

Production Scheduling for Hybrid Flow Shop Systems with Heterogeneous Parallel Machines and Integrated Work-in-Progress Inventory

Phong-Nhat Nguyen^{1,2}, Truong Pham-Nguyen-Dan^{1,2}, Quyen Le-Thi-Ngoc^{1,2}

¹ Ho Chi Minh City University of Technology, Ho Chi Minh City, Viet Nam

² Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Viet Nam

Abstract

To secure a larger market share in the dynamically evolving personalized market, enterprises must adopt more flexible production modes. One critical challenge in this context is the optimization of production scheduling within hybrid flow shop systems featuring heterogeneous parallel machines. In such systems, machines differ in capabilities, setup requirements, and processing speeds, and not all machines are qualified to process every job - adding complexity to scheduling decisions. This study proposes a multi-objective hybrid flow shop scheduling model that integrates both time and material flow considerations. The model is designed to minimize two key objectives: the minimum of the makespan and the Work-in-Progress (WIP) inventory, which together influence overall system efficiency and responsiveness. By leveraging the strengths of traditional scheduling strategies, the proposed approach supports better planning and execution under increasing demand conditions. A comprehensive scheduling model incorporating time and cost constraints is developed, and numerical experiments are conducted to validate its effectiveness. The results demonstrate that the proposed model significantly improves production efficiency, reduces operational costs, and increases adaptability to market variations. Furthermore, the study provides actionable insights for decision-makers in complex manufacturing environments, offering a scalable framework for dynamic scheduling optimization. These findings contribute to advancing research in production scheduling and support practical applications in industries seeking to enhance competitiveness through agile and cost-effective operations.

Keywords: production scheduling, hybrid flow shop, heterogeneous parallel machines, work-in-progress inventory.

Received on 09 July 2025, accepted on 26 November 2025, published on 27 November 2025

Copyright © 2025 Phong-Nhat Nguyen *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsmre.9693

1. Introduction

In modern manufacturing, companies must deliver high-quality services at optimal costs to maintain competitive advantages. A failure to balance these factors can lead to market exclusion. Production speed and delivery time are critical determinants of product quality and pricing. To ensure profitability and customer satisfaction, companies must complete orders on time, highlighting the essential role of production scheduling and sequencing within production systems.

With the rise of customized ordering systems, firms increasingly rely on flexible and adaptive production environments to meet diverse customer demands. As a result, large-scale production frequently employs hybrid production systems, consisting of multiple machine environments operating in parallel. In industries such as semiconductors and automotive manufacturing, production involves a wide variety of product types, each requiring specific machine setups and parameter configurations. A machine is typically qualified to process only a subset of products within a given setup, further contributing to the

system's complexity. This leads to heterogeneous parallel machine systems, where machines differ in processing capabilities, speeds, and setup requirements. These factors introduce significant challenges in achieving efficient scheduling, workload balancing, and resource allocation.

Optimizing scheduling in heterogeneous parallel machine systems has become a key challenge for enhancing productivity and operational efficiency. This challenge becomes more complex when considering inventory factors and integrating hybrid production strategies into scheduling models. Recent studies indicate that addressing these issues can help balance productivity, cost, and market demand fulfilment. One primary approach to tackling this issue is scheduling within heterogeneous parallel machine systems, where machines differ in processing speed, operational capabilities, or costs. This model is illustrated in Figure 1, where jobs move through multiple processing stages with heterogeneous parallel machines i, k at each stage (i : machine, k : stage), requiring optimal assignment strategies to balance workloads, minimize makespan, and control production costs.

In addition to machine heterogeneity, manufacturers currently face significant cost optimization pressures in scheduling problems. Various cost components are associated with flow shop scheduling, including production costs, inventory costs, setup costs, earliness penalties, tardiness penalties, and delivery costs [1, 2]. Earliness or tardiness penalties arise when jobs are completed before or after the customer-specified due date, whereas delivery costs reflect transportation expenses [3]. Inventory costs can be classified into three categories: raw material inventory, work-in-progress (WIP) inventory, and finished goods inventory.

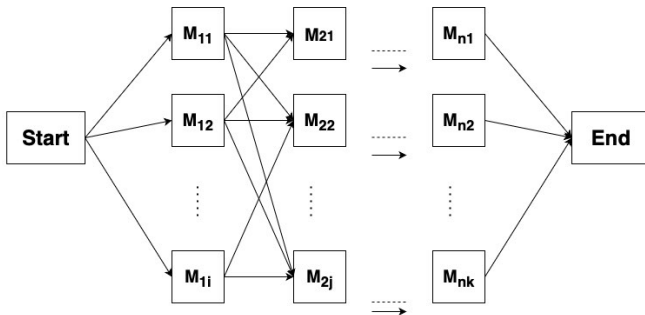


Figure 1. Hybrid Flowshop production line

As implied by their names, raw material and finished goods inventory costs are incurred when purchasing materials and storing completed products, respectively. WIP inventory costs, on the other hand, arise from holding semi-finished products between consecutive processing stages. In hybrid flow shop scheduling, where each machine can process only one job at a time, subsequent jobs must wait in queues until prior operations are completed and machines are reset, leading to increased WIP inventory costs.

Given this context, several integrated cost models have been developed to optimize production schedules by

accounting for multiple cost factors. However, hybrid flow shop scheduling models that explicitly address WIP inventory costs in relation to setup times remain underexplored. This gap is critical because setup time is a major driver of WIP accumulation. Therefore, this study proposes an integrated cost model for hybrid flow shop scheduling that incorporates both WIP inventory costs and setup time dependencies, aiming to enhance schedule efficiency in heterogeneous production environments.

2. Literature Review

In modern manufacturing environments, companies must not only deliver high-quality services at competitive costs but also ensure production progress and delivery times to maintain a market advantage. A lack of attention to these factors can result in a loss of competitiveness and eventual market exclusion. Production speed and service time play a crucial role in determining product quality and pricing. Therefore, production planning and scheduling are essential to ensuring on-time order fulfillment [4]. With increasing market competition and the diversification of customer demands, customized small-batch production models have become more prevalent. However, traditional distribution methods often fail to optimize production schedules as they do not fully consider workshop capacity, leading to infeasible or suboptimal planning for the entire process [5, 6]. To address this limitation, scheduling in Hybrid Shop Scheduling environments has gained significant interest from researchers and enterprises due to its flexibility and optimization capabilities. This model integrates multiple classical scheduling approaches such as flow shop, job shop, parallel machine scheduling, and multiprocessor scheduling [7].

Mozdgir *et al.* [8] investigated scheduling in two-stage assembly flow shop systems, integrating setup time constraints and machine heterogeneity. Their findings highlight the importance of accounting for machine-specific characteristics - such as processing capabilities and setup times - to enhance scheduling efficiency and applicability in real-world production systems. Furthermore, Murrieta-Cortés *et al.* [9] emphasized the importance of minimizing lead time and optimizing resource allocation in flow shops with parallel machines, aiming to improve flexibility and operational effectiveness. Obeid *et al.* [10] contributed by studying job family scheduling on heterogeneous parallel machines, where they explored the trade-offs between scheduling performance and quality constraints. Similarly, Chu *et al.* [11] explored integrated planning and scheduling in hybrid flow shop environments, demonstrating that combining these two processes significantly improves efficiency in make-to-order production. Collectively, these studies establish a critical foundation for addressing the complex challenges posed by machine heterogeneity in multi-stage manufacturing systems.

However, most prior studies consider machine heterogeneity in isolation, often overlooking the

compounded effects of sequence-dependent setup times and the accumulation of work-in-progress (WIP) inventory between stages. Additionally, few address the combined complexity of heterogeneous machine configurations and multi-stage processing within hybrid flow shop environments, particularly from a cost-optimization perspective. This study builds upon existing research by introducing an integrated scheduling model that simultaneously captures machine heterogeneity, setup-related delays, and inter-stage inventory dynamics. The model supports cost-effective job allocation across non-identical machines, offering a more comprehensive and realistic approach to hybrid production scheduling.

In today's competitive landscape, minimizing costs related to production scheduling remains a significant challenge for manufacturers. Major cost components include production costs, inventory costs, setup costs, earliness penalties, tardiness penalties, and delivery costs. Several studies have proposed integrated cost models in multi-machine scheduling systems, particularly in flow shop environments, to optimize production scheduling.

Bülbül *et al.* [1] studied job scheduling problems to minimize tardiness costs, earliness penalties, and intermediate inventory costs in mobile production. They developed an integer programming model and a heuristic search method combining Dantzig-Wolfe Decomposition and Lagrangian Relaxation to optimize costs. Their results showed that this approach provided strong lower bounds for optimal integer solutions and generated near-optimal solutions quickly. Lee *et al.* [12] proposed production control policies based on due dates in semiconductor manufacturing, balancing production goals, due dates, and inventory. Their findings indicate that these models outperform traditional methods.

Wang and Wei [13] examined parallel machine scheduling with deteriorating maintenance, where maintenance time increased if delayed. Their study focused on minimizing the total absolute deviation of completion times and waiting times. The authors demonstrated that this problem could be solved using a polynomial-time approach, offering optimal solutions with reasonable computation time. This work extends previous research by integrating deteriorating maintenance factors into parallel machine scheduling to improve production efficiency. Bozorgirad and Logendran [14] examined integrated flow shop scheduling, where parallel machines at multiple stages had varying processing times. Their objective was to reduce production costs by optimizing inventory levels and order tardiness. Navaei *et al.* [2] addressed a two-stage flow shop problem with heterogeneous assembly machines and setup-dependent constraints, proposing hybrid metaheuristic algorithms to minimize delay and inventory costs. Soltani and Karimi [15] applied a cyclic policy to flow shop systems, reducing WIP inventory through a mixed-integer programming model combined with a Simulated Annealing (SA) algorithm.

While these studies make important contributions, most focus on specific cost types or simplified shop models, and

few explicitly address WIP inventory across multiple stages in hybrid flow shops. In particular, the interaction between setup time and WIP accumulation is often overlooked, despite its significance in modern, high-mix production settings. This research fills that gap by integrating WIP inventory costs and setup time dependencies into a single scheduling model. By doing so, it offers a more realistic and holistic approach to optimizing job flows and reducing intermediate inventory in complex hybrid production environments.

Based on the theoretical foundation of previous studies, this research focuses on solving scheduling problems in hybrid flow-shop environments with heterogeneous parallel machines, a challenging yet highly applicable area. The study aims to minimize work-in-progress (WIP) inventory costs at different stages, thereby optimizing production performance and enhancing system competitiveness. Integrating advanced scheduling methods with inventory management strategies not only addresses real-world scheduling problems but also contributes to refining the theoretical framework for production scheduling in modern industrial contexts.

Author(s) & Year	Hybrid Flow Shop	Parallel Machines	Setup Time
Mozdgir <i>et al.</i> (2013)	✓	✓	✓
Murrieta-Cortés <i>et al.</i> (2012)	✓	✓	
Obeid <i>et al.</i> (2014)		✓	
Chu <i>et al.</i> (2022)	✓	✓	
Bülbül <i>et al.</i> (2004)		✓	
Lee <i>et al.</i> (2008)		✓	
Wang & Wei (2011)		✓	
Bozorgirad & Logendran (2013)	✓	✓	
Navaei <i>et al.</i> (2014)	✓	✓	✓
Soltani & Karimi (2015)	✓	✓	
This study	✓	✓	✓

Author(s) & Year	WIP Inventory	Multi-Objective	Cost Minimization
Mozdgir <i>et al.</i> (2013)			✓
Murrieta-Cortés <i>et al.</i> (2012)		✓	
Obeid <i>et al.</i> (2014)			✓
Chu <i>et al.</i> (2022)		✓	✓

Bülbül <i>et al.</i> (2004)	✓		✓
Lee <i>et al.</i> (2008)	✓		
Wang & Wei (2011)			✓
Bozorgirad & Logendran (2013)			✓
Navaei <i>et al.</i> (2014)	✓		✓
Soltani & Karimi (2015)	✓		
This study	✓	✓	✓

Table 1. Summary matrix of related studies and their addressed characteristics

3. Mathematical Model

This study investigates a production scheduling problem within an industrial bearing manufacturing system comprising four consecutive grinding stages: face grinding, outer diameter (OD) grinding, raceway grinding, and bore grinding. The system features a hybrid flow shop-job shop configuration, where the first and final two stages operate sequentially, while the second stage consists of three parallel intermediate stages. Products follow a flow shop pattern with sequential workstation routing but also exhibit job shop characteristics, diverging to specific machines based on product specifications. Semi-finished products between stages are classified as work-in-process (WIP) inventory, a critical factor influencing system performance.

Key challenges include significant setup times of 1–2 hours during product changeovers, heterogeneous processing times across workcenters due to product-specific requirements, and order-driven production with non-identical batches and volatile demand. These factors collectively contribute to WIP imbalances between parallel production lines, delivery delays, and excessive WIP accumulation, particularly during sudden demand fluctuations. A primary concern for manufacturers is the uncontrolled surge in WIP inventory caused by abrupt demand declines, which disrupts production efficiency and increases holding costs.

The WIP inventory convention in this study is described in Figure 2. The production system includes two machines, M1 and M2, with intermediate inventory points I1 and I2. Under this convention, WIP at each machine is determined based on the total quantity of products awaiting processing or currently being processed at that machine. Specifically, WIP at M1 includes products being processed at M1 and completed products awaiting transfer at I1 before moving to M2.

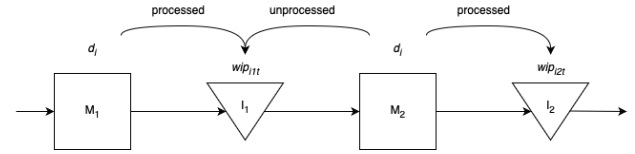


Figure 2. Inventory convention model for the problem

In this section, a Mixed-Integer Linear Programming (MILP) model is proposed for scheduling in a hybrid flow-shop environment with heterogeneous parallel machines, considering work-in-progress (WIP) inventory factors in a multi-product, multi-stage production system. This model is designed to simultaneously address two primary objectives: optimizing job completion time (makespan) and minimizing WIP inventory levels during production.

Notably, this model not only provides solutions for fundamental scheduling problems but also integrates practical factors such as inventory constraints and machine capacity, enabling production managers to develop efficient scheduling plans aligned with business objectives and real-world manufacturing limitations. Additionally, the model's results can support manufacturing enterprises in reducing operational costs, improving order fulfillment capability, and enhancing supply chain sustainability.

Index:

i, j Order index ($i, j = 1, 2, \dots, N$)

t Time period index ($t = 1, 2, \dots, T$)

k Workstation index ($k = 1, 2, \dots, K$)

m Machine index at workstation k ($m = 1, 2, \dots, M_k$)

E_k Set of orders to be processed in workstations k

where $|E_k| = e_k$

Input parameters:

$pt_{i,k,m}$ Processing time of order i at workstation k on machine m

$st_{i,k,m}$ Setup time of order i at workstation k on machine m

d_i Demand of order i

dd_i Delivery time of order i

T_t Time to consider inventory at time point t

i_{max} Maximum inventory cost of WIP

$c_{i,k,m}$ Inventory cost per unit of order i at workstation k on machine m

p_i Penalty cost of order i

Decision variables:

$S_{i,k,m}$ Start time of order i at workstation k on machine m

$x_{ijkm} = \begin{cases} 1, & \text{if order } j \text{ is processed after order } i \\ & \text{at workstation } k \text{ on machine } m \\ 0, & \text{otherwise} \end{cases}$

$y_{i,k,m,t} = \begin{cases} 1, & \text{if order } i \text{ is being processed} \\ & \text{at workstation } k \text{ on machine } m \text{ at time } t \\ 0, & \text{otherwise} \end{cases}$

$$a_{i,k,m} = \begin{cases} 1, & \text{if order } i \text{ is processed at workstation } k \\ & \text{on machine } m \\ 0, & \text{otherwise} \end{cases}$$

$$z_{j,k,m} = \begin{cases} 1, & \text{if order } j \text{ is processed at workstation } k \\ & \text{on machine } m \\ 0, & \text{otherwise} \end{cases}$$

Dependent variables:

$C_{i,m,k}$ Completion time of order i at workstation k machine m

C_i Completion time order i at the final workstation K

C_{max} Total production time for all orders i

$wip_{i,k,m,t}$ Actual inventory level of order i at workstation k on machine m at time period t

Objective function:

$$Min C_{max} \quad (1)$$

$$Min \sum_k \sum_i \sum_t (wip_{i,k,m,t} \times c_{i,k,m}) \quad (2)$$

The first objective function (1) minimizes the makespan while the second objective function (2) minimizes end-of-day inventory costs at each machine, optimizing inventory management.

Constraints (3)–(10) are established to ensure feasibility, production scheduling optimization, and inventory level calculations at workstation k on machine m . This will be divided into 4 types. First type is a lot of constraints on the timing and production order of the machines with four constraints.

$$S_{i,k+1,m} \geq C_{i,k,m} \quad \forall i \in N, \forall i \in \{1, \dots, K-1\}, \forall m \in M_k \quad (3)$$

$$C_{max} \geq C_{i,k,m} \quad \forall i \in N, \forall m \in M_k \quad (4)$$

$$C_{i,k,m} \geq S_{i,k,m} + a_{i,k,m} \times (pt_{i,k,m} \times d_i + st_{i,k,m}) \quad (5)$$

$$\forall i \in N, \forall k \in K, \forall m \in M_k$$

$$C_{i,K,m} < dd_i \quad i \in N, \forall m \in M_k \quad (6)$$

The constraint (3) requires that production of order i on the next machine begin only when order i has been completely completed on the previous machine. With the constraint (4), the total production time for all orders i will have to be greater than the completion time of order i (at the last machine). The paper also gives the formula for calculating the completion time at machines k at constraint (5). Of course, with the completion time of the last machine K in the system having to meet on-time delivery, this requirement falls under constraint (6).

Then is second type, there is the production order between orders i on machines k .

$$S_{j,k,m} \geq C_{i,k,m} - M \times (1 - x_{i,j,k,m}) \quad (7)$$

$$\forall i, j \in N, i \neq j, \forall k \in K, \forall m \in M_k$$

$$S_{j,k,m} \geq C_{i,k,m} - M \times x_{i,j,k,m} \quad (8)$$

$$\forall i, j \in N, i \neq j, \forall k \in K, \forall m \in M_k$$

With M is the large enough constant. Constraint (7) and (8) ensures the validity of the execution order between orders i and j , maintaining a logical processing sequence. If $x_{i,j,k,m} = 1$, it means order i will be produced before order j at workstation k on machine m and vice versa when $x_{i,j,k,m} = 0$.

While the third type is using the HFS model creates moving points that require consideration of options and constraints for smooth management.

$$\sum_{m=1}^M z_{j,k,m} = 1 \quad \forall j \in E_k, \forall k \in K \quad (9)$$

Constraint (9) ensures that orders assignment to machines at each workstation, ensuring a reasonable and feasible allocation of work among parallel machines.

In the Final type, we set a lot of the inventory constraints. The special feature of the model is that it considers the practical problem of inventory. Enterprises prioritize inventory in terms of capital constraints and inventory costs incurred rather than being concerned with the quantity and area of inventory.

$$\sum_k \sum_i \sum_t (wip_{i,k,m,t} \times c_{i,k,m}) \leq i_{max} \quad \forall m \in M \quad (10)$$

$$c_{i,k,m} \geq T_t - M \times (1 - y_{i,k,m,t}) \quad (11)$$

$$\forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T$$

$$s_{i,k,m} \leq T_t + M \times (1 - y_{i,k,m,t}) \quad (12)$$

$$\forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T$$

$$wip_{i,k,m,t} \leq \frac{Tt - S_{i,k,m} - st_{i,k,m}}{pt_{i,k,m}} \times y_{i,k,m,t} \quad (13)$$

$$\forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T$$

$$wip_{i,k-1,m,t} \geq (d_i - wip_{i,k,m,t}) \times y_{i,k,m,t} \quad (14)$$

$$\forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T$$

Constraint (10) relates to inventory costs, aiming to minimize WIP storage costs while maintaining necessary inventory levels to avoid workflow interruptions. Time logic constraints for inventory review (11) and (12) related to order start and end times. If $y_{i,k,m,t} = 1$ then the inventory review time T_t lies between the start and end times of the order at workstation k machine m and vice versa when $y_{i,k,m,t} = 0$. Constraint (13) and (14) determines the actual inventory levels at workstation k on machine m and workstation $k-1$ on machine m at time to consider inventory T_t .

To enhance the practical applicability of the scheduling model, it is essential to convert all objectives into a unified comparable metric, thereby facilitating effective decision-making in real-world industrial settings. Specifically, this study concurrently optimizes two key objectives: the total production completion time and the cost of work-in-progress inventory. In practice, these indicators are measured on different scales and units, so the model proposes a composite objective function with adjustable weights to dynamically balance their trade-offs according to managerial priorities.

Objective function:

$$\text{Min} \sum_i p_i \cdot (c_i - dd_i) \quad (15)$$

$$\text{Min} \sum_k \sum_i \sum_t (wip_{i,k,m,t} \times c_{i,k,m}) \quad (16)$$

Constraints:

$$S_{i,k+1,m} \geq C_{i,k,m} \quad \forall i \in N, \forall i \in \{1, \dots, K-1\}, \forall m \in M_k \quad (17)$$

$$C_{i,k,m} \geq S_{i,k,m} + a_{i,k,m} \times (pt_{i,k,m} \times d_i + st_{i,k,m}) \quad \forall i \in N, \forall k \in K, \forall m \in M_k \quad (18)$$

$$C_i \leq C_{i,K,m} + M \cdot (1 - z_{j,K,m}) \quad \forall i \in N, \forall m \in M_k \quad (19)$$

$$C_i \geq C_{i,K,m} - M \cdot (1 - z_{j,K,m}) \quad \forall i \in N, \forall m \in M_k \quad (20)$$

$$S_{j,k,m} \geq C_{i,k,m} - M \times (1 - x_{i,j,k,m}) \quad \forall i, j \in N, i \neq j, \forall k \in K, \forall m \in M_k \quad (21)$$

$$S_{j,k,m} \geq C_{i,k,m} - M \times x_{i,j,k,m} \quad \forall i, j \in N, i \neq j, \forall k \in K, \forall m \in M_k \quad (22)$$

$$\sum_{m=1}^M z_{j,k,m} = 1 \quad \forall j \in E_k, \forall k \in K \quad (23)$$

$$\sum_k \sum_i \sum_t (wip_{i,k,m,t} \times c_{i,k,m}) \leq i_{\max} \quad \forall m \in M \quad (24)$$

$$c_{i,k,m} \geq T_i - M \times (1 - y_{i,k,m,t}) \quad \forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T \quad (25)$$

$$s_{i,k,m} \leq T_i + M \times (1 - y_{i,k,m,t}) \quad \forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T \quad (26)$$

$$wip_{i,k,m,t} \leq \frac{Tt - S_{i,k,m} - st_{i,k,m}}{pt_{i,k,m}} \times y_{i,k,m,t} \quad \forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T \quad (27)$$

$$wip_{i,k-1,m,t} \geq (d_i - wip_{i,k,m,t}) \times y_{i,k,m,t} \quad \forall i \in N, \forall k \in K, \forall m \in M_k, \forall t \in T \quad (28)$$

The proposed model incorporates a series of constraints designed to ensure feasible, efficient scheduling and accurate inventory management in a hybrid flow shop system with heterogeneous parallel machines. Constraints (15) and (16) define the objective functions: minimizing the penalty cost related to order completion deviations from scheduled delivery times and minimizing total work-in-progress (WIP) inventory holding costs across all products, machines, and time periods. Constraints (17) and (18) ensure that each order begins processing at a workstation only after completion at the previous one and that completion times include processing and setup durations. Constraints (19) and (20) govern the sequencing at the final workstation, ensuring logical order precedence among jobs assigned to the same machine. Constraints (21) and (22) enforce the processing order between pairs of jobs at each machine using binary sequencing variables to prevent overlap. Constraint (23) requires that each order be assigned uniquely to one machine at each workstation. Constraint (24) limits the total WIP inventory costs to remain within predefined maximum values, reflecting operational and capital constraints. Constraints (25) and (26) tie inventory costs to actual processing times via

binary variables indicating when each order is being processed during specific time intervals. Constraints (27) and (28) relate inventory levels dynamically to processing times, setup times, and demands, reflecting the flow and accumulation of WIP between successive stages. Together, these constraints form a comprehensive framework that ensures temporal consistency, resource allocation effectiveness, sequence feasibility, and inventory control, enabling robust and realistic production scheduling aligned with business goals in complex manufacturing environments.

4. Experiment

In this study, an actual simulation model was developed to solve the problem of production scheduling in a mechanical production environment. Specifically, the system is built to simulate the process of planning and coordinating the operation of grinding machines in an industrial bearing production line, as illustrated in Figure 1. The model allows the evaluation of the efficiency of the scheduling algorithm through a small-scale dataset specifically designed for testing purposes.

The problem focuses on optimizing the cost of intermediate inventory (WIP), which arises during the process of unfinished products being temporarily stored in the line. The system consists of four main groups of machines: surface grinding, clear groove grinding, beveled edge grinding, and cylindrical grinding. In particular, a group of machines is configured with two devices operating in parallel. While the overall WIP inventory limit of the system is set at $i_{\max} = 11,000$ units.

The problem is designed to process three orders during a three-day production phase. Detailed delivery schedules, including required output and delivery times, are described in Appendix 4 and Appendix 5, respectively. Each shift in Appendix 6, including overtime, lasts 10 hours per day (equivalent to about 36,000 seconds). WIP inventory is counted at the end of each business day, at times such as dd_i delivery times in Appendix 5.

Each order is assigned a specific production process, which passes through groups of machines along a defined technological route (Appendix 1). The processing time on each type of machine is described in detail in Appendix 2, while the machine installation time before the execution of each order is presented in Appendix 3. The production process is designed according to a hybrid structure, combining the flow shop – with a fixed stage order for a part of the cycle and the job shop – with a flexible route at a number of specialized processing stations (Appendix 1).

The processing time of each order is determined by the total unit operation time (Appendix 2) multiplied by the number of products to be produced, plus the time to install the machine (Appendix 3). Each machine is subject to a different level of inventory cost when storing unfinished products, with details presented in Appendix 7.

The goal of the problem is to simultaneously optimize two quantities: the total production time and the WIP

inventory cost. To integrate these two objectives in a composite goal function, the weights are selected as 15% for time and 85% for inventory costs, respectively, which in turn translates into a single evaluation function that serves the optimization process. The company is looking to reduce the costs of keeping goods in the system so that investment resources can be used continuously instead of spending money on inventory in the system. Besides, on-time delivery is still a factor to consider, but in the short term is still understandable.

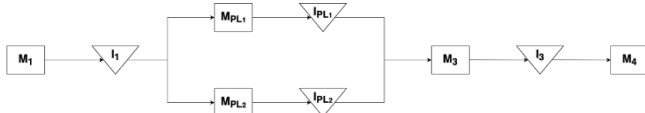


Figure 3. Production line model for the Industrial bearing manufacturing system

A flexible manufacturing system that combines both Flowshop and Jobshop production processes requires careful scheduling to address potential bottlenecks (which can be understood as wait times for orders to arrive at specific machines) or prioritize certain orders (e.g., to reduce inventory time in the process or ensure on-time delivery). In addition to production process considerations, the system also has parallel machines that are not homogeneous across machine groups. Even if there are only two heterogeneous parallel machines in the same group, the system creates a variety of possible scheduling alternatives for machine selection and order processing sequences.

The research model solved the problem by integrating the advantages of the two production processes on the Hybrid Flowshop. Compared to Flow Shop, HFS offers a balanced solution between a clear linear structure and the ability to dynamically allocate resources, which is suitable for more modern production systems. Compared to Job Shop, HFS reduces scheduling complexity but still retains a significant level of flexibility, making it the optimal choice in medium to large-scale production systems. Initially, the problem is solved by determining the order completion time i on machines m , which creates a dataset as shown in Table 1. Based on this time-based data, the study evaluates and removes order sequences that do not meet the expected delivery deadline. This is done by calculating the average processing time for each fleet of machines and then determining the total expected average processing time for the entire production line (illustrated in Table 1).

Table 2. Total production time and setup time of order i at workstation k on machine m (Unit: Seconds)

i/k	1	2	2	3	4
		$m = 1$	$m = 2$		
1	11.500	13.100	14.400	18.000	15.000
2	18.900	24.200	21.300	0	29.100
3	12.600	24.100	23.200	20.800	0

The three orders i can be arranged into six distinct production sequences. The production sequence directly determines both the completion times and delivery schedule of these orders. This sequencing is constrained by each order's due date (dd_i), while the production scheduler may also consider trade-off opportunities. Specifically, the scheduler could intentionally allow certain orders to exceed their due dates within acceptable limits (depending on predefined tolerance thresholds) to achieve system-wide benefits. Such trade-offs may include reducing the total completion time of all orders or decreasing the overall inventory holding costs of the production system. Table 2 below presents all possible production sequences for the three orders i .

Table 3. Order sequence scenarios, completion times, and feasibility evaluation

Scenario	Sequence
1	1 - 2 - 3
2	1 - 3 - 2
3	2 - 1 - 3
4	2 - 3 - 1
5	3 - 1 - 2
6	3 - 2 - 1

In addition to determining the production sequence of orders i , the production scheduler must also select appropriate machines for processing each order. The production line in this problem consists of four machine groups: (1) bore grinding, (2) outer diameter grinding (with two parallel machines PL_1 & PL_2), (3) raceway grinding, and (4) face grinding. Notably, the outer diameter stage features two heterogeneous parallel machines (PL_1 & PL_2), with different processing times and production capacities. This necessitates careful selection of which machine (PL_1 & PL_2), should process each order i . For the three orders and their machine selection at the second stage, the model has generated eight possible machine assignment configurations for each order (presented in Table 3).

Table 4. Machine assignment options for orders i in each major order sequence scenario

Sequence	i/m	1	2	2	3	4
			$m = 1$	$m = 2$		
1	1	x	x		x	x

	2	x		x		x
	3	x		x	x	
2	1	x		x	x	x
	2	x	x			x
	3	x	x		x	
3	1	x	x		x	x
	2	x	x			x
	3	x		x	x	
4	1	x		x	x	x
	2	x		x		x
	3	x	x		x	
5	1	x		x	x	x
	2	x	x			x
	3	x		x	x	
6	1	x	x	x	x	x
	2	x				x
	3	x	x		x	
7	1	x	x		x	x

Table 5. Production time calculation at workstations k for the First sub-scenario of the 1 - 2 - 3 order sequence (Unit: Seconds)

i/m	1	2 m = 1	2 m = 2	3	4
Starting time					
1	0	11500	0	24600	42600
2	11500	0	30400	0	57600
3	30400	0	51700	74900	0
Completion time					
1	11500	24600	0	42600	57600
2	30400	0	51700	0	86700
3	43000	0	74900	95700	0

Table 6. Production time calculation at workstations k for the First sub-scenario of the 1 - 2 - 3 order sequence (Unit: Seconds)

T_t	i/m	1	2 m = 1	2 m = 2	3
36000 (s)	1	0	220	0	280
	2	714	0	186	0
	3	262	0	0	0
72000 (s)	1	0	0	0	0
	2	0	0	525	0
	3	112	0	688	0

Table 7. WIP inventory cost calculation at time period t for workstations k in the First sub-scenario of the 1 - 2 - 3 order sequence (Unit: \$)

	2	x	x			x
	3	x	x		x	
8	1	x		x	x	x
	2	x		x		x
	3	x		x	x	

From the major sequencing scenarios, eight sub-scenarios emerge. For each of these sub-scenarios, it is necessary to calculate the start and completion times of each order. Simultaneously, the work-in-process (WIP) inventory levels at specific time periods and the corresponding WIP inventory costs must also be determined.

Below, the study presents a sample solution for the first sub-scenario of the 1 - 2 - 3 order sequence. The key values to be considered for this case are calculated and shown in tables 5 – 7 about processing time, WIP inventory levels, WIP inventory .

T_t	i/m	1	2 m = 1	2 m = 2	3
36000 (s)	1	0	1760	0	1680
	2	2855	0	1491	0
	3	1569	0	0	0
72000 (s)	1	0	0	0	0
	2	0	0	4200	0
	3	669	0	4131	0

Based on the six possible order sequencing arrangements and eight machine assignment configurations for each order, a total of 48 potential solutions were generated. Among these, only 6 solutions fully satisfy all specified constraints, while the remaining 42 solutions violate at least one constraint. However, as previously noted, the production scheduler may consider certain constraint-violating solutions from these 42 cases to evaluate potential trade-offs. Consequently, the study has selected 8 machine assignment configurations from two order sequencing scenarios (1-2-3 and 1-3-2) for further analysis.

From the sample calculations of the first sub-scenario in the 1 - 2 - 3 order sequence, similar calculations are performed for the remaining sub-scenarios in the model. The aggregated results of these cases are presented in Table 8. Additionally, the final evaluation score is provided by applying the objective function weights specified in the problem. Furthermore, production schedulers can enhance the analysis by considering values such as the average (AVERAGE) and minimum (MIN) inventory and production costs, allowing for a more refined trade-off decision-making process.

Table 8. Summary of total production time for orders i in the production line and inventory costs at evaluated time periods, including final scoring for two objectives

Sequence for order	Sequence for product station	C_{max} (s) (a)	wip_t (t = 36000s)	wip_t (t = 72000s)	TOTAL wip_t (b)	RATE (15%a + 85%b)
1 - 2 - 3	1	95700	9355	9000	18355	299568
	2	99500	<u>8637</u>	8248	16885	292773
	3	<u>87000</u>	9077	7766	16843	273666
	4	88000	8914	<u>7458</u>	<u>16372</u>	<u>271162</u>
	5	88000	<u>8637</u>	8091	16728	274188
	6	87900	9354	7258	16612	273052
	7	99500	9077	7923	17000	293750
	8	95700	8915	9371	18286	298981
1 - 3 - 2	1	97200	11568	4564	16132	282922
	2	<u>98100</u>	<u>10711</u>	<u>4022</u>	<u>14733</u>	<u>272381</u>
	3	98600	11096	4864	15960	283560
	4	99000	10851	4500	15351	278984
	5	<u>98100</u>	<u>10711</u>	5400	16111	284094
	6	97200	11568	4500	16068	282378
	7	98600	11095	4152	15247	277500
	8	99000	10851	5400	16251	286634
AVERAGE		95444	10026	6407	16433	282849
MIN		87000	8637	4022	14733	271162

This table provides the production time data for 3 orders i of 16 cases (column a), the inventory value at 2 points in time (36000s and 72000s in the next 2 columns) added in column b. Then there is the processing of multiplying the weights to produce the result of the target function. In addition to comparing the values of the options with each other, the problem will calculate the average and smallest values of 16 cases. That helps dispatchers see the most feasible options when comparing multiple targets.

It is evident that there are two key differences between the two major sequencing scenarios. As mentioned earlier, the 1 - 2 - 3 sequence results in a delayed delivery for order $i = 3$. However, considering the study's evaluation criteria, this scenario remains acceptable as the model produces favorable results. The set of sub-scenarios under the 1 - 2 - 3 sequence has a lower evaluation score compared to those under the 1 - 3 - 2 sequence.

Overall, the 1 - 2 - 3 sequence group exhibits shorter production times but higher WIP inventory costs, along with the drawback of delayed order delivery. Conversely, the 1 - 3 - 2 sequence group results in longer production times but lower WIP inventory costs.

The two most effective cases identified by the model for both sequencing scenarios are:

- The fourth sub-scenario of the 1 - 2 - 3 sequence, which accepts a short delay for order $i = 3$
- The second sub-scenario of the 1 - 3 - 2 sequence, which meets all delivery deadlines but has a higher evaluation score.

5. Conclusion

This study evaluates a modest simulation problem involving three orders i , four machine groups k , and only two heterogeneous parallel machines in the second machine group. Despite these limited conditions, the model generates up to 48 possible scenarios for evaluation. This highlights the complexity of scheduling in flexible production environments, including Flow Shop, Hybrid Flow Shop, and Job Shop systems with heterogeneous parallel machines.

Decision-making for selecting order sequencing and machine scheduling involves multiple criteria and varying priority weights, depending on managerial objectives. The proposed model effectively computes and captures all possible scenarios, ensuring that decision-makers have comprehensive data to facilitate quick and efficient comparisons when selecting the optimal scheduling approach.

6. Future Research

In future studies, the authors aim to examine production quantity decisions for each machine at different time periods, while still accounting for inventory constraints as a key factor. The HFS models will be integrated to reassess production decisions, rather than simply setting production quantities equal to forecasted demand. Implementing Push and Pull policies at different stages while determining the

optimal production quantity for each machine in each time period is expected to enhance scheduling efficiency.

With the flexibility of HFS, and the scheduling model presented in this study, future research will introduce additional sub-scenarios within the existing framework. A more detailed evaluation of production quantity allocation and processing time will contribute to a more effective scheduling process and a clearer understanding of WIP inventory management in response to the risks associated with costs and time delays.

Acknowledgements

This research is funded by Ho Chi Minh City University of Technology (HCMUT), VNU-HCM under grant number SVKSTN-2025-CK-45. We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.

References

- [1] Bülbül K, Kaminsky P, Yano C. Flow shop scheduling with earliness, tardiness, and intermediate inventory holding costs. *Naval Research Logistics*. 2004;51(3):407-45.
- [2] Navaei J, Ghomi SMTF, Jolai F, Mozdgir A. Heuristics for an assembly flow-shop with non-identical assembly machines and sequence dependent setup times to minimize sum of holding and delay costs. *Computers Operations Research*. 2014;44:52-65.
- [3] Federgruen A, Mosheiov G. Heuristics for multimachine scheduling problems with earliness and tardiness costs. *Management Science*. 1996;42(11):1544-55.
- [4] Ramezani R, Fallah Sanami S, Shafiei Nikabadi M. A simultaneous planning of production and scheduling operations in flexible flow shops: case study of tile industry. *The International Journal of Advanced Manufacturing Technology*. 2017;88:2389-403.
- [5] Aouam T, Geryl K, Kumar K, Brahimi N. Production planning with order acceptance and demand uncertainty. *Computers operations research*. 2018;91:145-59.
- [6] Tan W, Khoshnevis B. Integration of process planning and scheduling—a review. *Journal of Intelligent Manufacturing*. 2000;11:51-63.
- [7] Fan K, Zhai Y, Li X, Wang M. Review and classification of hybrid shop scheduling. *Production Engineering*. 2018;12:597-609.
- [8] Mozdgir A, Fatemi Ghomi S, Jolai F, Navaei J. Two-stage assembly flow-shop scheduling problem with non-identical assembly machines considering setup times. *International Journal of Production Research*. 2013;51(12):3625-42.
- [9] Cortés BM, García JCE, Hernández FR. Multi-objective flow-shop scheduling with parallel machines. *International journal of production research*. 2012;50(10):2796-808.
- [10] Obeid A, Dauzère-Pérès S, Yugma C. Scheduling job families on non-identical parallel machines with time constraints. *Annals of Operations Research*. 2014;213(1):221-34.
- [11] Chu H, Dong K, Li R, Cheng Q, Zhang C, Huang K, et al. Integrated modeling and optimization of production planning and scheduling in hybrid flow shop for order production mode. *Computers Industrial Engineering*. 2022;174:108741.
- [12] Lee B, Lee Y, Yang T, Ignisio J. A due-date based production control policy using WIP balance for implementation in semiconductor fabrications. *International Journal of Production Research*. 2008;46(20):5515-29.
- [13] Wang J-B, Wei C-M. Parallel machine scheduling with a deteriorating maintenance activity and total absolute differences penalties. *Applied Mathematics Computation*. 2011;217(20):8093-9.
- [14] Bozorgirad MA, Logendran R. Bi-criteria group scheduling in hybrid flowsheets. *International Journal of Production Economics*. 2013;145(2):599-612.
- [15] Soltani SA, Karimi B. Cyclic hybrid flow shop scheduling problem with limited buffers and machine eligibility constraints. *The International Journal of Advanced Manufacturing Technology*. 2015;76:1739-55.

Appendix

A.1. Technological process sequence of orders

	k	1	2	3	4
i	1	1	1	1	1
	2	1	1	0	1
	3	1	1	1	0

A.2. Processing time at machines for each order (Unit: Seconds)

pt_{ik}	k	1	2	2	3	4
			$m=1$	$m=2$		
i	1	15	22	24	30	25
	2	18	25	22	0	28
	3	13	27	26	22	0

A.3. Setup time at machines for each order (Unit: Seconds)

st_{ik}	k	1	2	2	3	4
			$m=1$	$m=2$		
i	1	4000	2100	2400	3000	2500
	2	2700	1700	1500	0	3900
	3	2200	2500	2400	3200	0

A.4. Forecasted order demand over time (in this problem, the forecasted quantity is considered the production quantity) (Unit: Units)

i	dd_1	dd_2	dd_3
1	500	0	0
2	0	0	900
3	0	800	0

A.5. Delivery time of orders (Unit: Seconds)

i	1	2	3
dd_i	36000	108000	72000

A.6. Time to consider inventory (Unit: Seconds)

t	1	2
T_1	36000	72000

A.7. Inventory cost per unit of order i at each machine (Unit: \$)

c_{ik}	k	1	2	2	3	4
			$m=1$	$m=2$		
i	1	2	8	6	6	1
	2	4	7	8	0	4
	3	6	5	6	4	0