

Article

Schedule-Informed Empty-Vehicle Dispatching for OHT Systems in Semiconductor Manufacturing

Jungwoo Ahn, Jihoon Bae , Hyojin Ahn, Seongjun Yu , In-Beom Park * and Kwanho Kim *

Department of Industrial and Systems Engineering, Dongguk University, Seoul 04620, Republic of Korea; woorabbit@dgu.ac.kr (J.A.); 2025126804@dgu.ac.kr (J.B.); 2026120264@dgu.ac.kr (H.A.); usj010426@dgu.ac.kr (S.Y.)

* Correspondence: ibpark@dgu.ac.kr (I.-B.P.); kahkim@dgu.ac.kr (K.K.)

Highlights

Please indicate how your work links to systems science via your contributions to systems practice, theory, and/or methodology.

- SOMOS provides an event-driven systems framework that couples semiconductor production execution with OHT transportation, allowing transport outcomes to propagate into subsequent manufacturing operations.
- SEVD introduces an execution-level spatial control mechanism that integrates realized production state and schedule-derived demand information while retaining conventional request assignment.

What are the main findings and/or the implications of the main findings?

- Proactive idle-vehicle repositioning reduces the post-assignment empty-approach travel and upper-tail pickup waiting time relative to conventional HUNGARIAN assignment while maintaining competitive production-level performance.
- Ablation and temporal-window analyses show that the clearest benefit arises from pre-request spatial repositioning itself rather than from precise temporal schedule prediction or any single demand component.

Abstract

Overhead hoist transport (OHT) performance affects the execution of semiconductor production schedules because transport delays postpone subsequent operations. This study develops the Scheduler and OHT Allocation Monitoring Simulator (SOMOS), an event-driven environment that integrates machine processing, transport request generation, OHT movement, and subsequent production execution. Based on SOMOS, we propose Schedule-informed Empty-Vehicle Dispatching (SEVD), which repositions idle OHT vehicles before pickup requests occur using immediate pickup demand, remaining processing information, and a schedule-derived spatial demand prior. SEVD retains the conventional HUNGARIAN assignment procedure for actual requests. Experiments compare SEVD with six benchmark dispatching methods over 20 paired seeds. Relative to HUNGARIAN assignment, SEVD reduces the mean post-assignment empty-approach travel from 96.82 s to 71.23 s and source wait p90 from 428.74 s to 399.34 s while maintaining production-level performance comparable to the strongest benchmarks. These gains are obtained by redistributing vehicle movement toward proactive repositioning, with total empty travel increasing by 17.8% and total travel distance by 6.8%. Fleet size and ablation experiments further show that the clearest benefit is associated with proactive spatial repositioning rather than precise temporal prediction. These results support interpreting SEVD as a framework for the proactive spatial repositioning of idle vehicles, while the incremen-



Academic Editors: Erik Kučera, Štefan Kozák, Oto Haffner and Lukáš Beňo

Received: 6 September 2026

Revised: 8 October 2026

Accepted: 16 October 2026

Published: 22 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

tal contribution of schedule-derived information was not statistically detectable in the present testbed.

Keywords: overhead hoist transport; empty-vehicle dispatching; semiconductor fabrication simulation

1. Introduction

Overhead hoist transport (OHT) is an automated material-handling system that transports lots between production machines along ceiling-mounted rails in semiconductor fabrication facilities [1–3]. In a semiconductor FAB with a complex re-entrant process flow, each lot must be transferred to its next assigned machine after completing the current operation [4,5]. A machine schedule determines the processing sequence and assigned machine for each lot, but it does not guarantee arrival at the next machine at the planned time [6,7]. If an empty OHT vehicle reaches the source machine late after processing is completed, the lot remains waiting for pickup and the subsequent operation is delayed [7,8]. Such transport delays contribute to machine idle time and deviations between actual production progress and the planned schedule [7–9]. Empty-vehicle dispatching is therefore closely related to the execution of the production schedule [7,8].

The performance of empty-vehicle dispatching depends not only on the number of OHT vehicles but also on their spatial dispatching within the FAB [3,10,11]. Even under the same fleet size, transport performance differs when idle vehicles are unevenly distributed across bays or production areas [3]. A pickup request generated in an area with limited idle-vehicle availability requires a longer empty trip before transportation begins [3,12]. This issue is particularly relevant in semiconductor manufacturing because transport requests arise from the progress and completion of machine operations [7,8]. The spatial availability of idle vehicles therefore affects the response to transport demand generated by the production flow [3,13].

Previous studies on OHT and AGV systems have investigated vehicle dispatching, routing, and idle-vehicle management using rule-based, optimization-based, and learning-based approaches [14–16]. Semiconductor scheduling studies have further shown that transportation delay and AMHS capability interact with machine utilization and production performance under re-entrant and uncertain production conditions [8], while related studies have examined production scheduling, lot dispatching, release control, machine allocation, cycle time, and WIP [4,17–21]. Discrete-event simulation has also been widely adopted to evaluate semiconductor AMHS under different layouts, fleet sizes, and operational policies [22–26]. However, relatively limited attention has been given to the use of production information for managing the spatial availability of idle OHT vehicles before transport requests are generated [3,27–30]. Transportation performance and its subsequent effect on production execution are also often considered at different decision levels [7,8,22].

Evaluating this relationship requires a simulation environment that explicitly links production operations with OHT transportation [2,7,8,22]. Machine completion generates transport demand, while transportation determines when the subsequent operation can begin [7,8]. Processing-time variation further changes the timing of transport requests and subsequent operations [8,31], so production execution and OHT transportation evolve through an interacting event-driven process [7,22]. To represent this interaction, this study develops the Scheduler and OHT Allocation Monitoring Simulator (SOMOS), which integrates re-entrant production execution, processing-time variation, transport request generation, vehicle assignment, OHT movement, and subsequent machine operations within a common simulation environment. SOMOS therefore enables OHT operating

strategies to be evaluated under identical production conditions using both transportation-level and production-level performance measures.

Based on SOMOS, this study evaluates representative rule-based dispatching methods, HUNGARIAN-based vehicle assignment [15,32], and the proposed Schedule-informed Empty-Vehicle Dispatching (SEVD). SEVD retains the HUNGARIAN assignment procedure for actual transport requests while additionally managing the spatial distribution of idle vehicles before subsequent requests occur. SEVD neither predicts the exact occurrence time of individual future pickup requests nor reserves specific vehicles for future tasks. Instead, SEVD combines the current production state with spatial demand information derived from the production schedule. Here, the remaining processing time provides a short-term signal from realized production execution, whereas the schedule-derived term serves as an auxiliary spatial prior rather than a predictor of exact request timing. Immediate pickup demand, operations approaching completion, and the planned spatial distribution of transport demand are integrated to estimate transport demand across the FAB. Idle vehicles are then repositioned toward areas with higher estimated demand while remaining available for normal vehicle assignment.

This study makes three main contributions. First, SOMOS provides an event-driven simulation environment that explicitly connects realized machine processing, transport request generation, OHT dispatching and movement, and the start of subsequent production operations. This structure enables transportation decisions to be evaluated according to their effects on actual production execution rather than through an isolated transport model. Second, SEVD provides a production-informed framework for proactive idle-vehicle repositioning that combines immediate pickup demand, realized processing progress, and a schedule-derived spatial prior while retaining the conventional HUNGARIAN assignment procedure for actual requests. The empirical contribution demonstrated in this study is the transportation-level benefit of adding pre-request repositioning to request-level assignment. The present study does not claim that the specific SEVD demand formulation is superior to alternative proactive repositioning strategies. Third, the proposed framework is evaluated through paired multi-seed experiments using both transportation- and production-level measures. The experiments examine the effects of SEVD on empty travel and pickup waiting, the variation in these effects across OHT fleet sizes, and the contribution of individual SEVD components through ablation analysis.

The organization of this paper is as follows: Section 2 reviews the related literature. The SOMOS and the event-driven interaction between semiconductor production execution and OHT transportation are discussed in Section 3. Section 4 presents the proposed SEVD, including spatial transport demand estimation and idle-vehicle dispatching. Section 5 describes the experimental settings and evaluates the proposed approach through comparative experiments, fleet size analysis, and ablation studies. Finally, the conclusions of this study and suggestions for future research are summarized in Section 6.

2. Literature Review

Vehicle dispatching, routing, and idle-vehicle positioning constitute major operational decisions in automated transport systems [1,2,33]. General AGV research distinguishes vehicle scheduling and dispatching, path planning, and traffic control as related but separate control problems [33]. Semiconductor-specific studies similarly identify dispatching, routing, vehicle allocation, and idle-vehicle management as important determinants of wafer-transport performance [1,2]. Early OHT dispatching studies mainly focused on request-level vehicle selection [13,33]. Dispatching rules were developed to reduce delivery time and vehicle interference by considering not only pickup distance but also interactions among vehicles [34]. Other studies extended this perspective to dynamic vehicle allocation,

adjusting the number or dispatching of available vehicles according to changing transport demand across the FAB [11,13,35]. These approaches differ from conventional dispatching in that they regulate vehicle availability at an area level rather than selecting a vehicle only after an individual request occurs [3,13,36]. More recent studies have incorporated real-time traffic information and learning-based methods [16,37,38]. Q-learning has been applied to joint vehicle assignment and route guidance [16], while $Q(\lambda)$ -learning and graph-based prediction models have been used for dynamic routing under congestion [37,38]. Recent studies have further extended OHT research through learning-based control and simulation-based system analysis. Chou et al. applied multiagent reinforcement learning to dynamic OHT dispatching under varying transport demand, while Sung et al. developed a discrete-event simulation framework to evaluate OHT system configurations and semiconductor logistics performance [39,40]. Li and Jin proposed a learning-based multi-objective dispatching framework for inter-bay AMHS that jointly considers transportation and production efficiency, while Koo et al. developed a reinforcement-learning-based dynamic link weight control method to alleviate OHT congestion in semiconductor FABs [41,42]. Overall, the literature has progressed from distance- or rule-based vehicle selection toward dynamic and traffic-responsive control. However, most approaches primarily rely on current transport requests, vehicle states, or network traffic conditions, whereas the spatial positioning of idle vehicles based on upcoming production demand has received comparatively less attention [3,12,27–30].

Production scheduling and material handling are strongly coupled because transportation determines when a completed lot becomes available for its subsequent operation [7,8,17]. Integrated scheduling studies have increasingly considered transportation operations together with production decisions [7]. Drießel and Mönch integrate machine scheduling and AMHS transport operations within an extended shifting bottleneck framework, so material-handling decisions are incorporated into the production scheduling process [7]. In contrast, SEVD takes the production schedule as given and operates at the execution level by repositioning idle vehicles without modifying the production schedule or the assignment procedure for actual transport requests. In semiconductor manufacturing, Kim et al. [8] incorporated AMHS capability into dynamic machine targeting and modeled re-entrant process flows together with processing and transportation time variations. Semiconductor scheduling studies likewise identify lot dispatching, release control, batching, and machine allocation as interdependent shop-floor decisions [17,43]. This interaction becomes particularly important in semiconductor FABs because of re-entrant flows and highly variable operating conditions [4,8,31]. Fluctuation-smoothing approaches seek to control competition among lots and reduce cycle-time variability [19], while other semiconductor scheduling studies show that local resource decisions can propagate delays throughout re-entrant production systems [4,20]. Transportation delay can similarly affect downstream machine availability because delayed movement changes the arrival time of lots at subsequent resources [7,8]. Production-control policies also influence the workload presented to the material-handling system [25]. Semiconductor workload-control studies examine dispatching and order release as mechanisms for regulating WIP and cycle time [18], while comparisons of push-and-pull systems emphasize the importance of WIP control in determining system behavior [21]. Thus, existing studies demonstrate substantial interaction between production control and material handling, although the two have often been optimized or evaluated at different decision levels [7,8].

Discrete event simulation is widely used to evaluate semiconductor AMHSs because production flow, vehicle movement, routing, blocking, and dispatching interact dynamically [22,23,25]. Full-FAB simulation models have incorporated manufacturing tools and material-handling constraints within a common environment [22]. Two-step simulation

has also been proposed to reduce computational effort while preserving the influence of manufacturing flow on AMHS performance [25]. Other studies have used metamodeling and simulation experiments to evaluate material handling performance with lower computational cost [11,24,26]. Reviews of a 300 mm semiconductor AMHS also identify simulation as a major method for evaluating layout, fleet size, dispatching, routing, and overall transport performance [2,23]. These studies demonstrate the importance of simulation for examining interactions that are difficult to isolate analytically. Wan and Shin use the expected arrival times of future jobs and the expected release times of occupied vehicles to support predictive vehicle assignment [27]. Benzoni et al. similarly incorporate estimated upcoming transport arrivals and vehicle availability into a look-ahead assignment framework [28]. In contrast, SEVD does not predict or reserve individual future transport requests. Instead, it uses production-side information to redistribute idle vehicles spatially before requests occur, while actual transport requests are still assigned through the conventional HUNGARIAN procedure. Previous studies have also examined this interaction from the perspective of manufacturing automation. Tyan et al. integrated tool and vehicle dispatching in a fully automated semiconductor FAB, while Koo et al. incorporated production-side bottleneck information into vehicle dispatching decisions. More broadly, simulation has been recognized as an important technology for evaluating alternative product, process, and manufacturing-system configurations in digital manufacturing [44–46].

Table 1 compares the related research streams in terms of pre-request vehicle positioning, the use of production-schedule information, and production-level evaluation. Previous studies have addressed these aspects from different perspectives, including request-level dispatching, dynamic vehicle allocation, predictive vehicle management, production-transport integration, and AMHS simulation. However, the combination of production-schedule information with pre-request idle-vehicle positioning and subsequent production-level evaluation remains comparatively limited. This study addresses this gap by using SEVD to adjust idle-vehicle dispatching before request generation and SOMOS to evaluate its effects on both transportation and production execution.

Table 1. Overview of previous research related to OHT.

Research Stream	Ref	Main Information	Pre-Request Positioning	Production Schedule	Production-Level Evaluation
Dynamic vehicle allocation	[3,11,13,35,36]	Regional workload	✓	△	△
Production-transport integration	[7,8,17]	Production state	–	✓	✓
Request-level dispatching	[14,15,34,35]	Request; vehicle state	–	–	–
Learning-based control	[16,37,38]	Vehicle; traffic state	–	–	–
AMHS simulation	[22–26]	Layout; fleet; transport state	△	△	✓
Predictive/look-ahead vehicle management	[27–30]	Future transport information	✓	△	△
This study	SOMOS + SEVD	Production state; schedule demand	✓	✓	✓

✓, explicitly considered; △, partially considered or dependent on the study; –, not a primary focus.

3. Scheduler and OHT Allocation Monitoring Simulator

To evaluate the interaction between semiconductor production execution and OHT transportation, this study develops the SOMOS. SOMOS integrates machine processing, transport request generation, OHT dispatching, vehicle movement, and subsequent ma-

chine operations within a common event-driven simulation environment. The key role of SOMOS is to propagate realized transportation outcomes into production execution. A transport request is generated when an operation is actually completed, and the subsequent operation starts only after the lot is delivered to its assigned machine. Transportation delay therefore changes the realized production trajectory rather than being treated as a fixed inter-operation delay. Let the set of manufacturing machines be denoted by Equation (1):

$$M = m_1, m_2, \dots, m_i \quad (1)$$

the set of OHT vehicles defined as Equation (2):

$$V = v_1, v_2, \dots, v_j \quad (2)$$

and the set of bays by Equation (3):

$$B = b_1, b_2, \dots, b_k \quad (3)$$

The semiconductor fabrication schedule specifies the processing sequence, assigned machine, and planned timing of each lot. For lot ℓ , its sequence of manufacturing operations is represented as Equation (4):

$$O_\ell = (o_{\ell,1}, o_{\ell,2}, \dots, o_{\ell,q_\ell}) \quad (4)$$

where each operation is associated with its assigned machine and scheduled processing information. The OHT simulation represents the rail network, machine locations, bay configuration, and the state and position of each OHT vehicle.

3.1. Integrated Simulation Architecture

Figure 1 presents the overall architecture of SOMOS, in which semiconductor production execution and OHT transportation are integrated within a common simulation environment. The production model determines the processing sequence and realized operation timing of wafer lots, while the OHT model represents the rail network, vehicle states, and physical transportation between consecutive operations. By connecting these two components, SOMOS allows transportation outcomes to directly affect the subsequent execution of the production schedule rather than treating transportation as a predetermined inter-operation delay.

The production model of SOMOS was developed with reference to the semiconductor FAB testbed reported by Kim et al. [8]. Their testbed consists of 100 production machines belonging to six machine types and represents a re-entrant production structure in which wafer lots repeatedly revisit machine types across multiple process layers. The testbed also incorporates processing-time variability through truncated normal distributions. SOMOS uses this production structure as the manufacturing backbone and integrates it with explicit OHT vehicle movement and dispatching. The production schedule provides the planned sequence of operations, assigned machines, and corresponding processing information, whereas the OHT model determines the actual transportation required between consecutive operations.

Processing times during simulation are not restricted to their nominal scheduled values. Let $p_{l,r}^{plan}$ denote the planned processing time of operation r for lot l , and let $p_{l,r}^{real}$ denote its realized processing time generated by the processing-time variation model. Because $p_{l,r}^{real}$ varies during simulation, the actual completion time of an operation can deviate from its planned completion time. This distinction is important because an OHT

transport request is generated according to the actual completion event, rather than the originally planned completion time.

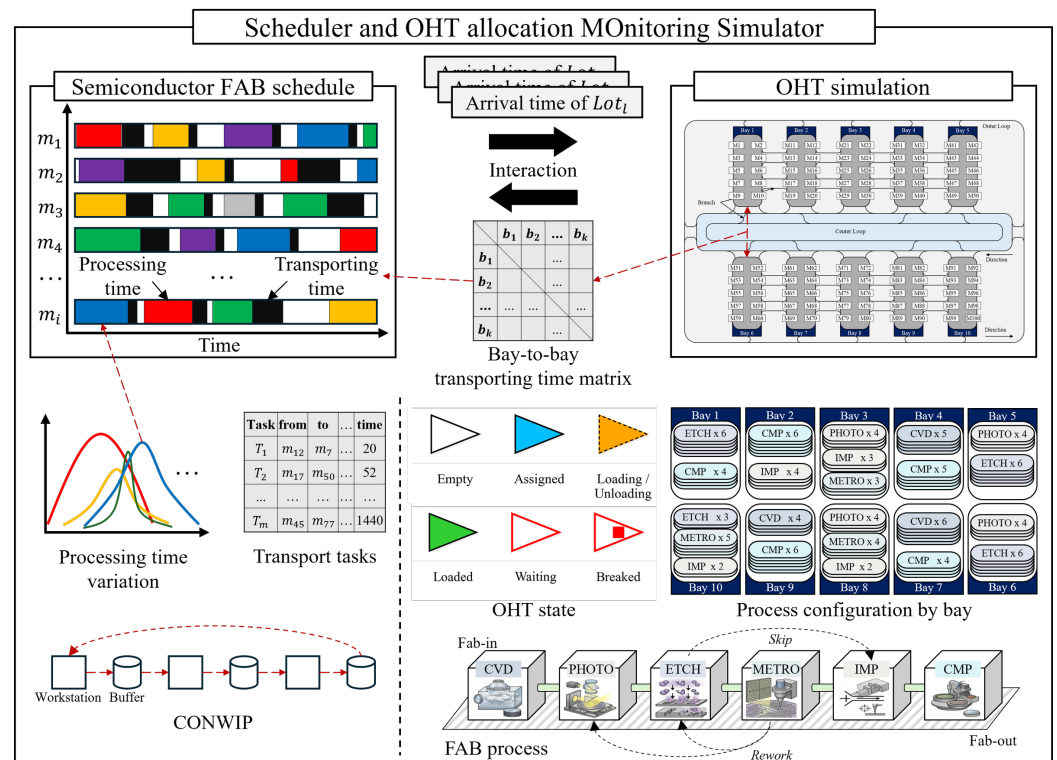


Figure 1. Overall architecture of SOMOS integrating semiconductor FAB scheduling and OHT simulation.

When operation r of lot l is completed at machine M_i and the subsequent operation is assigned to machine M_j , the corresponding transport request is defined by the source machine M_i , the destination machine M_j , and the request-generation time determined by the realized completion event. The OHT dispatching module then assigns an eligible vehicle according to the configured dispatching method. SOMOS continuously maintains the dynamic state of the transportation system, including vehicle locations and operating states such as idle, empty travel, loading, loaded travel, unloading, and waiting. Through this architecture, the planned production schedule provides the reference production sequence, while realized processing and transportation determine the actual execution trajectory. The event-driven interaction among these components is described in Section 3.2.

3.2. Event-Driven Interaction Between Production and OHT Operations

Figure 2 illustrates the event-driven interaction between production execution and OHT transportation in SOMOS. Each operation first undergoes realized processing according to the processing-time variation described in Section 3.1. When the operation is completed, a pickup request is generated for transporting the lot to the machine assigned to its subsequent operation. The OHT dispatching module then selects an eligible empty vehicle according to the configured dispatching method. The assigned vehicle travels to the source machine, loads the lot, and transports it to the destination machine. The subsequent production operation starts only after transportation is completed and the lot becomes available at the destination. This sequence is repeated throughout the simulation as lots progress through their manufacturing routes.

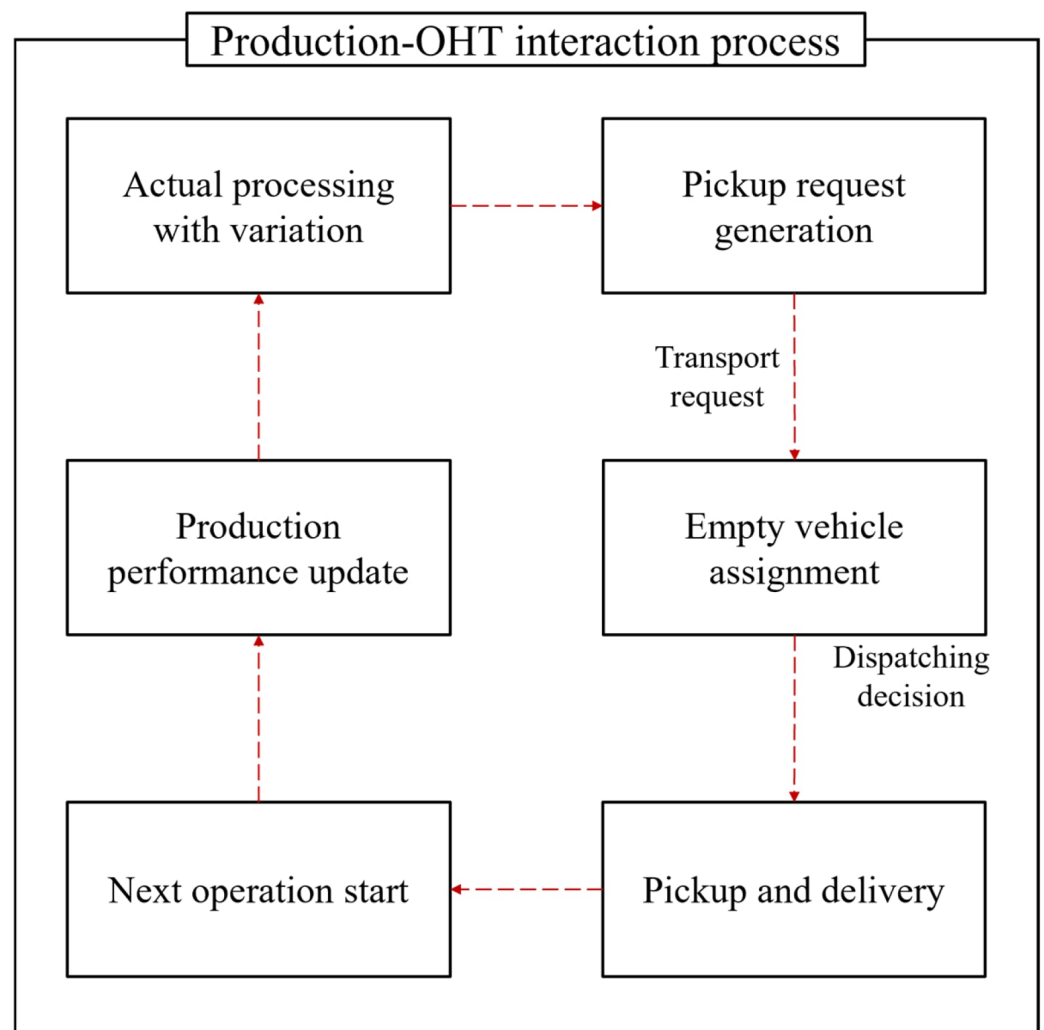


Figure 2. Event-driven interaction between production operations and OHT transportation in SOMOS.

This event structure creates a direct interaction between production variability and OHT operation. A change in realized processing time shifts the occurrence time of the corresponding transport request. The transportation state at that time, including vehicle location and availability, subsequently determines the pickup and delivery process experienced by the completed lot. Any resulting transportation delay changes the time at which the lot becomes available for its next operation and therefore propagates into the subsequent production trajectory. SOMOS consequently records transportation and production performance from the same sequence of simulation events rather than evaluating transportation as an independent post-processing step. In particular, the simulator records the empty travel required to reach a source machine and the time for which a completed lot remains at its source while waiting for pickup, which are used as transportation performance measures in Section 5.

Figure 3 presents the integrated visualization interface of SOMOS for observing this interaction during simulation. The OHT rail network and machine Gantt chart are synchronized according to the same simulation clock. The rail view displays the locations and movement states of OHT vehicles, while the Gantt chart represents the corresponding execution state of production operations. System indicators such as executed operations, completed lots, and active OHT movements are updated on the same time basis. The interface therefore enables the spatial state of OHT transportation and the progress of production execution to be examined at the same simulation time, providing a system-level view of how the two processes evolve together.

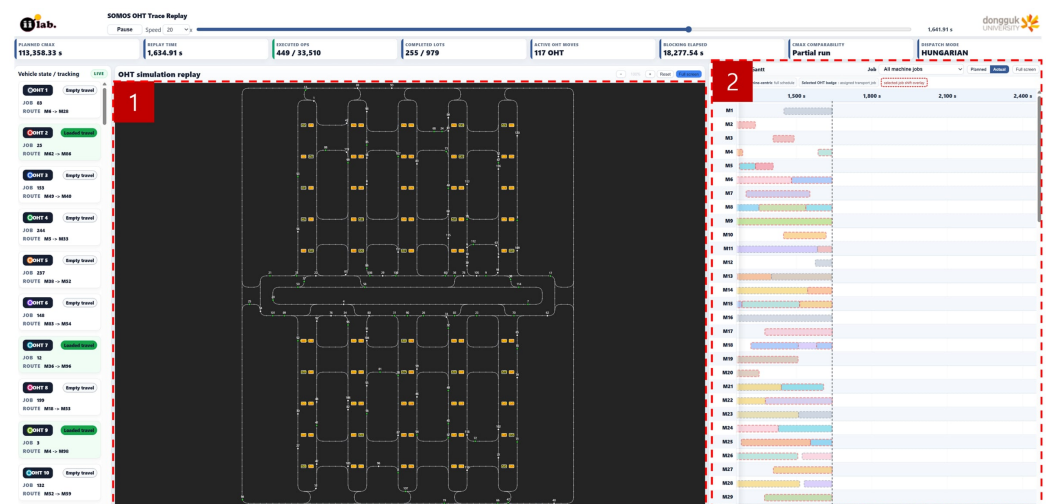


Figure 3. Integrated visualization interface of SOMOS for synchronized OHT transportation and production-schedule monitoring.

A more detailed relationship between an individual OHT operation and its associated production job is provided in Figure 4. For a selected vehicle, SOMOS records the assigned transport job, source and destination information, travel route, and empty or loaded movement state. The corresponding production job is simultaneously identified in the machine Gantt chart. This enables a transport operation to be followed from vehicle assignment and movement to the production operation associated with the transported lot. Figures 3 and 4 therefore provide complementary levels of observation. Figure 3 represents the system-wide evolution of OHT transportation and production execution, whereas Figure 4 supports detailed tracing of individual vehicle and job trajectories.

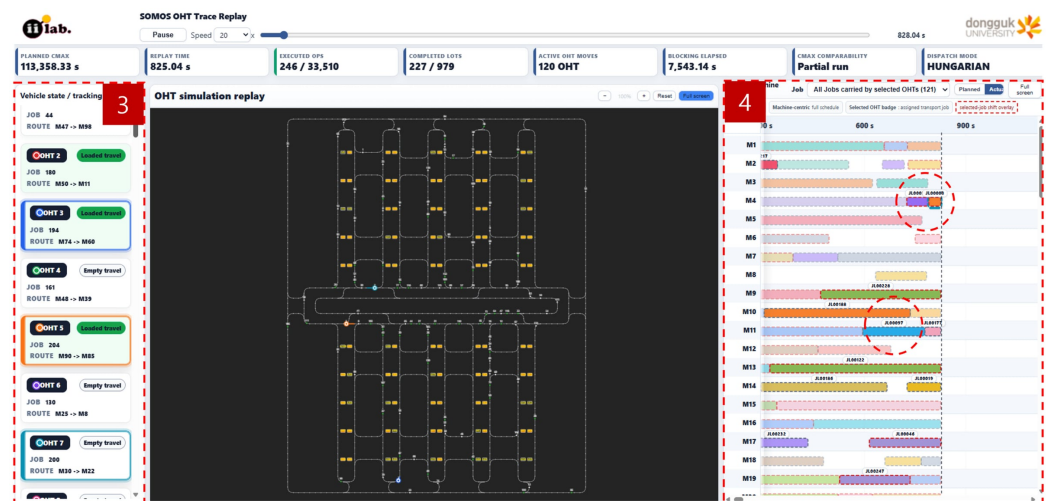


Figure 4. Detailed OHT and production job trace interface in SOMOS.

Together, the event-driven simulation and visualization functions of SOMOS provide a common environment for examining how OHT decisions are associated with subsequent production execution. The visualization functions are used for monitoring and trace analysis and do not modify the dispatching decisions or production logic executed during simulation. A demonstration of the synchronized OHT movement, vehicle state tracking, and production schedule visualization in SOMOS is available online (<https://youtu.be/esarS0voKZ8> (accessed on 29 June 2026)).

To verify the internal consistency of SOMOS, the simulator was examined at both the production and transportation event levels. The production-side configuration follows

the re-entrant semiconductor FAB testbed described in Section 3.1, including its machine structure and processing-time variation. Event traces were also checked to confirm that transport requests are generated upon realized operation completion, that pickup and delivery are completed before the subsequent operation starts, and that transportation delays are propagated to the realized production trajectory. These checks constitute model verification of the internal simulation logic and event consistency, rather than external validation against industrial production or OHT data.

4. Proposed Method

Figure 5 illustrates the overall procedure of the proposed SEVD. SEVD uses the current production state and production-schedule information obtained from SOMOS to estimate the spatial distribution of transport demand. Based on the estimated demand, idle OHT vehicles are repositioned toward spatial cells requiring greater vehicle availability. SEVD does not assign a specific vehicle to a future pickup request or predict the exact occurrence time of that request. Instead, it adjusts the spatial distribution of idle vehicles before subsequent requests occur. When an actual pickup request is generated, vehicle assignment is performed using the conventional HUNGARIAN procedure. SEVD therefore operates as an idle-vehicle positioning mechanism that precedes request-based vehicle assignment. Unlike the predictive and look-ahead approaches of Wan and Shin and Benzoni et al., which incorporate estimated future transport arrivals and vehicle availability into the assignment decision, SEVD does not predict or reserve individual future transport requests [27,28]. Instead, it performs pre-request spatial repositioning of idle vehicles, while actual request assignment remains unchanged. Accordingly, SEVD is designed to improve transportation responsiveness through idle-vehicle positioning rather than to optimize lot-level production priorities such as due-date urgency, tardiness, slack, or downstream criticality.

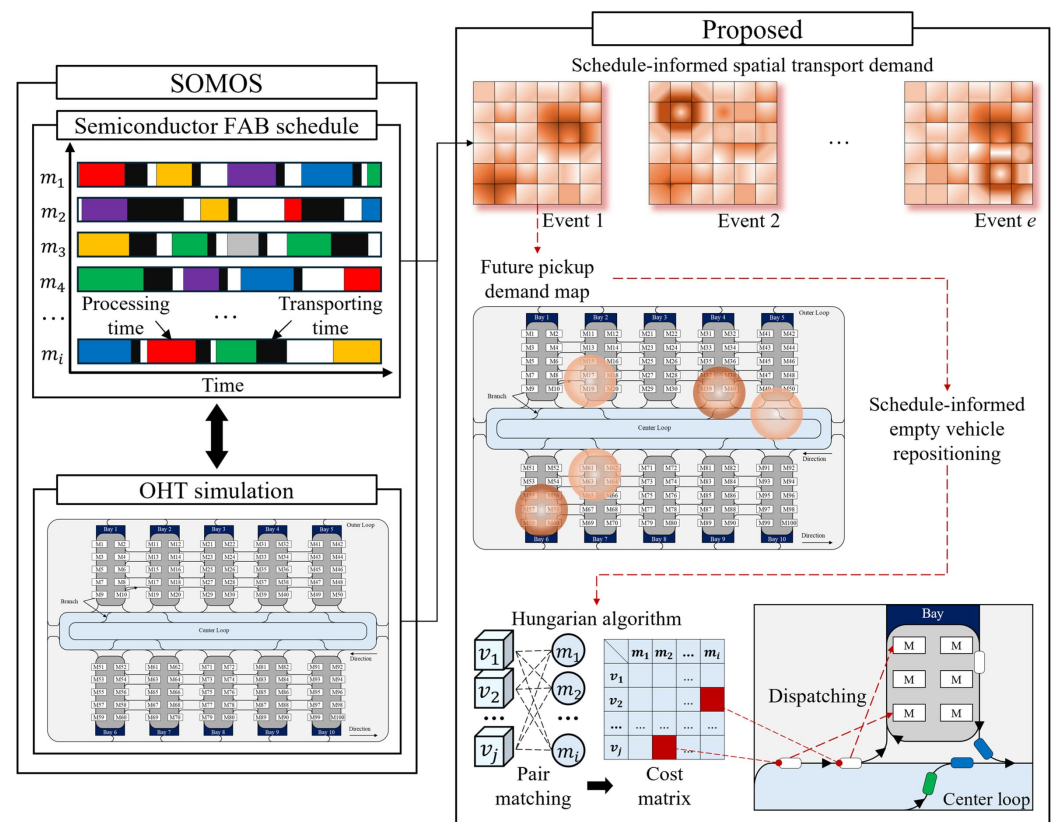


Figure 5. Overall framework of the proposed Schedule-informed Empty-Vehicle Dispatching.

4.1. Spatial Transport Demand Estimation

As shown in Figure 5, SEVD estimates spatial transport demand using three sources of information. The first represents immediate pickup demand at the output buffer of each machine. The second represents demand associated with the operation currently approaching completion. The third represents a spatial demand prior obtained from the production schedule. For machine M_i , the demand weight at simulation time t is defined as Equation (5):

$$w_i(t) = \text{buffer_num} \cdot \mathbf{1}[\text{out_buffer}(M_i, t) \neq \emptyset] + \left(1 - \frac{\rho_i(t)}{H}\right)^+ \cdot \mathbf{1}[\rho_i(t) \leq H] + \lambda \frac{g_i}{g_{\max}} \quad (5)$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. The first term represents immediate transport demand at machine M_i . When a completed lot is present in the output buffer, a fixed demand weight of *buffer_num* is assigned. In the experiments, *buffer_num* is set to 3. Under the default $\lambda = 1$, the second and third terms in Equation (5) are each bounded by 1. Therefore, setting *buffer_num* = 3 ensures that a machine with an existing pickup request receives greater demand weight than the maximum combined contribution of the two anticipated-demand terms. The second term represents transport demand expected to arise from the operation currently being processed. Here, $\rho_i(t)$ denotes the remaining processing time of the current lot at M_i when a subsequent transport operation is required, and H denotes the remaining processing time window. If M_i is idle or its current operation does not require a subsequent transport operation, the second term is defined as zero.

In the present simulation, $\rho_i(t)$ is computed from the realized processing time sampled when the ongoing operation starts. Consequently, once an operation has started, SEVD has access to its exact remaining duration. Therefore, $\rho_i(t)$ should be interpreted as simulator-side remaining-processing information rather than as an operational estimate. This information is used only in the repositioning step of SEVD and is not used by the benchmark methods. As $\rho_i(t)$ decreases, the corresponding demand weight increases, giving greater importance to machines whose current operations are approaching completion. The parameter H therefore has two roles in this term. It determines whether an ongoing operation is included through the condition $\rho_i(t) \leq H$, while also controlling the relative weighting of eligible operations through $1 - \rho_i(t)/H$. The third term provides a schedule based spatial prior. The value g_i denotes the number of planned transport tasks for which M_i appears as the source machine in the production schedule, and $g_{\max}$ denotes the maximum value of g_i among all machines. The parameter λ determines the contribution of this schedule-based information. Unlike the second term, this component does not estimate the exact occurrence time of an individual future pickup request. Instead, it provides information on how transport demand is spatially distributed across the FAB according to the production plan. Accordingly, this term should be interpreted as an auxiliary spatial prior rather than a temporal predictor; its incremental contribution is evaluated directly in the ablation analysis in Section 5.3. The machine level demand is subsequently aggregated into spatial cells. Let $\mathcal{C}$ denote the set of spatial cells and $c \in \mathcal{C}$ denote an individual cell. In the default configuration, the cell size is determined as one fourth of the maximum spatial span of the machine layout. The transport demand of cell c is obtained by summing the demand weights of the machines located within the cell as defined in Equation (6):

$$D_c(t) = \sum_{M_i \in c} w_i(t) \quad (6)$$

A larger $D_c(t)$ therefore indicates a greater concentration of current and schedule-based transport demand around cell c . The resulting cell level demand $D_c(t)$ is used to determine the target spatial dispatching of idle OHT vehicles, as described in Section 4.2.

4.2. Idle-Vehicle Dispatching

Based on the spatial transport demand obtained in Section 4.1, SEVD determines the target spatial dispatching of idle OHT vehicles and repositions vehicles from surplus cells toward deficit cells. Algorithm 1 summarizes the overall procedure. Let $N^I(t)$ denote the total number of idle OHT vehicles at simulation time t , and let $N_c(t)$ denote the number of idle vehicles currently located in cell c . The target number of idle vehicles for cell c is determined according to the relative transport demand of the cell as:

$$N_c^*(t) = \frac{D_c(t)}{\sum_{c' \in C} D_{c'}(t)} N^I(t) \quad (7)$$

Equation (7) distributes the available idle-vehicle capacity according to the relative spatial transport demand. Cells with greater $D_c(t)$ are consequently assigned a greater proportion of the idle OHT vehicles. Because $N_c^*(t)$ is generally not an integer, the implementation uses the floor value of the target when identifying surplus vehicles and the ceiling value when determining the number of vehicles required by a deficit cell. In particular, idle vehicles exceeding $\lfloor N_c^*(t) \rfloor$ are treated as available for repositioning. All idle vehicles located in cells with zero transport demand are also included in the surplus vehicle set. After the surplus vehicles are identified, deficit cells are considered in descending order of $D_c(t)$. This gives cells with greater estimated transport demand priority when additional idle vehicles are distributed. For each deficit cell, SEVD selects the nearest vehicle from the available surplus set. Let $\mathcal{V}^S(t)$ denote the set of surplus idle vehicles and let ρ_c denote the representative OHT node of cell c . The repositioning vehicle is determined as Equation (8):

$$V^* = \underset{V_j \in \mathcal{V}^S(t)}{\operatorname{argmin}} d(\operatorname{pos}(V_j), \rho_c) \quad (8)$$

where $d(\cdot, \cdot)$ denotes the travel time between two OHT nodes obtained from the rail network using A* path search. The representative node ρ_c is the machine node with the highest demand weight, $w_i(t)$, within cell c . The selected surplus vehicle is repositioned toward ρ_c , and this procedure is repeated according to the required number of additional vehicles in each deficit cell. The implemented procedure therefore considers both the magnitude of spatial transport demand and the rail travel time required to redistribute the available idle vehicles.

Algorithm 1 determines only the repositioning targets of idle vehicles and does not reserve a vehicle for a specific future transport request. A repositioning vehicle remains available for normal vehicle assignment while moving toward its target cell. Therefore, when an actual pickup request is generated, the repositioning movement can be interrupted and the vehicle can immediately participate in the conventional assignment procedure. This design changes the spatial dispatching of idle vehicles without modifying the subsequent vehicle assignment rule.

Algorithm 1. Schedule-informed Empty-Vehicle Dispatching

Input Spatial cell set: C , cell demand; $D_c(t)$, idle vehicle set; $V_{idle}(t)$, current vehicle positions.

```

1:   Obtain the total number of idle vehicles  $N^I(t)$ 
2:   for each cell  $c \in C$  do
3:       Calculate the target idle-vehicle count  $N_c^{(t)}$  using Equation (7)
4:       Determine the current idle-vehicle count  $N_c(t)$ 
5:   end for
6:   Construct the surplus vehicle set  $\mathcal{V}^S(t)$ 
7:   Rank deficit cells in descending order of  $D_c(t)$ 
8:   for each deficit cell  $c$  do
9:       Determine the required number of additional idle vehicles
10:      while additional idle vehicles are required do
11:          Select the nearest vehicle  $V^*$  from  $\mathcal{V}^S(t)$  using Equation (8)
12:          Set  $\rho_c$  as the repositioning target of  $V^*$ 
13:          Keep  $V^*$  available for normal vehicle assignment
14:          Remove  $V^*$  from  $\mathcal{V}^S(t)$ 
15:      end while
16:   end for
17:   return updated repositioning targets

```

Output Updated repositioning targets

5. Experiments

5.1. Experiment Settings

All experiments were conducted using SOMOS under the same production schedule, machine configuration, rail layout, and transport task information. The primary comparison used 120 OHT vehicles over a 24 h simulation horizon. Each dispatching method was evaluated using 20 replications with random seeds from 1 to 20, and the same seed set was applied to all methods to support paired comparisons. Realized processing times were sampled in each run according to the step-specific processing-time distribution described in Section 3.1, with the stochastic sampling controlled by the simulation seed. The common seed set was used to maintain a consistent replication basis for seed-matched comparisons across the dispatching methods. This design allows performance differences to be evaluated under common production and system configurations while accounting for processing-time variability.

Figure 6 illustrates the OHT rail layout used in the experiments. The simulated system consists of 10 bays and 100 production machines connected through branch rails, a center loop, and an outer loop. Each bay contains ten machines connected to the main rail network through local branch segments. OHT vehicles move along predefined rail directions and travel between bays through the center and outer loops. This layout is used consistently across all dispatching methods so that the observed performance differences originate from vehicle assignment and idle-vehicle dispatching rather than from changes in the physical transport network.

Table 2 summarizes the simulation and SEVD parameter settings used in the experiments. The simulated FAB consists of 10 bays and 100 machines belonging to six process types. The primary experiment was conducted using 120 OHT vehicles with a WIP limit of 250. Twenty independent replications were performed using seeds from 1 to 20, and each dispatching method was evaluated using the same seed set to enable paired comparisons. The OHT vehicles travel at 5.0 m/s on straight rail segments and 2.5 m/s on curved segments. Loading and unloading require 25 s each, and vehicle routes are determined using A* path search based on rail travel time. Unresolved transport requests are reconsidered at 2 s intervals. The execution WIP limit was set to 250, and machine breakdowns were not considered in the

experiments. For SEVD, the immediate pickup demand weight, *buffer_num*, was set to 3. The remaining processing time window, *H*, was set to 900 s, and the schedule demand weight, λ , was set to 1.0. The spatial cell size, γ , was automatically determined as one fourth of the maximum spatial span of the machine layout. These values are treated as default operating settings rather than uniquely optimized parameters. Their robustness was evaluated through the one-factor-at-a-time sensitivity analysis in Section 5.3, including *span*/2 and *span*/6 for the spatial cell sizes and 15 s and 120 s for the repositioning intervals. Idle-vehicle repositioning was updated every 45 s. These values correspond to the default SEVD configuration used in the primary experiments. SOMOS was implemented in Python 3.11, and the complete software dependencies and version information required for reproducing the simulation are provided in the publicly available GitHub repository.

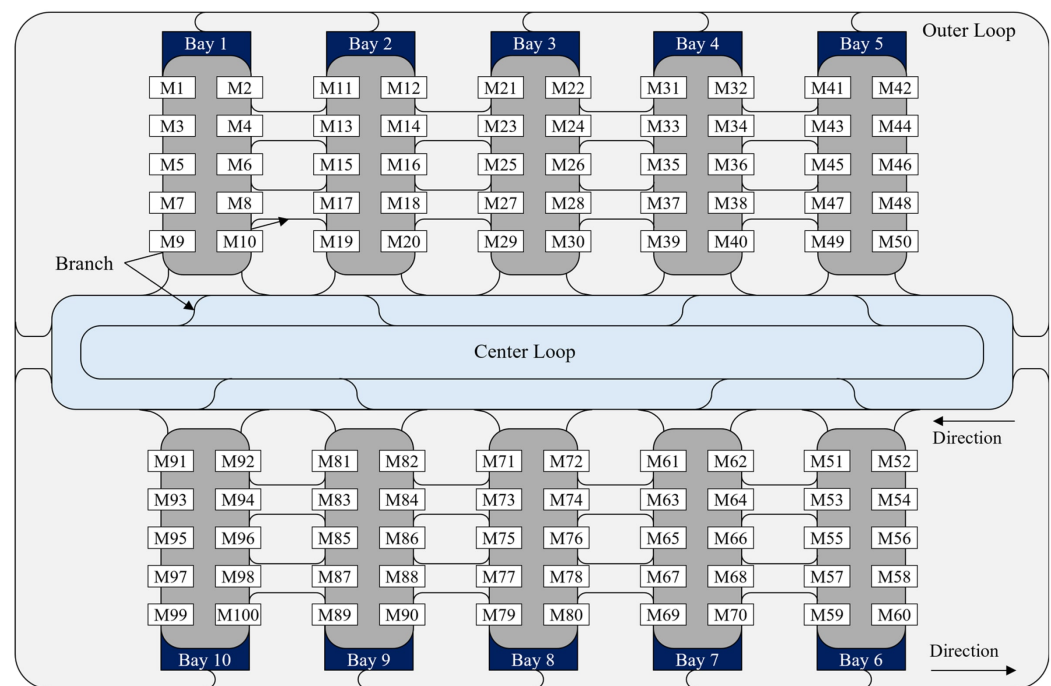


Figure 6. OHT rail layout and machine configuration used in the simulation experiments.

Table 2. Simulation and SEVD parameter settings used in the experiments.

Category	Parameter	Symbol	Setting
Simulation	Number of bays		10
	Number of machines		100
	Number of process types		6
	Simulation horizon		24 h
	Number of OHT vehicles	<i>N</i>	120
	Number of replications		20
	WIP limit		250
	OHT speed on straight rail		5.0 m/s
	OHT speed on curved rail		2.5 m/s
	Loading time		25 s
	Unloading time		25 s
	Path search		A*
SEVD	Dispatch retry interval		2 s
	Immediate pickup demand weight	<i>buffer_num</i>	3
	Remaining processing time window	<i>H</i>	900 s
	Schedule demand weight	λ	1.0
	Spatial cell size	γ	span divided by 4
	Repositioning interval	<i>d_t</i>	45 s

As the number of OHT vehicles increases, interactions among vehicles become more frequent at branch connections and shared rail segments [47]. Figure 7 illustrates representative traffic conditions observed in the simulation. A local deadlock occurs when vehicle movements mutually restrict one another around a bay branch, whereas congestion occurs when several vehicles approach or pass through the same rail connection within a short period. When persistent vehicle blocking prevents further system progress, the simulation watchdog terminates the event loop before the predefined 24 h horizon, and the run is classified as collapsed. This collapse criterion is distinct from the 75% production attainment criterion, which is used separately to evaluate production progress. The likelihood of such conditions generally increases as the number of OHT vehicles increases because more vehicles simultaneously compete for shared rail segments and branch connections. In this study, these collapsed runs are retained as system-level operational outcomes rather than excluded from the performance evaluation. These traffic conditions are treated as outcomes of vehicle movement in the simulated rail network rather than as separate control objectives of the proposed method.

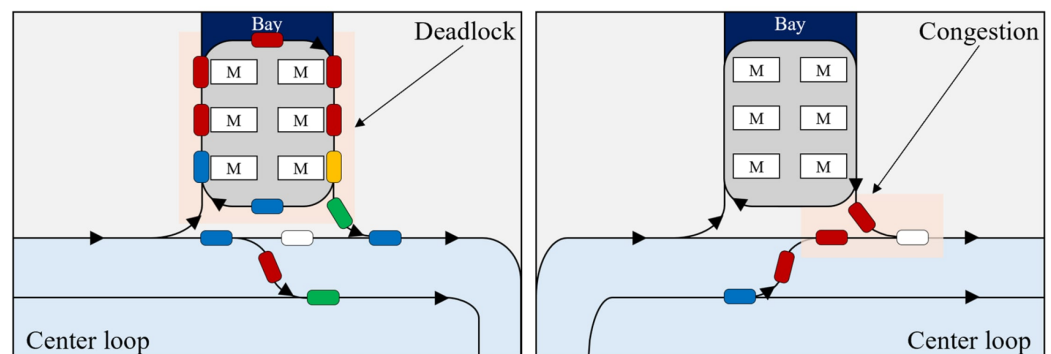


Figure 7. Examples of OHT traffic conditions observed in the simulated rail network.

The primary experiment evaluated SEVD against six benchmark dispatching methods consisting of five rule-based methods and the HUNGARIAN assignment method. NVF prioritizes the vehicle task pair with the shortest empty travel time to the pickup location [48]. STD considers the sum of empty and loaded travel times and selects assignments with shorter total transportation time [14,48]. EDD prioritizes transport requests with earlier planned due times [49], whereas FIFO gives priority to requests that became ready earlier [48,49]. PRIORITY considers both request tardiness and the number of lots waiting at the source machine, thereby favoring requests with greater delays and queue pressures [43]. These five methods determine vehicle task assignments sequentially according to their respective priority criteria. HUNGARIAN differs from the rule-based methods by determining multiple vehicle task assignments simultaneously [15]. At each dispatching decision, a cost matrix is constructed between the currently available vehicles and ready transport requests, and the assignment that minimizes the total transportation cost is obtained by solving a linear assignment problem [15,32]. SEVD uses the same HUNGARIAN assignment procedure for actual transport requests but additionally adjusts the spatial distribution of idle vehicles according to the transport demand estimated in Section 4. Therefore, the comparison between HUNGARIAN and SEVD isolates the effect of the proposed idle-vehicle repositioning mechanism while maintaining the same vehicle assignment procedure. However, this comparison does not establish that the specific SEVD demand formulation is superior to alternative proactive repositioning strategies.

Performance was evaluated using both production-level and transportation-level measures. Production performance was assessed using throughput, machine utilization, attainment rate, and the makespan attainment index. Throughput is defined as the num-

ber of lots completing the fabrication process within the 24 h simulation horizon. Let N_{real} denote the number of lots completed within this horizon and $P(N_{\text{real}})$ denote the planned completion time of the corresponding lot in the reference production schedule. The makespan attainment index converts the realized production shortfall into the time scale of the reference schedule and is defined as Equation (9):

$$M_{\text{att}} = 24\text{h} + [24\text{h} - P(N_{\text{real}})] = 48\text{h} - P(N_{\text{real}}) \quad (9)$$

When the realized production progress exactly reaches the level planned for 24 h, $P(N_{\text{real}}) = 24\text{h}$ and $M_{\text{att}} = 24\text{h}$. If production falls behind the reference schedule, $P(N_{\text{real}}) < 24\text{h}$, resulting in $M_{\text{att}} > 24\text{h}$. Conversely, $M_{\text{att}} < 24\text{h}$ indicates production progress ahead of the 24 h reference level. Thus, a lower M_{att} indicates better production performance. The corresponding attainment rate is defined as Equation (10):

$$R_{\text{att}} = \frac{P(N_{\text{real}})}{24\text{h}} \times 100 \quad (10)$$

Accordingly, the 75% attainment criterion corresponds to $M_{\text{att}} \leq 30\text{h}$. Machine utilization represents the proportion of simulation time during which the production machines are actively processing lots. Higher throughput, attainment rate, and machine utilization indicate better production performance, whereas a lower makespan attainment index indicates better performance. Transportation performance was evaluated using source-waiting time and empty travel time to the source machine.

Source-waiting time is defined as the interval between completion of a machine operation and pickup of the corresponding lot by an OHT vehicle. Both the mean and the 90th percentile of source-waiting time were considered to represent average pickup delay and upper tail waiting conditions, respectively. Source-waiting time is recorded only for transport requests whose deliveries are completed within the observed simulation horizon. Requests that remain undelivered when a run terminates, either because the lot has not yet been picked up or because it is still in transit, therefore do not yield a completed source-waiting observation and are excluded from the reported source-wait statistics. This treatment applies to both full-horizon and collapsed runs.

The post-assignment empty approach time measures the travel time of an assigned empty OHT vehicle from assignment until arrival at the pickup machine. This metric excludes pre-request repositioning and undirected idle roaming, which are accounted for separately in the travel decomposition. Total empty travel is defined as the sum of approach, repositioning, and roaming movement. These measures directly characterize the transportation process affected by idle-vehicle dispatching. Lower source-waiting time and empty travel time indicate better transportation performance. For runs terminating before the 24 h horizon because of system collapse, transportation measures were calculated using the transport observations recorded up to the termination time. Because the same 20 random seeds were used for all methods, method comparisons were performed on a paired basis. Two-sided Wilcoxon signed-rank tests were used to evaluate paired differences, and 95% bootstrap confidence intervals of the paired mean differences were obtained using 20,000 resamples. These analyses were applied to both transportation- and production-level measures. For the parameter-sensitivity comparisons involving multiple alternative configurations, p -values were adjusted using the Holm procedure to control for multiple comparisons.

5.2. Experiment Results

Table 3 summarizes the overall production and transportation performance of the seven dispatching methods across the 20 common simulation seeds. Values are reported

as mean $\pm$ standard deviation over 20 simulation replications. EDD and FIFO show substantially lower production progress, while PRIORITY provides intermediate performance. NVF, STD, HUNGARIAN, and SEVD form the competitive group in terms of production-level measures. Among these methods, STD records the lowest mean makespan attainment index of 31.85 h and the highest unconditional mean throughput of 455.4 lots, whereas SEVD records 32.17 h and 446.9 lots, respectively. SEVD therefore does not uniformly outperform the strongest benchmarks at the production-level. SEVD nevertheless satisfies the 75% attainment criterion in 13 of the 20 runs, compared with 12 for STD and 10 for both NVF and HUNGARIAN. Taken together, these results indicate that the advantage of SEVD does not lie in uniformly outperforming the strongest benchmarks on every production-level measure, but in maintaining competitive production progress while providing a clearer improvement in transportation performance.

Table 3. Overall performance comparison of the seven dispatching methods.

Method	$R_{att} \geq 75\%$	M_{att} (h)	Throughput	Machine Utilization (%)	Source Wait p90 (s)	Empty Travel to Source (s)
NVF	10/20	34.06 (± 6.54)	394.7 (± 178.9)	49.5 (± 23.0)	442.7 (± 29.4)	91.4 (± 5.6)
STD	12/20	31.85 (± 4.55)	455.4 (± 124.1)	57.5 (± 16.0)	438.3 (± 12.5)	91.5 (± 3.0)
EDD	0/20	47.26 (± 0.35)	26.2 (± 14.5)	2.9 (± 1.5)	689.7 (± 193.7)	228.4 (± 29.5)
FIFO	1/20	46.41 (± 4.15)	48.3 (± 115.8)	5.9 (± 14.5)	506.0 (± 87.0)	211.8 (± 38.4)
PRIORITY	8/20	38.98 (± 8.87)	257.9 (± 244.0)	32.6 (± 31.4)	483.3 (± 47.8)	140.8 (± 45.2)
HUNGARIAN	10/20	33.90 (± 6.58)	398.4 (± 180.6)	50.1 (± 23.2)	428.7 (± 28.3)	96.8 (± 14.4)
SEVD	13/20	32.17 (± 6.06)	446.9 (± 168.8)	56.5 (± 21.6)	399.3 (± 11.0)	71.2 (± 17.1)

A further decomposition of throughput according to run outcome helps explain these aggregate differences. Among the runs satisfying the attainment criterion, SEVD records the highest mean throughput of 557.5 lots, compared with 548.4 for STD, 545.3 for HUNGARIAN, and 550.1 for NVF. The lower unconditional throughput of SEVD relative to STD is therefore largely associated with differences among the runs not meeting the attainment criterion rather than with lower production progress in the runs satisfying the criterion.

Here, a collapsed run refers to a simulation realization that terminates before the 24 h horizon because persistent vehicle blocking prevents normal transport operations and, consequently, subsequent production progress, as described in Section 5.1. The collapsed SEVD runs record a mean throughput of 241.6 lots and terminate between approximately 1.3 and 19.1 h, whereas the collapsed STD runs record a higher mean throughput of 315.8 lots and terminate between approximately 6.1 and 19.0 h. Early collapse therefore has a substantial influence on the unconditional production measures. A similar but more pronounced pattern is observed for PRIORITY, FIFO, and EDD. This decomposition explains the large seed-to-seed variation in production-level performance and supports the use of the unconditional all-seed results as the primary basis for comparison, with criterion-conditioned results used only to interpret the source of the variation. This result indicates that the lower unconditional throughput of SEVD relative to STD is primarily

associated with run-level collapse behavior rather than with poorer production progress when the system maintains sufficient attainment.

In contrast, the transportation-level results show a clearer advantage for SEVD. SEVD records the lowest source wait p90 of 399.3 s and the lowest empty travel to the source of 71.2 s. The corresponding values for HUNGARIAN are 428.7 s and 96.8 s, while NVF and STD also show substantially larger empty travel values of 91.4 s and 91.5 s, respectively. These results indicate that the primary advantage of SEVD is observed at the transportation-level, where proactive idle-vehicle dispatching directly reduces empty-approach travel and upper-tail pickup waiting while preserving competitive production-level performance.

Figure 8 further examines the seed-level variation underlying the aggregate results in Table 3. Panel (a) shows that the production outcomes are not simply separated by small shifts in their mean values. Instead, the results show substantial variation among simulation realizations. EDD and FIFO are concentrated almost entirely in the high-makespan-attainment region, whereas NVF, HUNGARIAN, STD, and SEVD produce a much larger number of runs near the lower-attainment region. SEVD satisfies the 75% attainment criterion in 13 of the 20 seeds, followed by STD with 12 and NVF and HUNGARIAN with 10 each, as summarized in Panel (b).

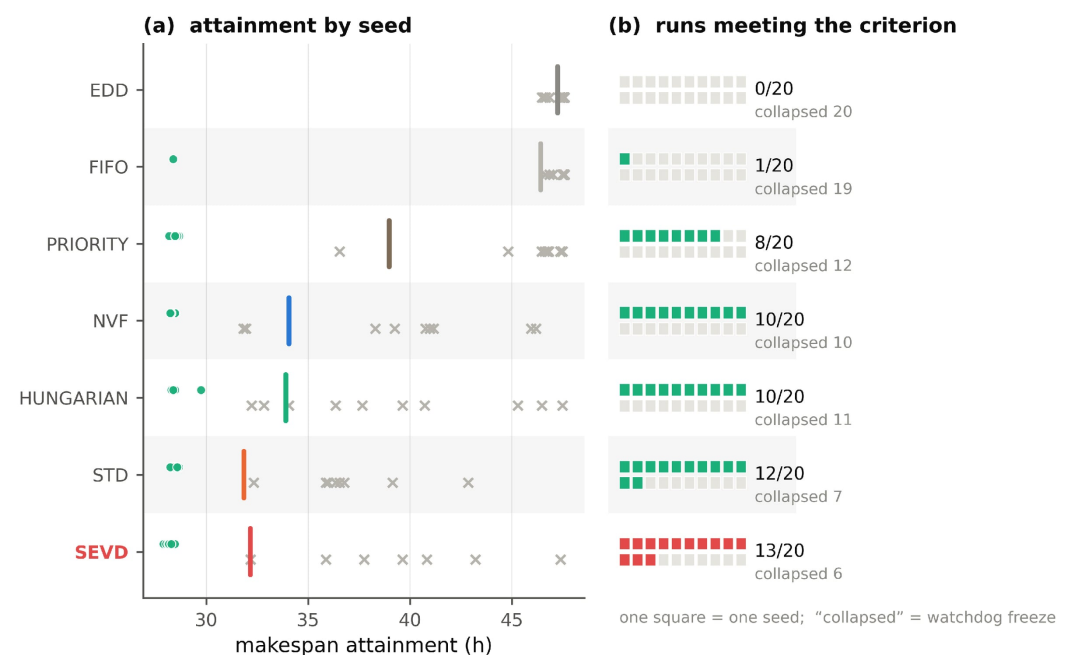


Figure 8. Production attainment of the seven dispatching methods across 20 common simulation seeds. In panel (a), each marker represents one seed: filled circles denote runs that satisfy the 75% attainment criterion, and crosses denote runs that do not. The vertical bar indicates the mean over all 20 seeds. In panel (b), each square represents one seed, and filled squares denote runs that satisfy the criterion. Collapsed gives the number of runs terminated by the gridlock watchdog.

However, Figure 8 also shows why the attainment count should not be interpreted as direct evidence of overall superiority. STD achieves a slightly lower mean makespan attainment index than SEVD, and the seed-level outcomes of the competitive methods overlap considerably. Paired statistical tests confirm that these production-level differences are not significant. The makespan attainment difference is +0.32 h relative to STD (95% CI [−2.68, 3.49], $p = 0.629$) and −1.73 h relative to HUNGARIAN (95% CI [−5.64, 2.01], $p = 0.294$). These results indicate that SEVD remains competitive with the strongest production-level benchmarks, while substantial run-to-run variation remains across the common seeds. This motivates a direct examination of the transportation-level effects of SEVD in Figure 9.

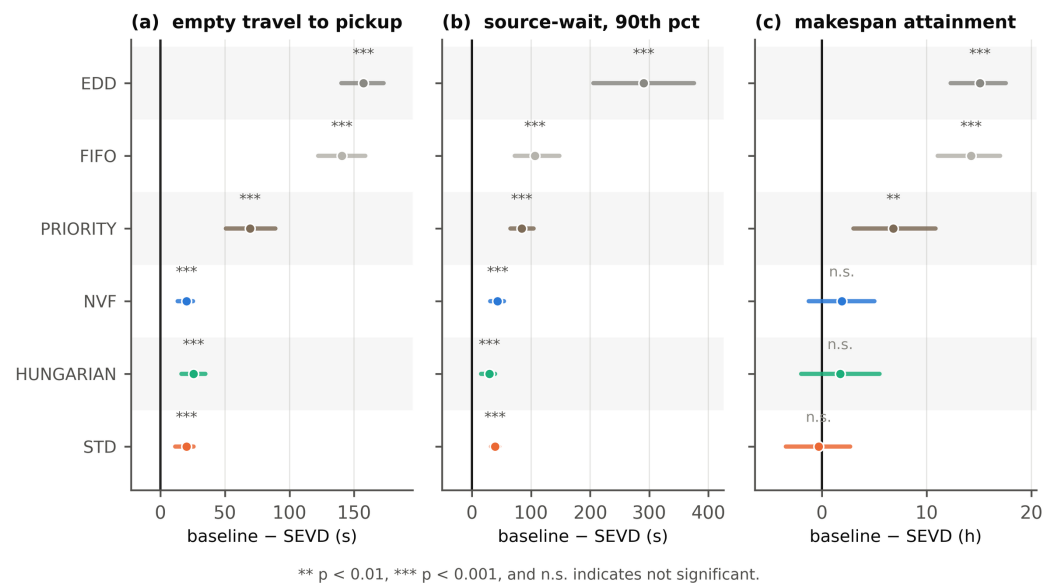


Figure 9. Paired comparison of SEVD and the benchmark dispatching methods across 20 common simulation seeds for transportation- and production-level performance measures.

Figure 9 provides the most direct evidence of the operational mechanism introduced by SEVD. Unlike the production-level measures in Figure 8, both transportation measures show improvements relative to all benchmark methods. The paired differences remain positive for empty travel to the source and for the 90th percentile of source-waiting time, indicating that SEVD generally places vehicles in more favorable locations before actual transport requests are assigned. The comparison with HUNGARIAN is particularly important because the two methods use the same HUNGARIAN assignment procedure after a pickup request is generated. Their primary difference is therefore the pre-request re-dispatching of idle vehicles performed by SEVD. Mean post-assignment empty-approach travel decreases from 96.8 s under HUNGARIAN to 71.2 s under SEVD, corresponding to a reduction of approximately 26.4%. This reduction refers specifically to the approach movement after assignment and does not represent a reduction in total empty-vehicle movement. Source wait p90 decreases from 428.7 s to 399.3 s, corresponding to a reduction of approximately 6.9%. SEVD performs better in 19 of the 20 paired seeds for both measures. The paired differences are statistically significant, with a 25.59 s reduction in post-assignment empty-approach travel relative to HUNGARIAN (95% CI [16.25, 35.04], $p = 0.0005$) and a 29.40 s reduction in source wait p90 (95% CI [14.89, 39.31], $p = 0.0007$).

The result against NVF provides an additional insight. NVF already selects a nearby vehicle after a request becomes available, yet its mean empty travel remains approximately 20 s higher than that of SEVD. This indicates that reactive selection of a nearby vehicle cannot fully compensate for an unfavorable idle-vehicle dispatching that already exists at the time of request generation. The principal effect of SEVD therefore occurs before request assignment by improving the spatial starting positions of idle vehicles. The larger relative reduction in empty travel than in source waiting also clarifies the propagation of this effect. SEVD directly changes the location from which an empty vehicle approaches the pickup machine. Reduced pickup waiting is a subsequent effect that is additionally influenced by vehicle availability, request timing, and other concurrent transport events.

Figure 10 further decomposes empty-vehicle movement into post-assignment approach, pre-request repositioning, and undirected roaming. Aggregated over all 20 seeds per vehicle-hour of realized operation, SEVD reduces approach travel by 28.0% and undirected roaming by 79.4% relative to HUNGARIAN, but the added repositioning increases total empty travel by 17.8%. Total travel distance also increases by 6.8%, while empty and

loaded blocking times decrease by 15.5% and 28.1%, respectively. These results indicate that SEVD redistributes empty movement toward purposeful repositioning rather than reducing the total amount of empty movement. Because the simulator does not include an explicit energy-consumption model, total travel distance is reported only as a distance-based proxy rather than as a direct energy estimate.

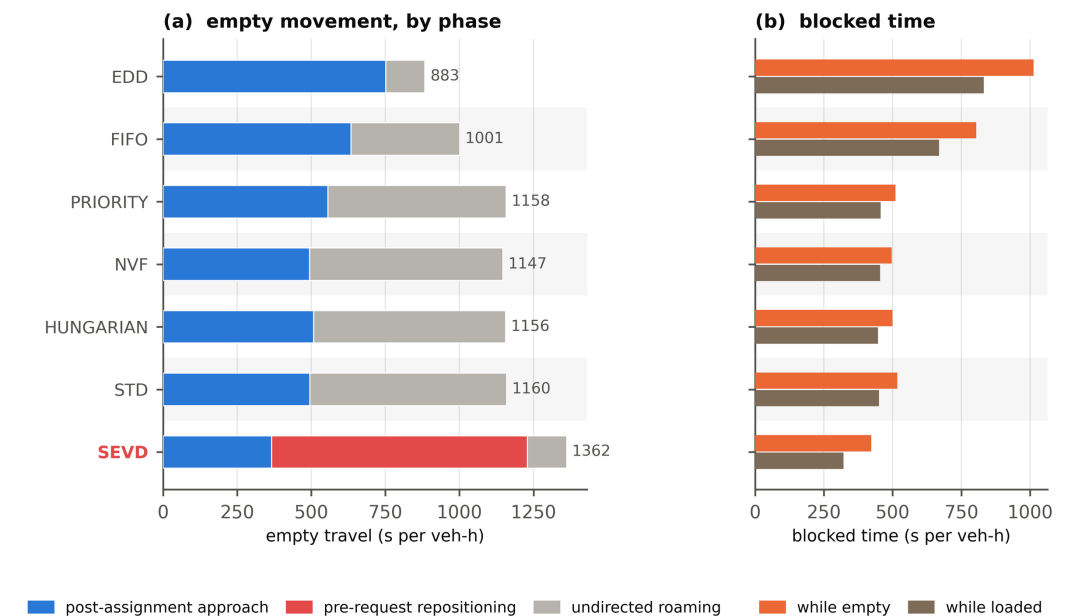


Figure 10. Decomposition of empty-vehicle movement and associated traffic effects.

Figure 11 examines whether the transportation advantage of SEVD is affected by unequal simulation durations. Panel (a) reports the termination times and collapse outcomes for each method. Panel (b) compares post-assignment empty-approach travel using two observation-horizon definitions. In the common-horizon comparison, both runs within each seed pair are truncated to the shorter realized duration before the transportation metric is recomputed. The full-horizon comparison includes only seed pairs in which both methods complete the 24 h simulation.

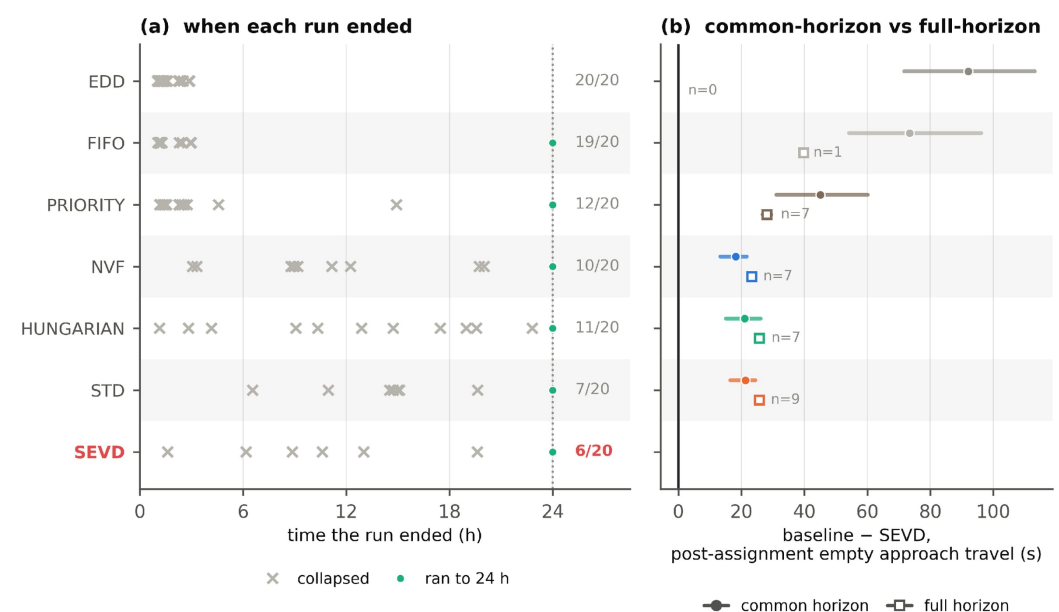


Figure 11. Run termination patterns and robustness of post-assignment empty-approach travel to differences in observation horizon.

Against HUNGARIAN, the mean advantage of SEVD is 21.0 s under the common-horizon comparison, with SEVD performing better in 18 of 20 paired runs. Among the seven pairs in which both methods complete the full 24 h horizon, the corresponding advantage is 25.6 s and SEVD performs better in all seven pairs. Similar positive differences are observed against the other competitive benchmarks. These results indicate that the transportation advantage of SEVD is retained after equalizing the observation horizon and is therefore not attributable to unequal simulation durations. The full-horizon subset is reported only as a supplementary robustness check because it conditions on both runs completing the simulation.

Because source-waiting time is observed only for completed deliveries, requests remaining undelivered at termination are excluded from the source-wait statistics, and this exclusion is not removed by equalizing the observation horizon. Across the 20 seeds, SEVD recorded 149.8 outstanding requests on average, compared with 179.5 for HUNGARIAN and 165.1 for STD. As a sensitivity analysis, each outstanding request was assigned a censored waiting value equal to the interval from operation completion to simulation termination. With these censored requests included, SEVD retained a lower source-wait p90 than HUNGARIAN in 15 of 20 paired seeds and lower than STD in 17 of 20 paired seeds. However, the mean differences became sensitive to early-collapsed runs. These results indicate that the relative advantage of SEVD is generally preserved across paired seeds, while also showing that source-wait statistics in collapsed runs remain affected by censoring.

The relative benefit of SEVD changes with OHT fleet size, as shown in Figure 12. At a fleet size of 80 vehicles, the difference relative to STD and HUNGARIAN is small, while NVF achieves shorter empty travel than SEVD. From 100 vehicles onward, however, the mean differences become positive for all three competitive benchmarks, and the advantage of SEVD increases further at 120 and 140 vehicles. Relative to HUNGARIAN, the mean empty travel difference increases from only a few seconds at 80 vehicles to approximately 48 s at 140 vehicles. This pattern indicates that the effectiveness of SEVD depends on the operating condition of the OHT fleet. In particular, the increasing advantage at larger fleet sizes is consistent with the interpretation that the spatial placement of idle vehicles becomes more influential when more vehicle capacity is present in the system. The result therefore identifies fleet size as an important operating condition for SEVD rather than demonstrating uniform superiority under every fleet configuration.

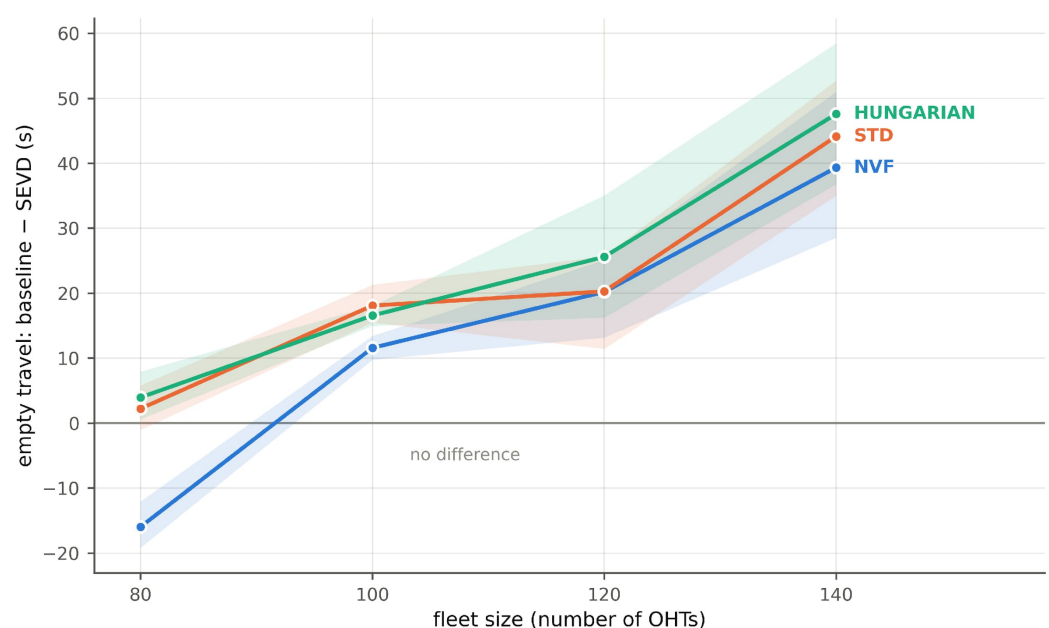


Figure 12. Effect of OHT fleet size on the empty travel reduction achieved by SEVD.

5.3. Ablation Study

To examine the contribution of individual SEVD components and the sensitivity of the proposed method to its parameter settings, Figure 13 presents the ablation and one-factor-at-a-time sensitivity results. Panel (a) reports post-assignment empty-approach travel, Panel (b) reports source wait p90, and Panel (c) shows the paired difference in post-assignment empty-approach travel relative to the default SEVD. Disabling the SEVD repositioning mechanism produces the clearest deterioration in both transportation measures. Post-assignment empty-approach travel increases from approximately 71.2 s under the default SEVD to 96.8 s under SEVD OFF, while source wait p90 increases from approximately 399.3 s to 428.7 s. The corresponding paired increase in post-assignment empty-approach travel is approximately 25.6 s, indicating that the pre-request repositioning mechanism itself accounts for the most pronounced transportation-level improvement.

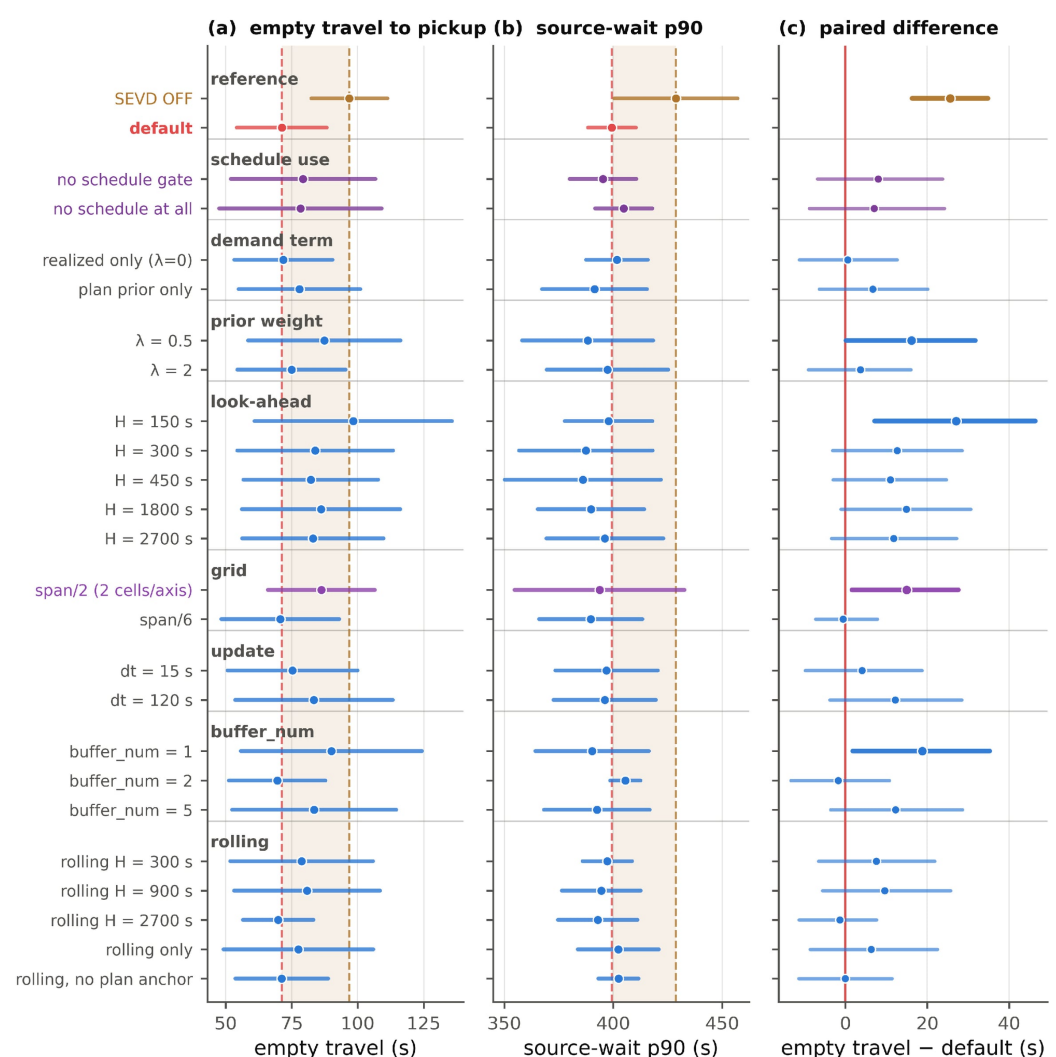


Figure 13. Ablation and parameter sensitivity analysis of SEVD. Each dot indicates the mean over 20 seeds and the horizontal line indicates standard deviation.

The demand-component ablations provide a more limited distinction among the individual information sources. The realized-only configuration ($\lambda = 0$), which removes the schedule-derived prior, does not significantly differ from the default SEVD in the all-seed comparison (mean difference = -0.58 s, 95% CI $[-12.63, 11.15]$, $p = 0.57$). Similarly, retaining only the schedule-derived prior does not provide a clearly distinguishable improvement

from the default in the all-seed results. These findings do not support attributing the overall transportation benefit to either demand component alone. Instead, the clearest contrast is between the active SEVD repositioning mechanism and SEVD OFF. In the supplementary full-horizon subset, removing the realized production-state terms while retaining the schedule prior increases post-assignment empty-approach travel by 2.36 s (95% CI [1.29, 3.43]). Because this comparison is conditional on both runs completing the simulation horizon, it is used only as supplementary evidence and not as the primary basis for attributing the overall SEVD benefit to the realized-state terms.

A configuration that removes all production-schedule information, including both the schedule-derived prior and the schedule-based eligibility condition, likewise does not differ significantly from the default SEVD in the all-seed comparison (mean difference = 7.07 s, 95% CI [−8.74, 24.03], $p = 0.99$). Therefore, in the present testbed, the incremental contribution of schedule information to the transportation benefit is not statistically detectable. The evidence instead supports the effectiveness of proactive repositioning within the SEVD framework rather than the superiority of its particular demand formulation.

The parameter-sensitivity results also indicate that most moderate changes in the prior weight, λ , temporal window, H , spatial resolution, repositioning interval, and immediate-demand weight remain closer to the default than SEVD OFF. Some more extreme settings show larger deterioration, particularly a short temporal window or a coarse spatial grid. Overall, Figure 13 indicates that the principal transportation benefit is associated with enabling proactive idle-vehicle repositioning, while the performance differences among individual demand components and moderate parameter settings are comparatively smaller.

The default parameter values should therefore be interpreted as operating settings within a stable region rather than as uniquely optimized values. For the spatial resolution, the course $span/2$ grid substantially reduces operational stability, with only 1 of 20 runs meeting the 75% attainment criterion, whereas $span/4$ and $span/6$ provide comparable all-seed transportation performance. For the repositioning interval, neither 15 s nor 120 s significantly outperforms the default 45 s setting after multiple-comparison correction. Similarly, alternative $buffer_num$ values do not significantly outperform the default, while $buffer_num = 3$ additionally preserves the structural priority of an existing pickup request described in Section 4.1.

Figure 14 examines whether the effectiveness of SEVD depends on accurately representing the timing of future transport demand. Panel (a) compares the default $H = 900$ s configuration with shorter and longer static windows and with alternative rolling implementations. Although most alternative settings show bootstrap confidence intervals that overlap zero, the shortest static window, $H = 150$ s, shows a clear deterioration in post-assignment empty-approach travel relative to the default $H = 900$ s. None of the tested alternatives provides a consistent improvement over the default configuration. Panel (b) provides an explanation for the limited sensitivity to larger static values of H . The planned processing times of the 33,510 operations have a maximum of approximately 676 s, while only 0.14% of realized operations exceed 900 s in the simulation runs. Consequently, the eligibility condition $\rho_i(t) \leq H$ is nearly saturated at the default $H = 900$ s. Increasing H to 1800 or 2700 s therefore does not introduce additional ongoing operations into this demand component. However, H continues to control the relative weighting of eligible operations through $1 - \rho_i(t)/H$. A larger H flattens this weighting function and reduces the contrast between operations with different remaining processing times. Thus, once the eligibility condition is saturated, H no longer changes which operations are included but still affects how strongly their imminence is differentiated.

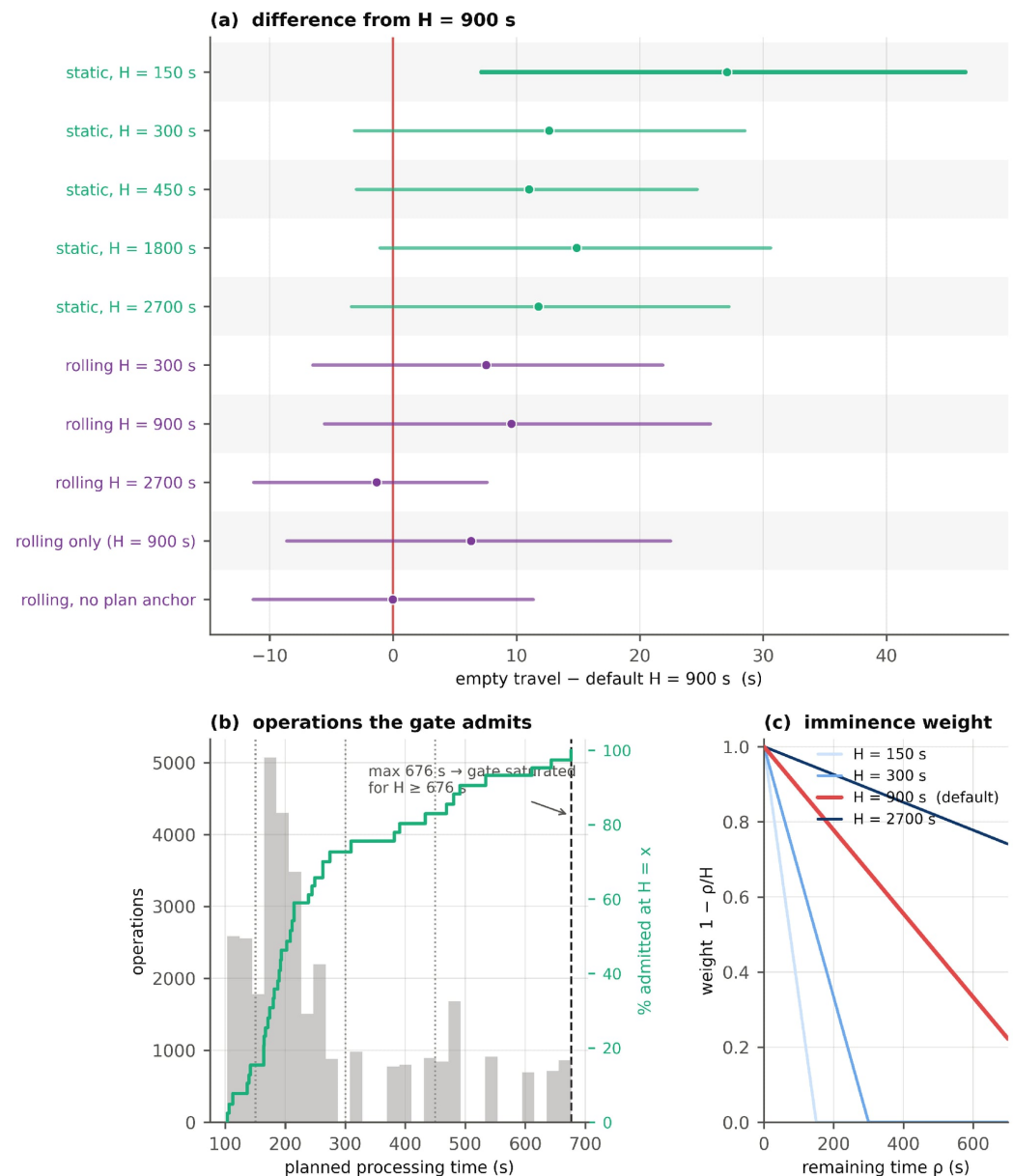


Figure 14. Effect of the temporal demand window and its relationship with the planned processing time variation.

More importantly, the rolling implementations were introduced to represent the temporal horizon more explicitly, yet they do not provide a consistent advantage over the simpler static implementation. The sensitivity results further show that this weighting effect does not create a sharply defined optimum, as no alternative H setting significantly outperforms the default after multiple-comparison correction. The shortest window, $H = 150$ s, shows a clear deterioration relative to the default, indicating that an excessively short temporal window can exclude relevant ongoing operations. However, no alternative H setting significantly outperforms the default after Holm correction. This result changes the interpretation of schedule awareness in SEVD. The benefit of using production-schedule information does not appear to originate primarily from precise prediction of when individual pickup requests will occur. The temporal-window results are consistent with the ablation findings. Increasing the temporal detail does not provide a consistent additional benefit. This indicates that SEVD does not rely on precise schedule timing. Instead, its observed advantage is primarily associated with proactive spatial repositioning rather than precise temporal schedule information.

6. Conclusions

This study developed SOMOS to represent the event-driven interaction between semiconductor production execution and OHT transportation and proposed SEVD for schedule-informed idle-vehicle dispatching. SEVD combines immediate pickup demand, remaining processing information, and a schedule-derived spatial demand prior to reposition idle vehicles before subsequent pickup requests occur, while retaining the conventional HUNGARIAN assignment procedure for actual transport requests. The experimental results show that the clearest benefit of SEVD appears at the transportation-level. Compared with HUNGARIAN, SEVD reduced mean post-assignment empty-approach travel from 96.82 s to 71.23 s and source wait p90 from 428.74 s to 399.34 s. This transportation benefit involves an operational trade-off. Aggregated over all seeds, proactive repositioning increased total empty travel by 17.8% and total travel distance by 6.8%, indicating that SEVD redistributes empty movement toward purposeful repositioning rather than reducing total vehicle movement. The production-level results showed greater run-to-run variation, and SEVD did not significantly outperform the strongest benchmark methods on production-level measures. Fleet size and ablation experiments further showed that the benefit of SEVD becomes more pronounced when spatial vehicle dispatching becomes operationally important, whereas detailed tuning of the temporal demand window provides limited additional improvement. These results indicate that SEVD's transportation benefit arises primarily from the proactive pre-request repositioning mechanism. The all-seed ablation does not identify either the realized production-state terms or the schedule-derived spatial prior as independently responsible for this benefit. In particular, the incremental contribution of the schedule-derived prior was not statistically detectable in the present testbed. Accordingly, the present results support the effectiveness of proactive, production-informed repositioning within the SEVD framework rather than establishing the proposed demand formulation as superior to alternative repositioning strategies. Comparison with other proactive repositioning baselines remains a direction for future work.

The present study has several limitations that define the direction of future work. The current SEVD formulation does not explicitly incorporate lot-level production priorities such as due-date urgency, tardiness, slack, or downstream criticality. Accordingly, SEVD is primarily intended to improve transportation responsiveness through idle-vehicle positioning, while incorporating production-criticality information into the demand formulation remains a direction for future work. In addition, source-waiting time is observed only for completed deliveries. Although outstanding requests are reported and a censored sensitivity analysis is provided, the reported source-wait statistics may underrepresent the longest waits, particularly in collapsed runs. Furthermore, SEVD uses the exact remaining processing time of ongoing operations, which is available in the simulator once the realized processing time has been sampled. In an industrial implementation, this quantity would need to be estimated from equipment-state or process-monitoring information, and estimation error may weaken the short-term imminence signal used by SEVD. A direct evaluation with estimated remaining processing times remains a direction for future work. Although deadlock and gridlock emerge from vehicle interactions in the simulated rail network, the evaluated dispatching methods do not include dedicated prevention or recovery mechanisms. Future research will therefore extend the OHT control framework to explicitly consider deadlock and gridlock prevention while preserving efficient vehicle dispatching and assignment. SEVD will also be extended toward adaptive dispatching in combination with reinforcement learning so that vehicle positioning and assignment policies can respond to changing production and transportation states. In addition, the current study evaluates OHT decisions using a predefined production schedule. Future work will integrate OHT operation with continuous rescheduling in which the Gantt schedule is

repeatedly updated according to realized processing and transportation states rather than adjusted through simple schedule shifting. This extension is expected to provide a tighter closed loop between production scheduling, OHT dispatching, and real time execution. In addition, SOMOS is evaluated using a literature-based semiconductor FAB testbed and has not been externally validated against industrial production or OHT data. Accordingly, the results should be interpreted within the simulated testbed, and validation using industrial data remains an important direction for future work.

Author Contributions: Conceptualization: J.A., I.-B.P. and K.K. Methodology: J.A. Software and Validation: J.A. and J.B. Formal Analysis and Investigation: J.A., J.B., H.A., S.Y., I.-B.P. and K.K. Data Curation: J.A., I.-B.P. and K.K. Writing—Original Draft Preparation: J.A., I.-B.P. and K.K. Writing—Review and Editing: I.-B.P. and K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Smart Manufacturing Innovation R&D program funded by the Korea Ministry of SMEs and Startups in 2024 (Project No. RS-2024-00433001).

Data Availability Statement: The data, simulation configurations, and source code supporting the findings of this study are publicly available in the GitHub repository at <https://github.com/Jungwu9/Scheduler-and-OHT-allocation-MONitoring-Simulator-SOMOS-> (accessed on 18 July 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aresi, L.; Dauzère-Pérès, S.; Yugma, C.; Ndiaye, M.A.; Ruliere, L. AMHS vehicle management policies in semiconductor manufacturing: A short review. In Proceedings of the 2019 International Conference on Industrial Engineering and Systems Management (IESM), Shanghai, China, 25–27 September 2019; pp. 1–6. [\[CrossRef\]](#)
2. Montoya-Torres, J.R. A literature survey on the design approaches and operational issues of automated wafer-transport systems for wafer fabs. *Prod. Plan. Control* **2006**, *17*, 648–663. [\[CrossRef\]](#)
3. Ahn, K.; Park, J. Cooperative zone-based rebalancing of idle overhead hoist transportations using multi-agent reinforcement learning with graph representation learning. *IIE Trans.* **2021**, *53*, 1140–1156. [\[CrossRef\]](#)
4. Mönch, L.; Fowler, J.W.; Dauzère-Pérès, S.; Mason, S.J.; Rose, O. A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations. *J. Sched.* **2011**, *14*, 583–599. [\[CrossRef\]](#)
5. Graves, S.C.; Meal, H.C.; Stefek, D.; Zeghmi, A.H. Scheduling of re-entrant flow shops. *J. Oper. Manag.* **1983**, *3*, 197–207. [\[CrossRef\]](#)
6. Uzsoy, R.; Lee, C.-Y.; Martin-Vega, L.A. A review of production planning and scheduling models in the semiconductor industry—Part II: Shop-floor control. *IIE Trans.* **1994**, *26*, 44–55. [\[CrossRef\]](#)
7. Drießel, R.; Mönch, L. An integrated scheduling and material-handling approach for complex job shops: A computational study. *Int. J. Prod. Res.* **2012**, *50*, 5966–5985. [\[CrossRef\]](#)
8. Kim, H.; Lim, D.-E.; Lee, S. Deep learning-based dynamic scheduling for semiconductor manufacturing with high uncertainty of automated material handling system capability. *IEEE Trans. Semicond. Manuf.* **2020**, *33*, 13–22. [\[CrossRef\]](#)
9. Park, H.; Lee, M.; Han, J. Heuristic Algorithms for the 1-m-1 Hybrid Flow Shop Scheduling Problem with Lot Streaming, No-Wait, Blocking, and Sequence-Dependent Setup Times. *Systems* **2026**, *14*, 900. [\[CrossRef\]](#)
10. Chang, K.-H.; Huang, Y.-H.; Yang, S.-P. Vehicle fleet sizing for automated material handling systems to minimize cost subject to time constraints. *IIE Trans.* **2014**, *46*, 301–312. [\[CrossRef\]](#)
11. Huang, C.-J.; Chang, K.-H.; Lin, J.T. Optimal vehicle allocation for an automated materials handling system using simulation optimisation. *Int. J. Prod. Res.* **2012**, *50*, 5734–5746. [\[CrossRef\]](#)
12. Shin, Y.; Jeong, Y.; Kim, G. New mathematical formulation for idle OHT vehicle repositioning: Maximization of repositioning success probability. *Eng. Optim.* **2026**, *58*, 934–959. [\[CrossRef\]](#)
13. Lin, J.T.; Wu, C.-H.; Huang, C.-W. Dynamic vehicle allocation control for automated material handling system in semiconductor manufacturing. *Comput. Oper. Res.* **2013**, *40*, 2329–2339. [\[CrossRef\]](#)
14. Egbelu, P.J.; Tanchoco, J.M.A. Characterization of automatic guided vehicle dispatching rules. *Int. J. Prod. Res.* **1984**, *22*, 359–374. [\[CrossRef\]](#)
15. Kim, B.-I.; Shin, J.; Jeong, S.; Koo, J. Effective overhead hoist transport dispatching based on the Hungarian algorithm for a large semiconductor FAB. *Int. J. Prod. Res.* **2009**, *47*, 2823–2834. [\[CrossRef\]](#)

16. Hong, S.; Hwang, I.; Jang, Y.J. Practical Q-learning-based route-guidance and vehicle assignment for OHT systems in semiconductor fabs. *IEEE Trans. Semicond. Manuf.* **2022**, *35*, 385–396. [\[CrossRef\]](#)
17. Pfund, M.E.; Mason, S.J.; Fowler, J.W. Semiconductor manufacturing scheduling and dispatching: State of the art and survey of needs. In *Handbook of Production Scheduling*; Springer: Boston, MA, USA, 2006; pp. 213–241. [\[CrossRef\]](#)
18. Fowler, J.W.; Hogg, G.L.; Mason, S.J. Workload control in the semiconductor industry. *Prod. Plan. Control* **2002**, *13*, 568–578. [\[CrossRef\]](#)
19. Lu, S.C.H.; Ramaswamy, D.; Kumar, P.R. Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants. *IEEE Trans. Semicond. Manuf.* **1994**, *7*, 374–388. [\[CrossRef\]](#)
20. Kumar, P.R. Scheduling semiconductor manufacturing plants. *IEEE Control Syst. Mag.* **1994**, *14*, 33–40. [\[CrossRef\]](#)
21. Spearman, M.L.; Zazanis, M.A. Push and pull production systems: Issues and comparisons. *Oper. Res.* **1992**, *40*, 521–532. [\[CrossRef\]](#)
22. Kiba, J.-E.; Lamiabale, G.; Dauzère-Pérès, S.; Yugma, C. Simulation of a full 300mm semiconductor manufacturing plant with material handling constraints. In Proceedings of the 2009 Winter Simulation Conference (WSC), Austin, TX, USA, 13–16 December 2009; pp. 1601–1609. [\[CrossRef\]](#)
23. Agrawal, G.K.; Heragu, S.S. A survey of automated material handling systems in 300-mm semiconductor fabs. *IEEE Trans. Semicond. Manuf.* **2006**, *19*, 112–120. [\[CrossRef\]](#)
24. Mackulak, G.T.; Savory, P. A simulation-based experiment for comparing AMHS performance in a semiconductor fabrication facility. *IEEE Trans. Semicond. Manuf.* **2001**, *14*, 273–280. [\[CrossRef\]](#)
25. Kong, S.H. Two-step simulation method for automatic material handling system of semiconductor fab. *Robot. Comput.-Integr. Manuf.* **2007**, *23*, 409–420. [\[CrossRef\]](#)
26. Batareseh, O.G.; Nazzal, D.; Wang, Y. An interval-based metamodeling approach to simulate material handling in semiconductor wafer fabs. *IEEE Trans. Semicond. Manuf.* **2010**, *23*, 527–537. [\[CrossRef\]](#)
27. Wan, J.; Shin, H. Predictive vehicle dispatching method for overhead hoist transport systems in semiconductor fabs. *Int. J. Prod. Res.* **2022**, *60*, 3063–3077. [\[CrossRef\]](#)
28. Benzoni, A.; Yugma, C.; Bect, P. Vehicle look-ahead dispatching for overhead hoist transport system in semiconductor manufacturing. *IEEE Trans. Semicond. Manuf.* **2023**, *36*, 130–138. [\[CrossRef\]](#)
29. Schmalzer, R.; Schmidt, T.; Schoeps, M.; Luebke, J.; Hupfer, R.; Schlaus, N. Simulation based evaluation of different empty vehicle management strategies with considering future transport jobs. In Proceedings of the 2017 Winter Simulation Conference (WSC), Las Vegas, NV, USA, 3–6 December 2017; pp. 3576–3587. [\[CrossRef\]](#)
30. Wertz, R.; Fischmann, C.; Böttinger, F.; Kasperczyk, M. Improved empty vehicle balancing in automated material handling systems. In Proceedings of the Tenth International Conference on Computer Modeling and Simulation (UKSim), Cambridge, UK, 1–3 April 2008; pp. 732–733. [\[CrossRef\]](#)
31. Jacobs, J.H.; Etman, L.F.P.; van Campen, E.J.J.; Rooda, J.E. Characterization of operational time variability using effective process times. *IEEE Trans. Semicond. Manuf.* **2003**, *16*, 511–520. [\[CrossRef\]](#)
32. Kuhn, H.W. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **1955**, *2*, 83–97. [\[CrossRef\]](#)
33. Qiu, L.; Hsu, W.J.; Huang, S.Y.; Wang, H. Scheduling and routing algorithms for AGVs: A survey. *Int. J. Prod. Res.* **2002**, *40*, 745–760. [\[CrossRef\]](#)
34. Im, K.; Kim, K.; Park, T.; Lee, S. Effective vehicle dispatching method minimising the blocking and delivery times in automatic material handling systems of 300 mm semiconductor fabrication. *Int. J. Prod. Res.* **2009**, *47*, 3997–4011. [\[CrossRef\]](#)
35. Liao, D.-Y.; Fu, H.-S. Speedy delivery—Dynamic OHT allocation and dispatching in large-scale, 300-mm AMHS management. *IEEE Robot. Autom. Mag.* **2004**, *11*, 22–32. [\[CrossRef\]](#)
36. Ben Chaabane, A.; Dauzère-Pérès, S.; Yugma, C.; Rullière, L.; Lamiabale, G. Analyzing the impact of key parameters of vehicle management policies in a unified AMHS. In Proceedings of the 2013 Winter Simulation Conference (WSC), Washington, DC, USA, 8–11 December 2013; pp. 3818–3828. [\[CrossRef\]](#)
37. Hwang, I.; Jang, Y.J. Q(λ) learning-based dynamic route guidance algorithm for overhead hoist transport systems in semiconductor fabs. *Int. J. Prod. Res.* **2020**, *58*, 1199–1221. [\[CrossRef\]](#)
38. Ahn, K.; Lee, K.; Yeon, J.; Park, J. Congestion-aware dynamic routing for an overhead hoist transporter system using a graph convolutional gated recurrent unit. *IIE Trans.* **2022**, *54*, 803–816. [\[CrossRef\]](#)
39. Chou, C.-W.; Chiu, W.-C.; Hsu, Y.-T. Multiagent reinforcement learning-based dispatching model for overhead hoist transfer in automated material handling system. *Comput. Ind. Eng.* **2025**, *204*, 111109. [\[CrossRef\]](#)
40. Sung, J.-H.; Ju, S.-H.; Cho, S.-W.; Joo, H.-J.; Seo, K.-M.; Kang, B.-G. Modeling and Exploratory Analysis of Discrete Event Simulations for Optimizing Overhead Hoist Transport Systems and Logistics in Semiconductor Manufacturing. *Mathematics* **2025**, *13*, 1167. [\[CrossRef\]](#)
41. Li, H.; Jin, Z. A new look of dispatching for multi-objective interbay AMHS in semiconductor wafer manufacturing: A T-S fuzzy-based learning approach. *Expert Syst. Appl.* **2025**, *262*, 125615. [\[CrossRef\]](#)

42. Koo, B.; Kim, Y.; Choi, J.; Jun, S.; Shin, Y. Alleviation of OHT vehicle congestion in semiconductor FAB with dynamic link weight control: A reinforcement learning approach. *Int. J. Prod. Res.* **2026**. [[CrossRef](#)]
43. Sarin, S.C.; Varadarajan, A.; Wang, L. A survey of dispatching rules for operational control in wafer fabrication. *Prod. Plan. Control* **2011**, *22*, 4–24. [[CrossRef](#)]
44. Tyan, J.C.; Du, T.C.; Chen, J.C.; Chang, I.-H. Multiple response optimization in a fully automated FAB: An integrated tool and vehicle dispatching strategy. *Comput. Ind. Eng.* **2004**, *46*, 121–139. [[CrossRef](#)]
45. Koo, P.-H.; Jang, J.; Suh, J. Vehicle dispatching for highly loaded semiconductor production considering bottleneck machines first. *Int. J. Flex. Manuf. Syst.* **2005**, *17*, 23–38. [[CrossRef](#)]
46. Mourtzis, D.; Papakostas, N.; Mavrikios, D.; Makris, S.; Alexopoulos, K. The role of simulation in digital manufacturing: Applications and outlook. *Int. J. Comput. Integr. Manuf.* **2015**, *28*, 3–24. [[CrossRef](#)]
47. Rahman, M.M.; Fan, J.; Tijan, E.; Grubišić, N. Reinforcement Learning for Congestion Mitigation in Inland Freight Terminals: A Simulation-Based Serial Mediation Analysis of Operational Learning Stability and Logistics Efficiency. *Systems* **2026**, *14*, 743. [[CrossRef](#)]
48. de Koster, R.M.B.M.; Le-Anh, T.; van der Meer, J.R. Testing and classifying vehicle dispatching rules in three real-world settings. *J. Oper. Manag.* **2004**, *22*, 369–386. [[CrossRef](#)]
49. Rose, O. Comparison of due-date oriented dispatch rules in semiconductor manufacturing. In Proceedings of the 2003 Industrial Engineering Research Conference (IERC), Portland, OR, USA, 18–20 May 2003.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.