

CIRP Conference on Manufacturing Systems (CIRP CMS)

AMR-Based Automated Material Handling in Semiconductor Fabs: A Discrete Event Simulation Study

Gaetano Muscatello^{a*}, Matteo Carabelli^a, Tullio Tolio^a, Maria Chiara Magnanini^a^a*Department of Mechanical Engineering, Politecnico di Milano, Milano 20156, Italy** Tel.: +39 3280830148; E-mail address: gaetano.muscatello@polimi.it

Abstract

Given their capital intensity, advanced semiconductor fabs are accelerating automation to improve control efficiency and flexibility in complex production systems. Within this context, intralogistics is pivotal for enhancing high throughput and minimizing lead time. An industrial case study is presented on the design and planned implementation of an Automated Material Handling System (AMHS) based on Autonomous Mobile Robots (AMR) in a production fab. To support this transition, a data-driven sizing of transport demand was first performed through the integration of heterogeneous data sources, and the layout of pick-up/drop-off stations was defined; these elements are then fed directly into a discrete-event simulation (DES) model of the AMHS, which captures wafer-transport dynamics. The proposed methodology leverages a highly reconfigurable and fast-executing DES model, coupled with a Genetic Algorithm (GA), to jointly optimize operational parameters, configuration variables, and routing/dispatching policies with the objective of cost minimization. The GA–DES loop identifies the best combination of AMR fleet setup and control rules by maximizing transport completion rate. The optimization explicitly accounts for battery and buffer capacity constraints, fab traffic, and the stochastic nature of AMR-operator interactions. This results in a simulation-based decision-support framework for smart logistics management and reconfiguration in semiconductor fabs, enabling more effective AMR deployment and higher operational performance.

© 2026 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 2025 CIRP CMS

Keywords: Semiconductor Manufacturing; Autonomous Mobile Robots; Stochastic Modelling; Discrete Event Simulation

1. Introduction

Front-end semiconductor manufacturing is a highly complex production environment, characterized by a job-shop configuration with a very high product mix, long cycle times, and re-entrant flows, where wafers repeatedly visit the same process areas throughout their fabrication [1, 2]. To ensure full utilization of the extremely costly equipment installed in the fab, optimized production planning and control, together with an efficient transport system that guarantees the continuous feeding of tools and prevents any risk of starvation are required [3, 4]. The equipment is typically organized into bays, each hosting tools of the same process family, and wafer lots must be transferred across multiple bays according to their process routes, typically travelling a total of about 20 km and visiting 500 machines [5, 6]. In this context, wafer movement is

generally classified into intra-bay (within the same bay) and inter-bay (between different bays) transport. Among the two, inter-bay transport is more critical, as it handles a larger portion of the material flow, higher transport demand, and a broader range of travel risks [7]. As a result, inter-bay material handling plays a central role in sustaining fab-wide performance, strongly influencing throughput, cycle time, and tool utilization. Automated Material Handling Systems (AMHS) have become essential for inter-bay transport in front-end fabrication [8]. Their importance increased significantly with the transition from 200 mm to 300 mm wafers because the heavier Front Opening Unified Pods (FOUPs) made manual handling ergonomically impractical [9]. Beyond this reason, AMHS were driven by productivity and economic benefits: despite the higher initial investment, automated transport has been shown to outperform manual handling in terms of net

present value cash-flow [10]. Among AMHS technologies, Overhead Hoist Transports (OHTs), automated vehicles moving on fixed overhead tracks, have become the most widely adopted solution in 300 mm semiconductor fabs [11]. However, OHT systems are rarely implemented in 200 mm fabs, as their layouts and physical infrastructures were not originally designed to accommodate overhead tracks, and most equipment did not support automated load/unload interfaces [6]. As a result, material handling in 200 mm facilities has historically relied on manual transport performed by operators [12]. AMRs can autonomously navigate within the fab without the need of fixed infrastructure, move collision-free to any reachable target point using onboard laser-scanning sensors, thus adapting to highly dynamic and evolving industrial environments [13]. These characteristics make AMRs particularly suitable for brownfield 200 mm facilities seeking cost-effective automation while preserving existing equipment and layouts. However, to successfully deploy an AMR fleet in a semiconductor fab, systematic design is required, setting configuration variables, routing and dispatching policies, as well as key operational parameters that jointly determine the overall logistics performance. In existing literature, these design aspects are typically analyzed individually, whereas an integrated approach is far more beneficial for guiding a correct and effective AMR system implementation, since the performance of each component is strongly interdependent with the others. To address this need, a highly reconfigurable discrete-event simulation (DES) model represents an effective solution, as it allows the joint evaluation of these design dimensions while accurately capturing the inter-bay transport dynamics and operational interactions within the fab [14, 15]. From a configuration perspective, fleet sizing has been extensively investigated in the literature through simulation-based approaches [16, 17]. Simulation has also been widely employed to optimize fleet management strategies, supporting the selection of the most effective routing and dispatching policies to enhance AMHS performance [18, 19]. However, configuration aspects and operational aspects are usually studied separately, even though they strongly influence one another in real fabs. This work addresses a case study involving the implementation of an AMR fleet in a 200 mm semiconductor fab. Operational and physical constraints are analyzed to size the loading and unloading stations and define a simplified behavioral model for the AMR fleet. Multiple data sources, including Radio-Frequency Identification (RFID) data, Manufacturing Execution System (MES) data, and the fab layout, are integrated to reconstruct production flows across bays, in order to estimate the steady-state inter-bay transport demand. Based on this information, a discrete-event simulation model is developed to reproduce the internal fab dynamics. The simulator is designed to be modular, reconfigurable, and computationally efficient, allowing rapid evaluation of different solutions. The simulation model is coupled with a genetic algorithm, which explores a large solution space by iteratively testing candidate configurations in terms of fleet size, routing and dispatching policies, and operational parameters such as speed and stop duration. By evaluating these elements together rather than separately, the framework enables a truly integrated joint optimization of configuration

choices, control strategies, and process parameters, reflecting their strong interdependencies in real fab operations. Each candidate solution is evaluated through simulation with respect to transport-demand fulfillment, while strictly enforcing AMR battery constraints and buffer capacity limits, ensuring that only feasible and operationally sustainable configurations are selected. The methodology is detailed in the following sections. Section 2.1 discusses the limitations of the industrial case under study. Section 2.2 presents the approach adopted for transport demand estimation. Section 2.3 outlines the alternative configurations, operational parameters, and policies considered in the simulation model-based evaluation. Finally, the last section reports the numerical results of the proposed approach.

2. Methodology

With the objective of jointly optimizing configuration, operations and policies (COP), Discrete Event Simulation serves as the environment for digital replica of the semiconductor fab, allowing for a reliable representation of the stochastic production conditions. The simulation model must be highly reconfigurable and computationally efficient, enabling extensive multi-scenario analyses. These analyses aim to evaluate overall transport efficiency and estimate queue levels, while explicitly accounting for AMR fleet battery consumption and charging behavior. In this context, simulator modularity is the property of having simulation model with easy re-initialization capability, with simple inputs and adjustable control logics both in terms of operation management and configuration. The efficiency of model evaluation enables the possibility to perform analyses on a wide range of scenarios. This methodology is applied to a real industrial study aimed at optimizing the AMR transportation system, defining the fleet size and proper layout features for operation, in a real 200 mm fab. For the selection of the solution of actual implementation, meta-heuristic optimization via genetic algorithm is exploited. Unlike most existing contributions that treat fleet sizing and dispatching separately, this work encodes configuration, operational parameters and routing rules into a single chromosome structure, allowing cross-interactions to emerge during optimization. This avoids suboptimal decoupled solutions and captures intertwined effects between dwell time, routing thresholds and fleet size.

2.1. System description and operational constraints

The case study involves a 200 mm semiconductor front-end fab characterized by high-volume, high-mix production. Wafers are processed in lots of 25 units, carried inside standardized cassettes. The fab layout develops along a main central corridor, from which multiple bays branch out on both sides. The AMRs are responsible for inter-bay transport along this main corridor. Some peripheral areas of the fab are not suitable for AMR circulation and must remain unaffected by the automated transport system. Each AMR is equipped with a unit holding device, allowing the handling of up to six lots simultaneously. Loading and unloading operations are performed by operators; therefore, the AMRs stop at dedicated pick-up/drop-off stations for a dwell time T_D to allow manual

handling. The interaction between AMRs and operators introduces additional stochasticity into the system. Operators may release an AMR earlier, if handling operations are completed ahead of time, or extend the dwell time when additional manipulation is required. In some cases, the operator may fail to notice the AMR's arrival and perform no action; this event is modelled through an operator miss probability p_{om} which is evaluated empirically for different T_D . Some of the stations, in addition to serving as pick-up and drop-off points, also function as battery-charging locations. The number and placement of stations are defined through structured brainstorming sessions to identify a suitable compromise: operators' walking distance outside the bay caps the number of stations; on the other hand, the stations cannot be too many to avoid AMR idling and corridor congestion. In fact, each station serves multiple adjacent bays, which were grouped to balance the transport flows evaluated in the next section. For balancing purposes, stations were initially arranged alternately on both sides of the main corridor, so that AMRs stop at roughly half of them on the forward path and the remaining half on the return path. However, an alternative configuration can also be assessed, in which selected stations are positioned on both sides of the corridor, allowing AMRs to serve them in both directions. Charging stations were located only where physical layout constraints allowed their installation. At the current stage, automated lot identification and destination tracking on board the AMRs is not feasible. Given the high complexity and interdependency of transport flows, a milk-run fleet strategy (hereafter called bus mode) is initially adopted: AMRs continuously travel back and forth along the main corridor, stopping at every station. This first implementation is selected for its simplicity and robustness, enabling a preliminary performance assessment before deploying more advanced routing strategies. To support the understanding of the system, Figure 1 shows a simplified schematic of a representative section of the fab.

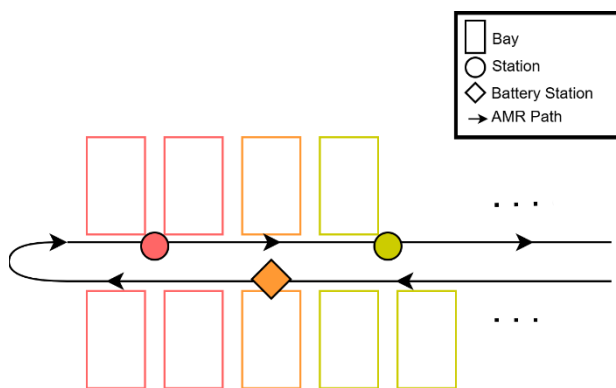


Figure 1. Representative fab section with AMR stations

2.2. Data-Driven Transport Demand Analysis

Multiple data sources are collected and analyzed to quantify the inter-bay transport demand to be fulfilled. One month of historical MES and RFID data is considered, as the two systems provide complementary information: MES records contain process-related events for each lot, while RFID logs provide location-based tracking. The two datasets are merged into a

single integrated database to reconstruct the complete movement history of each lot. In addition, the RFID location coordinates are associated with the coordinates extracted from layout file. This integrated database enables data cleaning and filtering to isolate relevant transport flows. First, movements occurring in portions of the fab that are not intended for AMR circulation are removed through location-based filtering, by excluding transfers associated with coordinates outside the AMR operating perimeter. Second, short-distance transfers between adjacent bays, typically carried out manually by operators within the bay, without using the main corridor, are excluded through an additional location-based filter, so to keep only movements that would realistically require AMR transport. Third, consecutive process steps performed within the same technological area are removed through a process-based filter, as these within-area transfers remain under the responsibility of the bay personnel. After filtering, one month of validated inter-bay moves is analyzed to compute the transport workload between bay pairs. The resulting daily average inter-bay flow matrix is derived and used as the baseline input for the AMHS design and subsequent simulation-optimization activities.

2.3. Simulation model-based mixed configuration-operations metaheuristic optimization

To fully characterize the DES model, systematic input preparation is needed. Inter-bay flow matrix and station demand characterize the initialization of server production details in SIEMENS Tecnomatix Plant Simulation environment setup. The reference notation for the methodology is reported in Table 1.

Table 1. Methodology notation

Symbol	Description
v_c	AMR cruise velocity
T_D	AMR dwell time
$p_{om,j}$	Probability of operator miss at station j
N_{st}	Threshold for standard taxi mode selection
N_{dt}	Threshold for dedicated taxi mode selection
FV_{st}	Boolean flag value for standard taxi mode activation
FV_{dt}	Boolean flag value for dedicated taxi mode activation
N_j	Number of lots headed to station j
C_{AMR}	Single AMR acquisition cost
N_{AMR}	Number of circulating AMR
c_{op}	Unitary cost of unfulfilled transport demand
r	Discount rate
tr_j	Daily number of requested moves from station j
tc_j	Daily number of completed moves from station j

The model sets a fixed number of AMR travelling across the plant layout, encoding the basic controls for transportation flows in the main corridor. The model input is externally sourced, to intuitively manage configuration, policies, and operational features, enabling easy reconfigurability of the simulation environment. Figure 3 outlines the overall framework of the methodology. Orange-colored blocks refer to AMR features, whereas light blue ones pertain to the fab. The initialization action of the simulation entails the setting based on each block related to policies, configuration and operational

parameters, as well as the offline synchronization with real data from the fab (MES, RFID and layout data).

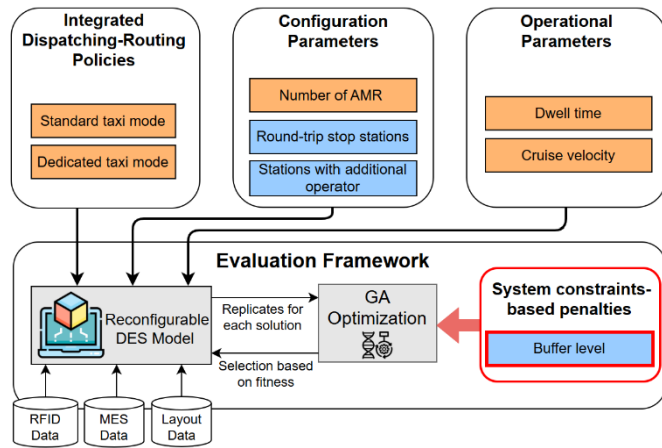


Figure 2. DES model-based evaluation framework

Multi-level inputs feed the simulation model, allowing for high level of model reconfigurability. The simulator is deployed by a Genetic Algorithm (GA) optimization module, which tests the alternatives based on a cost-based fitness function. Penalty is given to the solutions where at least one outlet buffer level of a station reaches the buffer capacity within the simulation period, hindering fab operability. Operational input parameters are upper- and lower-bounded numerical variables, that is, the dwell time (T_D) and cruise velocity (v_c). Configuration parameters include AMR fleet size and the allocation of pick-up/drop-off points on both sides of the corridor near each station. Another option is characterized by the allocation of a single operator of the current workforce to a specific station j . Given a system configuration, this setting is not allowed for more than half of the total stations, due to production constraints. This option has the beneficial effect of reducing $T_{D,j}$ and $p_{om,j}$, while increasing $p_{om,i \neq j}$.

Dispatching policies have been tightly coupled with routing policies. Although scientific literature highlights the efficiency of high-level dynamic dispatching strategies in AMHS [20], the constraints of the industrial case study led to adopting a simple, yet effective heuristic rule derived from expert insights. Whenever AMR_k stops at station, the lots in the pick-up buffer are grouped by destination. The destination groups are then ordered in decreasing size. Within each destination group, lots are loaded according to a First-In First-Out logic. Table 2 describes the routing procedure depending on active policies. The default policy is *Bus mode*, which entails a cyclic visit of stations along the loop. Alternative routing policies are evaluated at every AMR stop, after the pick-up operation. The pick-up ends when the AMR is full or the buffer is empty. If the number of lots onboard directed to station, namely N_j , exceeds the specific threshold, the condition for selecting an alternative policy is satisfied. Both alternative policies are activated only if their corresponding Boolean flag value is set to true. The first policy is *Standard taxi mode* (active if FV_{st} is set to true), where the AMR is scheduled to reach the selected destination (station j) by the shortest path, as fast as traffic allows. On the other hand, *dedicated taxi mode* (active if FV_{dt} is set to true) represents standard taxi mode specifically set for

the lowest traffic station, i.e., the one requesting the lowest daily number of moves. Dedicated taxi mode is considered for the peculiarity of the case study: in fact, the lowest traffic station requests significantly less moves compared to the other stations. This station is visited by an AMR only if the selection condition is satisfied. If none of the policies are activated or selected, standard Bus mode is used. Table 2 reports the policy selection algorithm executed for each AMR stop.

Table 2. Pseudocode of routing policy selection

Algorithm 1

```

Initialize dispatching list, AMR content,  $FV_{st}$ ,  $N_{st}$ ,  $FV_{dt}$ ,  $N_{dt}$ 
Perform lots pick-up in AMR
Compute  $N_j \forall j$ 
If  $FV_{dt} = \text{true}$ 
    If  $N_j \geq N_{dt}$ ,  $j = \text{lowest-traffic station}$ 
        Select Dedicated taxi mode and leave
If  $FV_{st} = \text{true}$ 
    For  $j = 1, \dots, \#stations$ 
        If  $N_j \geq N_{st}$ 
            Collect  $j$  as candidate destination
    If  $\#candidate\ destinations > 1$ 
        Pick  $j$  of the lot with earliest loading time
        Select Standard taxi mode with destination  $j$  and leave
    Elseif  $\#candidate\ destinations = 1$ 
        Select Standard taxi mode with destination  $j$  and leave
    Else
        Select Bus mode and leave

```

The discrete-event simulator can quickly execute the selected COP scenario feeding directly the GA module, whose objective function is expressed below:

$$\min OF = N_{AMR} \cdot C_{AMR} + \sum_{j=1}^2 365 \cdot c_{op} \cdot (\sum_k tr_{avg,k} - \sum_k tc_{avg,k}) \cdot \frac{1}{(1+r)^j} \quad (1)$$

The scope is to minimize the total AMHS-related cost. The first term represents the fixed investment needed to acquire the AMR fleet, while the second term accounts for two years of discounted operating costs generated by unfulfilled transportation. Missing transports delay lots in reaching their next processing step, increasing their lead time and, by Little's Law, the average WIP of the fab. This translates into additional stock-holding costs, captured through the parameter C_{op} , which represents the marginal cost of delaying an average lot and is derived from historical company data. It multiplies the gap between the total transportation requests and those fulfilled, evaluated with reference to the worst daily performance observed over the simulation week. This gap corresponds to the average number of pending requests at day-end and quantifies the economic penalty due to transportation unfulfillment. The simulator models AMR battery consumption dynamics: as soon as the AMR goes on reserve, routing to out-of-the-loop recharge station occurs. Buffer blocking is handled through a penalty-based constraint approach: when any station buffer exceeds its capacity, a large additive penalty is applied to fitness function value, to significantly lower the probability of selection of the solution; otherwise, no penalty is applied. The next section describes the results of the evaluation.

3. Numerical results

The Genetic Algorithm module has been set in accordance with the proposed framework, considering the large space of solutions. Table 3 lists the optimization variables as they are formalized in the simulator environment.

Table 3. Optimization parameters setting description

Area	Optimization parameter	Range	Encoding
Configuration	Number of AMR	5-10	Discrete
	Round-trip stop stations	true/false $\forall$ station	Boolean
	Station with fixed operator	true/false $\forall$ station	Boolean
Operations	v_c [m/s]	0.4-0.9	Discretized continuous: $\Delta=0.1$
	T_D [min]	3-7	Discretized continuous: $\Delta=0.5$
Routing Policies	N_{st} [lots]	1-6	Discrete
	N_{dt} [lots]	1-6	Discrete
	FV_{st}	true/false	Boolean
	FV_{dt}	true/false	Boolean

A highly combinatorial decision space results from problem formulation. Each individual encodes the complete COP solution in a mixed integer–binary chromosome, including fleet size, station-related Boolean variables, discretized continuous operational parameters (cruise velocity and dwell time), routing thresholds and activation flags. The maximum cruise velocity is bounded by an industrially defined quality threshold. Values exceeding v_{c-max} were identified by the company as increasing the risk of product quality degradation beyond acceptable limits. The explored fleet-size range reflects an industrially defined feasibility window derived from preliminary sizing assessments, excluding configurations deemed either insufficient to satisfy demand or impractical in terms of investment and congestion. In the simulation each AMR has been modelled with reliability behavior characterized by frequent micro-stops slowing down operations, due to safety stoppages made because of the frequent interaction with people and obstacles in the fab. The availability parameters were derived from a reference fab operating with a similar AMR-based transport system under comparable operational conditions. These parameters define the stochasticity of the system, along with the actual dwell time at stations, and the oscillating daily transport demand at stations. Preliminary convergence tests showed that beyond the fifth run the estimated standard deviation was stabilized. Therefore, five replications were considered sufficient to ensure robust performance comparison. The simulation length has been set to one week plus two hours of ramp-up. The algorithm is implemented in Tecnomatix Plant Simulation, version 2404.0009. Table 4 outlines GA features.

Table 4. GA parameters

GA parameter	Value
Number of generations	30
Generation size	20
Number of best individuals	3
Number of individual replicates	5
Crossover rate	0.8
Mutation rate	0.5

Uniform crossover enables recombination across the mixed discrete–Boolean COP variables, while mutation is applied independently to each gene, flipping Boolean values and randomly re-sampling discrete parameters. The relatively high mutation rate was selected to prevent premature convergence given the limited population size and the highly combinatorial COP decision space. For the same reason, elitism has not been applied. Parents set was defined by roulette-wheel selection among the top individuals in both parents' and children's arrays, with a controlled selection pressure. The number of generations was fixed a priori as a computational budget; although the best solution emerged at generation 18, additional generations confirmed solution stability. Figure 3 shows the results of the evaluation.

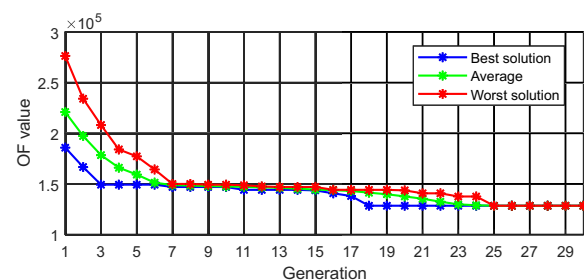


Figure 3. GA evolution plot

The evolution of the best, average and worst solutions shows the enhancement produced by the GA iterations. The overall runtime of the genetic algorithm was approximately two hours on a workstation equipped with an Intel® Core™ Ultra 9 185H processor, confirming the computational feasibility of the approach for industrial-scale optimization. AMHS total cost minimization, across the whole generations' frame, resulted in the COP solution described in Table 5.

Table 5. Best COP solution

COP parameter	Value
N_{AMR}	8
Round-trip stop stations	true for 23% of stations
Stations with fixed operator	true for 30% of stations
v_c [m/s]	0.9
T_D [min]	3.5
N_{st} [lots]	3
N_{dt} [lots]	1
FV_{st}	true
FV_{dt}	true

GA results displayed a near-optimal region in the COP decision space, where multiple configurations show overlapping confidence intervals. Nevertheless, in stochastic optimization context, the objective is to minimize the expected value of the cost function, and the selected COP configuration

is not dominated in expectation by any alternative candidate in the region of near optimality. It is worth to be noted that COP configurations of other top near-optimal solutions show the same features of the selected one, except slight variations. Although the numerical values have been scaled for industrial confidentiality, these relative differences indicate that the selected solution provides a superior cost-performance trade-off across the entire search space, with objective function value 1.5% lower than the second-best configuration and approximately 31.8% lower than the average OF value across all explored alternatives. In the best-performing configuration, the dwell time at stations tends to the lowest feasible value, minimizing unnecessary stops and reducing congestion along the corridor. Regarding routing parameters, the threshold for the Standard taxi mode is set to 3, while for the Dedicated taxi mode it is set to 1, reflecting a selective prioritization strategy. Moreover, the station-allocated operator and round-trip stop station logics are activated in stations with the highest transport demand. The near-optimal COP solution of the optimization problem jointly (i) guarantees the correct operation of the system, (ii) designs the configuration for the inter-bay areas, (iii) suggests the routing policies to implement, satisfying the battery and capacity limitations.

4. Conclusions and future work

The proposed simulation-optimization framework, coupling a highly reconfigurable discrete-event simulator with a genetic algorithm, proved to be computationally efficient in exploring a wide range of COP solutions within reasonable times. The integrated approach successfully identified near-optimal combinations of configuration variables, routing policies, and operational parameters that minimize cost while satisfying battery and buffer constraints. The AMR fleet is currently entering its operational phase, and real-fab data collection will enable the refinement of stochastic parameters such as operator miss probability and station availability. Preliminary observations suggest a correlation between the frequency of micro-stops caused by obstacle encounters and the AMR's cruise velocity. Once sufficient statistical evidence is gathered, this relationship will be incorporated into the model, allowing the identification of non-trivial velocity settings that balance efficiency and reliability. The developed modular framework will serve as a digital decision-support tool, allowing the continuous evaluation and updating of near-optimal operating conditions and control policies as the system evolves over time. As future work, the framework could upgrade into a Digital Twin architecture, continuously fed by real-time data and enabling state-dependent, dynamically updated routing policies instead of static rule sets. Advanced visualization tools such as Digital Shadows or AR/VR environments may further support operator interaction and real-time decision-making.

Acknowledgements

This work was funded by European Union NextGenerationEU, M4C2 I1.1, Progetto PRIN 2022 "MAGELLANO", Prot. 2022NRMS4X, CUP D53D23005440001.

References

- [1] Magnanini MC, Djurdjanovic D, Pomi R, Tolio T. System-level evaluation of productivity and quality in semiconductor frontend fabrication integrating product and process models. *CIRP J Manuf Sci Technol* 2025;60:296-306.
- [2] Mönch L, Fowler JW, Mason SJ. *Production planning and control for semiconductor wafer fabrication facilities: modeling, analysis, and systems*. 2nd ed. New York: Springer; 2013.
- [3] Hong T-Y, Chien C-F, Chen H-P. UNISON framework of system dynamics-based technology acquisition decision for semiconductor manufacturing and an empirical study. *Comput Ind Eng* 2023;182:109012.
- [4] A total standard WIP estimation method for wafer fabrication — Y. H. Lin & C.-E. Lee, *European Journal of Operational Research*, vol. 131, no. 1, pp. 78–94, 2001.
- [5] Wang, K. J., Lin, J. T., & Weigert, G. (2007). Agent-based interbay system control for a single-loop semiconductor manufacturing fab. *Production Planning & Control*, 18(2), 74–90.
- [6] Vehicle Look-Ahead Dispatching for Overhead Hoist Transport System in Semiconductor Manufacturing, A. Benzoni, G. C. Yugma & P. Bect, *IEEE Transactions on Semiconductor Manufacturing*, Feb 2023.
- [7] Pan, C., Zhang, J. & Qin, W. Real-time OHT Dispatching Mechanism for the Interbay Automated Material Handling System with Shortcuts and Bypasses. *Chin. J. Mech. Eng.* 30, 663–675 (2017).
- [8] G. K. Agrawal and S. S. Heragu, "A Survey of Automated Material Handling Systems in 300-mm Semiconductor Fabs," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 1, pp. 112–125, February 2006.
- [9] Wang, MJ, Chung, HC and Wu, HC. 2003. The evaluation of manual FOUP handling in 300-mm wafer fab. *IEEE Trans. Semicon. Manuf.*, 16 : 551 – 554.
- [10] Montoya-Torres, J. R. (2006). A literature survey on the design approaches and operational issues of automated wafer-transport systems for wafer fabs. *Production Planning & Control*, 17(7), 648–663.
- [11] Chatterjee, S., Jeong, T., Kim, Y., & Lee, S. (2025). A multi-objective study on OHT scheduling with fleet sizing and waiting strategies for semiconductor fabs through evolutionary algorithms. *IIE Annual Conference.Proceedings*, 1-6.
- [12] M. Haddadin and W. Moreno, "Automated Material Handling Systems: System of Systems Architecture Examination Semiconductor Manufacturing Perspective," *2021 32nd Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, Milpitas, CA, USA, 2021, pp. 1-6.
- [13] Fragapane, G., Sgarbossa, F., Strandhagen, J.O., & Strandhagen, J.W., "Review of Autonomous Mobile Robots in Intralogistics: State-of-the-Art, Limitations and Research Gaps", 2024.
- [14] Terkaj, W., Tolio, T., & Urgo, M. (2015). A virtual factory approach for in situ simulation to support production and Maintenance Planning. *CIRP Annals*, 64(1).
- [15] Fowler JW, Mönch L, Ponsignon T. Discrete-event simulation for semiconductor wafer fabrication facilities: a tutorial. *Int J Ind Eng* 2015;22:661-682.
- [16] Chang, K. H., Y. H. Huang, and S. P. Yang. 2014a. "Vehicle Fleet Sizing for Automated Material Handling Systems to Minimize Cost Subject to Time Constraints." *IIE Transactions* 46(3): 301–312.
- [17] Chang, K.-H. 2022. "Applying Simulation Optimization for Agile Vehicle Fleet Sizing of Automated Material Handling Systems in Semiconductor Manufacturing." *International Journal of Industrial Engineering: Theory, Applications and Practice* 29(2): 179–198.
- [18] Sung, C. L. 2025. "Modeling and Exploratory Analysis of Discrete-Event Simulations for Optimizing Overhead Hoist Transport Systems and Logistics in Semiconductor Manufacturing." *Journal of Simulation Modelling Practice and Theory*, 145: 103123.
- [19] Kovács, B., Tassel, P., and Gebser, M. 2023. "Optimizing Dispatching Strategies for Semiconductor Manufacturing Facilities with Genetic Programming." *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '23)*. ACM, 1374-1382.
- [20] Wei, Q., Zilong, Z., Yaoming, Z., Yinbin, S. 2021. "Dynamic dispatching for interbay automated material handling with lot targeting using improved parallel multiple-objective genetic algorithm", *Computers & Operations Research*, Volume 131, 1052.