supply network capacity planning for semiconductor manufacturing with uncertain demand and correlation in demand considerations, *Int. J. Prod. Econ.*, **134** (2011), 322–332. https://doi.org/10.1016/J.IJPE.2009.11.006
6. C. Y. Lee, V. Charles, A robust capacity expansion integrating the perspectives of marginal productivity and capacity regret, *Eur. J. Oper. Res.*, **296** (2022), 557–569. https://doi.org/10.1016/j.ejor.2021.04.003
7. J. M. Swaminathan, Tool capacity planning for semiconductor fabrication facilities under demand uncertainty, *Eur. J. Oper. Res.*, **120** (2000), 545–558. https://doi.org/10.1016/S0377-2217(98)00389-0
8. J. M. Swaminathan, Tool procurement planning for wafer fabrication facilities: A scenario-based approach. *IIE Trans.*, **34** (2002), 145–155. https://doi.org/10.1080/07408170208928857
9. S. Karabuk, S. D. Wu, Coordinating strategic capacity planning in the semiconductor industry. *Oper. Res.*, **51** (2003), 839–849. https://doi.org/10.1287/opre.51.6.839.24917
10. N. Geng, Z. Jiang, Capacity planning for semiconductor wafer fabrication with uncertain demand and capacity, in *IEEE Int. Conf. Autom. Sci. Eng.*, **29** (2007). https://doi.org/10.1109/COASE.2007.4341669
11. J. T. Lin, T. L. Chen, H. C. Chu, A stochastic dynamic programming approach for multi-site capacity planning in TFT-LCD manufacturing under demand uncertainty, *Int. J. Prod. Econ.*, **148** (2014), 21–36. https://doi.org/10.1016/j.ijpe.2013.11.003
12. J. Weston, P. Escalona, A. Angulo, R. Stegmaier, Strategic capacity expansion of a multi-item process with technology mixture under demand uncertainty: An Aggregate Robust MILP Approach, in *Proc. 6th Int. Conf. Oper. Res. Enter. Sys. (ICORES 2017), SciTePress,* (2017), 181–191. https://doi.org/10.5220/0006202201810191

13. H. A. Kuo, C. F. Chien, Semiconductor capacity expansion based on forecast evolution and minimax regret strategy for smart production under demand uncertainty, *Comput. Ind. Eng.*, **177** (2023), 109077. https://doi.org/10.1016/j.cie.2023.109077
14. J. T. Lin, Y. Y. Chen, A multi-site supply network planning problem considering variable time buckets–A TFT-LCD industry case, *The Int. J. Adv. Manuf. Technol.*, **33** (2007), 1031–1044. https://doi.org/10.1007/s00170-006-0537-z
15. J. Stray, J. W. Fowler, W. M. Carlyle, A. P. Rastogi, Enterprise-wide semiconductor manufacturing resource planning, *IEEE Trans. Semicond. Manuf.*, **19** (2006), 259–268. https://doi.org/10.1109/TSM.2006.873399
16. B. T. Denton, J. Forrest, R. J. Milne, IBM solves a mixed-integer program to optimize its semiconductor supply chain, *Interfaces*, **36** (2006), 386–399. https://doi.org/10.1287/inte.1060.0238
17. T. Ponsignon, L. Mönch, Heuristic approaches for master planning in semiconductor manufacturing, *Comput. Oper. Res.*, **39** (2012), 479–491. https://doi.org/10.1016/j.cor.2011.06.009
18. T. Ponsignon, L. Mönch, Simulation-based performance assessment of master planning approaches in semiconductor manufacturing, *Omega*, **46** (2014), 21–35. https://doi.org/10.1016/j.omega.2014.01.005
19. J. J. Lowe, S. J. Mason, Integrated semiconductor supply chain production planning, *IEEE Trans. Semicond. Manuf.*, **29** (2016), 116–126. https://doi.org/10.1109/TSM.2016.2544202
20. D. Kim, Y. S. Park, H. W. Kim, K. S. Park, I. K. Moon, Inventory policy for postponement strategy in the semiconductor industry with a die bank, *Simul. Model. Pract. Theory*, **117** (2022), 102498. https://doi.org/10.1016/j.simpat.2022.102498
21. J. C. Chen, T. L. Chen, H. Y. Hu, P. Peng, T. Lin, Multiobjective order promising for outsourcing supply network of IC design houses, *IEEE Trans. Semicond. Manuf.*, **35** (2022), 680–697. https://doi.org/10.1109/TSM.2022.3200128
22. J. C. Chen, T. L. Chen, W. C. Huang, P. Peng, T. Lin, Supply chain planning for IC design house back-end production network with turnkey service, *IEEE Trans. Semicond. Manuf.*, **36** (2023), 458–475. https://doi.org/10.1109/TSM.2023.3281485
23. A. Goli, Efficient optimization of robust project scheduling for industry 4.0: A hybrid approach based on machine learning and meta-heuristic algorithms, *Int. J. Prod. Econ.*, **278** (2024), 109427. https://doi.org/10.1016/j.ijpe.2024.109427
24. A. Ala, A. Goli, Incorporating machine learning and optimization techniques for assigning patients to operating rooms by considering fairness policies, *Eng. Appl. Artif. Intell.*, **136** (2024), 108980. https://doi.org/10.1016/j.engappai.2024.108980
25. A. Goli, Optimization of renewable energy project portfolio selection using hybrid AIS-AFS algorithm in an international case study, *Sci. Rep.,* **14** (2024)*,* 17388. https://doi.org/10.1038/s41598-024-68449-w
26. A. Hosseini, A. Goli, Energy-Aware Dual-Channel agricultural supply chain network design under uncertainty. *Comput. Ind. Eng.*, **207** (2025), 111294. https://doi.org/10.1016/j.cie.2025.111294

27. S. Rahmani, A. Goli, Robust sustainable canola oil-based biodiesel supply chain network design under supply and demand uncertainty, *Environ. Sci. Pollut. Res.,* **30** (2023), 86268–86299. https://doi.org/10.1007/s11356-023-28044-4
28. I. Shahsavani, A. Goli, A systematic literature review of circular supply chain network design: application of optimization models. *Environ. Dev. Sustain.*, **27** (2025), 23483–23514. https://doi.org/10.1007/s10668-023-03362-2
29. A. Goli, E. B. Tirkolaee, A. M. Golmohammadi, Z. Atan, G. W. Weber, S. S. Ali, A robust optimization model to design an IoT-based sustainable supply chain network with flexibility, *Cent. Eur. J. Oper. Res.*, **33** (2025), 1025–1046. https://doi.org/10.1007/s10100-023-00870-4
30. A. Ghodratnama, R. Hassannia, A. Mirzazadeh, A fuzzy goal programming approach to solve a novel supply chain multi-objective mathematical model under uncertainty, *Int. J. Bus. Perform. Supply Chain Model.*, **9** (2017), 280–314. https://doi.org/10.1504/IJBPSCM.2017.091321
31. A. Özmen, E. Kropat, G. W. Weber, Robust optimization in spline regression models for multi-model regulatory networks under polyhedral uncertainty, *Optimization*, **66** (2017), 2135–2155. https://doi.org/10.1080/02331934.2016.1209672
32. F. Yilmaz, B. Karasözen, An all-at-once approach for the optimal control of the unsteady Burgers equation, *J. Comput. Appl. Math.*, **259** (2014), 771–779. https://doi.org/10.1016/j.cam.2013.06.036
33. F. Sahin, E. P. Robinson, Flow coordination and information sharing in supply chains: review, implications, and directions for future research, *Decis. Sci.*, **33** (2002), 505–536. https://doi.org/10.1111/j.1540-5915.2002.tb01654.x
34. P. Brandimarte, Multi-item capacitated lot-sizing with demand uncertainty, *Int. J. Prod. Res.,* **44** (2006), 2997–3022. https://doi.org/10.1080/00207540500435116
35. A. Norouzi, R. Uzsoy, Modeling the evolution of dependency between demands, with application to inventory planning, *IIE Trans.*, **46** (2014), 55–66. https://doi.org/10.1080/0740817X.2013.803637
36. E. Albey, A. Norouzi, K. G. Kempf, R. Uzsoy, Demand modeling with forecast evolution: An application to production planning, *IEEE Trans. Semicond. Manuf.*, **28** (2015), 374–384. https://doi.org/10.1109/TSM.2015.2453792
37. T. Ziarnetzky, L. Mönch, R. Uzsoy, Rolling horizon, multi-product production planning with chance constraints and forecast evolution for wafer fabs, *Int. J. Prod. Res.*, **56** (2018), 6112–6134. https://doi.org/10.1080/00207543.2018.1478461
38. T. Ziarnetzky, L. Mönch, R. Uzsoy, Simulation-based performance assessment of production planning models with safety stock and forecast evolution in semiconductor wafer fabrication, *IEEE Trans. Semicond. Manuf.*, **33** (2020), 1–12. https://doi.org/10.1109/TSM.2019.2958526
39. D. Kopp, L. Mönch, Hierarchical decision-making for qualification management in wafer fabs: A simulation study, *IEEE Trans. Autom. Sci. Eng.*, **20** (2022), 320–333. https://doi.org/10.1109/TASE.2022.3149735
40. A. Forel, M. Grunow, Dynamic stochastic lot sizing with forecast evolution in rolling-horizon planning, *Prod. Oper. Manag.*, **32** (2023), 449–468. https://doi.org/10.1111/poms.13881

41. H. Barman, M. Pervin, S. K. Roy, G. W. Weber, Analysis of a dual-channel green supply chain game-theoretical model under carbon policy, *Int. J. Syst. Sci. Oper. Logist.*, **10** (2023), 2242770. https://doi.org/10.1080/23302674.2023.2242770
42. S. Ergün, P. Usta, S. Z. Alparslan Gök, G. W. Weber, A game theoretical approach to emergency logistics planning in natural disasters, *Ann. Oper. Res.*, **324** (2023), 855–868. https://doi.org/10.1007/S10479-021-04099-9
43. İ. Özcan, S. Z. Alparslan Gök, G. W. Weber, Peer group situations and games with fuzzy uncertainty, *J. Ind. Manag. Optim.*, **20** (2024), 428–438. https://doi.org/10.3934/jimo.2023084