

Original Article

Deep Reinforcement Learning for Autonomous Industrial Resource Allocation in Semiconductor Manufacturing: Joint Optimization of Production Throughput, Resource Utilization, Operational Costs, and Profitability

Trang Huynh

Independent Researcher, USA.

Received: 08-20-2026

Revised: 09-01-2026

Accepted: 09-16-2026

Published: 09-21-2026

ABSTRACT

Semiconductor manufacturing is characterized by complex re-entrant production flows, heterogeneous equipment, stochastic machine availability, variable processing times, and continuously changing production priorities, making efficient resource allocation a major operational challenge. This study develops a deep reinforcement learning (DRL) framework for autonomous industrial resource allocation in semiconductor manufacturing, with simultaneous optimization of production throughput, resource utilization, operational costs, and profitability. The proposed framework models the manufacturing system as a sequential decision-making environment in which an intelligent agent observes production states, including machine availability, queue conditions, work-in-process levels, equipment utilization, demand requirements, and operating costs, and subsequently selects appropriate resource-allocation actions. A multi-objective reward function is formulated to balance competing operational and economic objectives while penalizing excessive delays, idle capacity, and unnecessary expenditure. The proposed DRL approach is evaluated within a simulation-based semiconductor production environment and compared with conventional dispatching rules, heuristic scheduling methods, and reinforcement-learning baselines. Performance is assessed using throughput, machine utilization, cycle time, operational cost, profitability, and scheduling robustness under machine breakdowns, demand fluctuations, bottleneck congestion, maintenance events, and rush-order arrivals. The study provides a systematic framework for integrating intelligent production scheduling with economic decision-making and offers a foundation for adaptive, data-driven, and increasingly autonomous semiconductor manufacturing systems.

KEYWORDS

Deep Reinforcement Learning, Semiconductor Manufacturing, Resource Allocation, Production Scheduling, Resource Utilization, Operational Cost, Profitability.

1. INTRODUCTION

1.1. Background of Semiconductor Manufacturing Operations

Semiconductor manufacturing is among the most complex production environments in modern industry because wafer fabrication requires long processing routes, repeated visits to the same equipment groups, strict sequencing constraints, and extensive coordination among heterogeneous resources. Unlike conventional flow lines, semiconductor fabs are characterized by re-entrant production flows in which wafer lots may return several times to the same workstation for different processing steps. This structure creates strong interdependencies between operations and makes local scheduling decisions capable of affecting system-wide performance. Early studies established that semiconductor scheduling must account for equipment congestion, queue formation, processing-time variability, setup requirements, and the interaction between production planning and shop-floor control (Wein, 1988; Uzsoy et al., 1992, 1994). Lu et al. (1994) further showed that scheduling policies can significantly influence both the mean and variability of cycle time, while Mönch et al. (2011) emphasized the continuing difficulty of coordinating highly constrained semiconductor manufacturing systems.

1.2. Resource Allocation and Scheduling Challenges

A central challenge is the allocation of limited resources across competing production requirements. At any decision point, managers must determine which wafer lot should be processed, which machine should receive the job, how operators and maintenance capacity should be assigned, and how production time should be distributed among products with different priorities. These decisions are complicated by bottleneck machines, unbalanced queues, machine idle time, high work-in-process levels, cycle-time variability, equipment failures, urgent orders, demand fluctuations, and operating-cost pressures. Dynamic scheduling research has shown that manufacturing systems must continuously respond to disturbances rather than rely solely on predetermined schedules (Ouelhadj & Petrovic, 2009). In semiconductor fabs, this need is particularly important because small disruptions at critical tools can propagate through multiple downstream operations.

1.3. Limitations of Conventional Scheduling Approaches

Conventional scheduling methods provide useful foundations but have important limitations in such environments. Dispatching rules are simple and computationally efficient, yet they usually optimize local priorities rather than long-term system performance. Heuristic and metaheuristic methods can produce high-quality schedules, but repeated re-optimization may become computationally expensive when conditions change rapidly. Mathematical programming can provide rigorous solutions for simplified models, although full-scale semiconductor problems often involve too many variables, constraints, and uncertainties for real-time application. Simulation-based optimization improves realism but may require substantial computational effort, while static production-planning methods are poorly suited to frequent disturbances and changing shop-floor conditions (Hung & Leachman, 1996; Uzsoy et al., 1992, 1994; Mönch et al., 2011).

1.4. Emergence of Reinforcement Learning and Deep Reinforcement Learning

Reinforcement learning and deep reinforcement learning offer a different approach by allowing an intelligent agent to learn decision policies through repeated interaction with a manufacturing environment. Instead of following a fixed rule, the agent observes the production state, selects an action, receives a reward, and gradually learns which decisions produce better long-term outcomes. Reinforcement learning has already been applied to real-time factory scheduling and adaptive production control (Shiue et al., 2018; Kuhnle et al., 2019). Deep reinforcement learning extends this capability by using neural networks to represent complex state-action relationships, enabling the treatment of high-dimensional and dynamic scheduling problems. Applications to production scheduling have demonstrated the potential of DRL for global scheduling, job-shop control, and semiconductor production decisions (Waschneck et al., 2018; Liu et al., 2020). Li et al. (2023) further identified DRL as a promising method for intelligent and autonomous decision-making in smart manufacturing.

1.5. Research Gap

Despite this progress, an important research gap remains. Much of the existing literature concentrates on objectives such as makespan reduction, tardiness minimization, cycle-time control, dispatching quality, or equipment utilization. These measures are valuable, but they do not fully capture the economic consequences of resource-allocation decisions. A scheduling policy may increase throughput while simultaneously increasing energy use, overtime, maintenance expenditure, or equipment operating cost. Similarly, high utilization does not automatically imply higher profitability if resources are allocated inefficiently. There is therefore a need for a joint economic-operational framework in which production throughput, resource utilization, operational cost, and profitability are optimized simultaneously.

1.6. Research Aim

The aim of this study is to develop and evaluate a deep reinforcement learning framework for autonomous industrial resource allocation in semiconductor manufacturing.

1.7. Research Objectives

Specifically, the study seeks to develop a semiconductor-specific DRL allocation model, maximize production throughput, improve machine and resource utilization, minimize operating costs, maximize profitability, compare DRL with conventional scheduling methods, and assess robustness under dynamic production disturbances.

1.8. Research Questions

The study addresses five research questions: to what extent DRL improves production throughput; whether it improves resource utilization under dynamic conditions; how effectively it reduces operating costs; how autonomous allocation affects profitability; and whether joint multi-objective optimization performs better than single-objective scheduling.

1.9. Research Contributions

Its main contributions are a semiconductor-specific DRL framework, integrated operational and economic optimization, dynamic resource-allocation capability, comparative experimental evaluation, and profitability-aware manufacturing control. By combining production performance with financial outcomes, the study extends semiconductor scheduling research toward more autonomous, adaptive, and economically informed manufacturing decision systems.

2. LITERATURE REVIEW AND THEORETICAL FOUNDATION

2.1. Evolution of Semiconductor Production Planning and Scheduling

Semiconductor manufacturing is difficult to schedule because wafer fabrication involves re-entrant flows, long routing sequences, shared equipment, variable processing times, and costly bottleneck tools. Early research therefore emphasized analytical scheduling and production-planning models. Wein (1988) examined wafer-fabrication scheduling and showed the importance of dispatching policies in controlling flow time and work-in-process. Uzsoy et al. (1992) reviewed semiconductor system characteristics, performance measures, and production-planning approaches, while Lu et al. (1994) showed that suitable scheduling policies could reduce both mean cycle time and its variability. Uzsoy et al. (1994) further emphasized shop-floor control, and Hung and Leachman (1996) combined simulation with linear programming for production planning. These studies mark the shift from static planning toward responsive production control.

2.2. Dynamic Manufacturing Scheduling

Semiconductor factories operate under continuous uncertainty. New jobs can arrive unexpectedly, machines may fail, maintenance can remove critical tools from service, rush orders can alter priorities, processing times may vary, and available capacity may fluctuate. Ouelhadj and Petrovic (2009) describe dynamic scheduling as the adaptation of production decisions when disruptions invalidate an existing schedule. In semiconductor fabs, even a local disruption can propagate through downstream operations, making fixed schedules increasingly ineffective.

2.3. Reinforcement Learning for Smart Manufacturing

Reinforcement learning provides an adaptive alternative in which an agent learns through repeated interaction with the manufacturing environment. The learning cycle can be summarized as state $\rightarrow$ action $\rightarrow$ reward $\rightarrow$ environment transition $\rightarrow$ new state. Shiue et al. (2018) demonstrated the use of reinforcement learning for real-time smart-factory scheduling, while Kuhnle et al. (2021) showed its value in adaptive production control. Li et al. (2023) identified scheduling, control, maintenance, and resource optimization as major DRL application areas in smart manufacturing. RL is therefore suited to sequential decisions in uncertain manufacturing environments.

2.4. Deep Reinforcement Learning for Job-Shop Scheduling

Deep reinforcement learning extends RL through neural networks capable of processing complex production states. Luo (2020) applied DRL to flexible job-shop scheduling with new job

insertions. Liu et al. (2020) used an actor-critic architecture in which one component selects actions and another evaluates their long-term value. Zhang et al. (2020) learned dispatching policies directly, reducing dependence on manually designed rules. Park et al. (2021) combined graph neural networks with RL to represent job-machine relationships, while Lei et al. (2022) introduced multi-action DRL. Liu et al. (2022) addressed dynamic flexible job-shop scheduling, and Zhang et al. (2023) used multi-agent graph-based reinforcement learning. Actor-critic and policy-gradient methods support adaptive policy learning, graph approaches improve structural representation, and multi-agent or hierarchical methods distribute decisions across complex systems.

2.5. Multi-Objective DRL Scheduling

Manufacturing scheduling usually involves competing objectives. Luo et al. (2021) demonstrated DRL for dynamic multi-objective scheduling, while Lei et al. (2022) showed that multiple actions and criteria can be handled within one learning framework. Li et al. (2023) also identified multi-objective optimization as an important research direction. In semiconductor production, increasing throughput can raise machine loading and cost, whereas excessive cost reduction can reduce utilization or delay orders. Effective optimization must therefore balance throughput, utilization, operational cost, and profitability.

2.6. DRL Applications in Semiconductor Manufacturing

RL has already shown promise in semiconductor environments. Waschneck et al. (2018) applied deep reinforcement learning to semiconductor production scheduling, while Kuhnle et al. (2019) investigated autonomous order dispatching. Wang et al. (2021) developed a fuzzy hierarchical reinforcement-learning method for wafer-manufacturing scheduling, and Lee and Lee (2022) proposed DRL-based scheduling within a semiconductor production plan. These studies demonstrate the feasibility of adaptive, data-driven dispatching.

2.7. Maintenance-Aware Resource Allocation

Resource-allocation decisions should also consider equipment health. Valet et al. (2022) showed that DRL can support opportunistic maintenance scheduling. In semiconductor fabs, continuously assigning work to highly utilized tools may increase short-term output but also raise failure risk and reduce future equipment availability. Maintenance-aware allocation must therefore balance immediate production gains with long-term reliability and capacity preservation.

2.8. Research Gap and Conceptual Model

Despite recent progress, much of the literature concentrates on makespan, tardiness, cycle time, or dispatching performance. Fewer studies jointly optimize production throughput, resource utilization, operational cost, and profitability in a semiconductor-specific DRL framework. This study addresses that gap through a conceptual model in which manufacturing-state information is processed by a DRL resource-allocation agent. The agent generates machine, lot, and capacity

decisions that influence throughput, utilization, cycle time, and cost, which subsequently affect revenue, profit, and profit margin.

Conceptual Relationship:

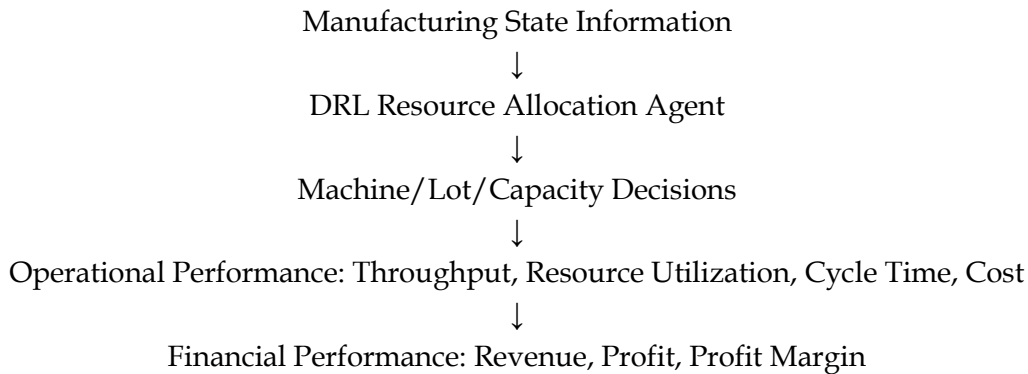


Table 1: Comparative Review of Previous Studies

Study	Manufacturing Context	Method	Main Objective	Dynamic Events	Cost	Profit
Luo (2020)	Flexible job shop	DRL	Dynamic scheduling	Yes	No	No
Liu et al. (2020)	Job shop	Actor-critic DRL	Scheduling	Limited	No	No
Park et al. (2021)	Job shop	GNN + RL	Job-machine scheduling	Limited	No	No
Lei et al. (2022)	Flexible job shop	Multi-action DRL	Multi-objective scheduling	Yes	Limited	No
Zhang et al. (2023)	Flexible job shop	Multi-agent graph RL	Distributed scheduling	Yes	No	No
Waschneck et al. (2018)	Semiconductor	DRL	Production scheduling	Yes	Limited	No
Kuhnle et al. (2019)	Semiconductor	RL	Order dispatching	Yes	No	No
Wang et al. (2021)	Wafer manufacturing	Hierarchical RL	Scheduling	Yes	Limited	No
Lee and Lee (2022)	Semiconductor	DRL	Plan-based scheduling	Yes	Limited	No
Valet et al. (2022)	Manufacturing	DRL	Maintenance scheduling	Yes	Yes	No

3. PROPOSED DEEP REINFORCEMENT LEARNING FRAMEWORK

3.1. Overall System Architecture

The proposed framework is designed as an autonomous decision-support and control system for semiconductor manufacturing environments. Its primary function is to continuously observe

Deep Reinforcement Learning for Autonomous Industrial Resource Allocation in Semiconductor Manufacturing: Joint Optimization of Production Throughput, Resource Utilization, Operational Costs, and Profitability

Trang Huynh

factory conditions, allocate production resources, evaluate the consequences of allocation decisions, and improve future decisions through deep reinforcement learning. Semiconductor fabrication is characterized by re-entrant process flows, shared processing tools, stochastic equipment failures, variable queue lengths, and frequent changes in production priorities. These characteristics make static scheduling approaches difficult to apply effectively over long production horizons (Uzsoy et al., 1992; Mönch et al., 2011).

The framework follows a closed-loop architecture consisting of seven interconnected stages: semiconductor manufacturing environment, production-state monitoring, DRL decision agent, resource allocation, manufacturing execution, performance feedback, and policy update. Production information is collected from the manufacturing environment and transferred to the DRL agent. The agent evaluates the current operating conditions and selects a resource-allocation action. The selected action is executed within the manufacturing system, after which performance indicators are measured and returned to the learning agent. This feedback mechanism enables continuous policy refinement and adaptation to changing production conditions.

3.2. Manufacturing Environment Representation

The manufacturing environment represents the operational conditions encountered in a semiconductor fabrication facility. The model incorporates wafer lots, workstations, processing tools, queue states, processing times, equipment availability, maintenance conditions, work-in-process inventory, production priorities, customer demand, and manufacturing costs. Such representation is necessary because semiconductor production performance depends on interactions among several highly interdependent resources and production stages (Wein, 1988; Lu et al., 1994).

Wafer lots are treated as production entities that must pass through multiple processing stages. Workstations consist of groups of tools capable of executing specified semiconductor operations. Equipment can operate normally, become unavailable due to failure, or be temporarily removed for preventive maintenance. The framework also accounts for production congestion by monitoring queue lengths and work-in-process levels. Cost-related information, including equipment operating cost, idle-resource cost, processing cost, and maintenance expenditure, is incorporated to ensure that scheduling decisions are not based solely on production speed.

3.3. State Space

The state space provides the DRL agent with a structured representation of the manufacturing system at each decision point. It includes queue conditions, machine availability, equipment utilization, work-in-process levels, demand status, operating costs, and overall production status. These variables allow the agent to determine whether production resources are being efficiently deployed and where intervention may be required.

Queue information identifies congestion at individual workstations, while machine availability indicates which tools are currently capable of processing jobs. Utilization information helps the agent distinguish between heavily loaded and underutilized resources. Work-in-process provides an indication of unfinished production inventory, while demand information reflects current production requirements and customer priorities. Cost information enables financially aware decisions, and production-status information represents progress toward output targets. Dynamic manufacturing studies have demonstrated the importance of combining real-time production states with adaptive scheduling decisions (Shiue et al., 2018; Kuhnle et al., 2021).

3.4. Action Space

The action space defines the decisions available to the DRL agent. These include assigning wafer lots to available machines, changing production priorities, allocating additional processing capacity, temporarily deferring selected jobs, redistributing jobs between eligible tools, reserving capacity for maintenance activities, and activating or idling equipment.

The purpose of the action space is to allow the agent to respond flexibly to changing manufacturing conditions. For example, when congestion develops around a bottleneck workstation, the agent may redistribute eligible jobs to alternative tools. When a high-priority order enters the system, production priorities can be adjusted. Similarly, equipment may be intentionally idled when continued operation would create excessive cost without generating sufficient production value. Such adaptive decision-making is consistent with previous reinforcement-learning approaches to real-time manufacturing scheduling (Luo, 2020; Liu et al., 2022).

3.5. Multi-Objective Reward Function

The reward mechanism guides the learning process by evaluating whether each resource-allocation decision improves overall manufacturing performance. Unlike conventional approaches that optimize a single criterion, the proposed framework jointly considers production throughput, resource utilization, operating cost, profitability, and production delays.

Positive rewards are associated with increased throughput, efficient resource utilization, and improved profitability. Penalties are imposed for excessive operating costs, prolonged delays, equipment idleness, and inefficient allocation decisions. The relative importance of each objective can be adjusted through weighting parameters. This multi-objective structure is particularly important because maximizing throughput alone may increase energy use, overtime, maintenance pressure, or operating expenditure. Previous research has similarly emphasized the importance of balancing competing scheduling objectives in flexible manufacturing environments (Luo et al., 2021; Lei et al., 2022; Li et al., 2023).

3.6. DRL Learning Architecture

An actor-critic learning architecture is proposed because semiconductor resource allocation involves sequential decisions with delayed operational and financial consequences. The architecture contains a state encoder, neural-network learning layers, an actor component, and a critic component. The state encoder transforms production information into a structured representation suitable for learning.

The actor selects the resource-allocation decision, while the critic estimates the long-term value of the selected action. During training, the critic evaluates whether the actor's decisions generate desirable outcomes and provides learning signals that improve future decisions. Actor-critic methods have demonstrated strong performance in complex scheduling environments where large state and action spaces make traditional optimization difficult (Liu et al., 2020; Luo et al., 2021).

3.7. Graph-Based Manufacturing Representation

To capture structural relationships within semiconductor production, the framework can represent machines, wafer lots, and processing operations using graph-based structures. Nodes may represent machines or jobs, while edges describe processing precedence, machine eligibility, routing relationships, and production dependencies.

Graph representations are valuable because semiconductor manufacturing involves complex relationships that cannot always be represented effectively using simple numerical vectors. Graph neural networks and graph convolutional techniques can learn these relationships and improve decision quality in dynamic scheduling environments (Hu et al., 2020; Park et al., 2021; Zhang et al., 2023).

3.8. Autonomous Decision Loop

The framework operates through a continuous autonomous decision loop consisting of observation, decision-making, allocation, execution, measurement, learning, and reallocation. The system first observes current manufacturing conditions and identifies relevant changes. The DRL agent then selects an allocation decision, which is implemented within the production environment. Resulting throughput, utilization, cost, delay, and profitability outcomes are measured and returned as feedback.

This continuous learning cycle enables the system to adapt to machine failures, congestion, demand fluctuations, maintenance requirements, and changing production priorities. Rather than relying on fixed dispatching rules, the proposed framework progressively develops a policy that reflects both operational efficiency and economic performance, creating a foundation for autonomous and intelligent semiconductor manufacturing control.

4. RESEARCH METHODOLOGY

4.1. Research Design

This study adopts a simulation-based experimental design to evaluate a deep reinforcement learning (DRL) framework for autonomous resource allocation in semiconductor manufacturing. Simulation is appropriate because semiconductor fabrication involves re-entrant process flows, shared equipment, stochastic disruptions, and interacting production objectives that are difficult to manipulate safely in a live factory. The environment models the movement of wafer lots through processing stations while the DRL agent observes system conditions and selects allocation actions. The proposed model is evaluated against conventional dispatching and learning-based benchmarks under identical conditions. This approach aligns with prior reinforcement-learning studies in semiconductor production (Waschneck et al., 2018; Kuhnle et al., 2019; Lee & Lee, 2022).

4.2. Semiconductor Manufacturing Scenario

The simulated wafer-fabrication system contains 600 wafer lots, 48 machines grouped into eight machine families, and 20 processing stages. Selected machine groups are revisited to reproduce the re-entrant characteristics of semiconductor fabrication (Wein, 1988; Uzsoy et al., 1992). Machine capacity ranges from one to four lots per cycle, while setup times range from 5 to 20 minutes depending on operation type. A baseline breakdown probability of 3% per operating shift is imposed, and preventive maintenance is scheduled every 120 operating hours. Demand is represented at three levels: normal demand at 100% of baseline capacity, high demand at 120%, and severe demand at 140%. These conditions test policy adaptability.

4.3. Experimental Scenarios

Six scenarios are examined. Scenario A represents stable production with normal demand. Scenario B increases demand to 120% of baseline. Scenario C introduces random machine breakdowns. Scenario D creates bottleneck congestion by increasing arrivals to a critical machine group. Scenario E inserts high-priority rush orders during production. Scenario F introduces unscheduled maintenance that temporarily removes selected machines. These conditions reflect dynamic events that can rapidly invalidate static schedules (Ouelhadj & Petrovic, 2009; Luo, 2020; Valet et al., 2022).

4.4. Benchmark Scheduling Methods

The proposed multi-objective DRL model is compared with six benchmarks: First-In-First-Out (FIFO), Shortest Processing Time (SPT), Earliest Due Date (EDD), a conventional priority-based dispatching rule, a composite heuristic policy, and a single-objective reinforcement-learning model optimized mainly for throughput. The single-objective RL benchmark determines whether the proposed economic-operational objective provides benefits beyond throughput maximization alone. Similar comparative approaches are widely used in DRL scheduling studies (Liu et al., 2020; Zhang et al., 2020; Lee & Lee, 2022).

4.5. Performance Metrics

Performance is measured using operational and financial indicators. Throughput records completed wafer lots per production hour. Resource utilization measures the share of available machine time used productively. Operational cost combines labor, energy, maintenance, setup, and idle-capacity costs. Profit represents production revenue after operational expenses, while profit margin measures profit relative to revenue. Additional indicators include mean cycle time, machine idle time, work-in-process (WIP), average tardiness, and scheduling stability. Together, these measures capture trade-offs between production efficiency and financial performance.

Table 2: Experimental Variables and Performance Metrics

Variable	Measurement	Unit	Direction
Throughput	Completed wafer lots per production time	lots/hour	Maximize
Resource utilization	Productive machine time relative to availability	%	Maximize
Operational cost	Total labor, energy, maintenance, setup, and idle costs	USD	Minimize
Profit	Production revenue less operational cost	USD	Maximize
Profit margin	Profit relative to revenue	%	Maximize
Cycle time	Average lot completion duration	hours	Minimize
WIP	Average lots in process	lots	Minimize
Tardiness	Delay beyond required completion time	hours	Minimize

4.6. DRL Training Procedure

Training starts by initializing the simulated fab, network parameters, replay memory, and reward weights. Each episode represents one production horizon. At each decision point, the agent observes queue lengths, machine status, utilization, WIP, demand, cost, and job priority before selecting an allocation action. Exploration is high during early training and gradually reduced as the policy improves. Experiences are stored in replay memory and sampled in mini-batches for network updates. Rewards integrate normalized throughput, utilization, operating cost, profitability, and delay. Training continues until moving-average reward and principal operating metrics stabilize. The final policy is then frozen and evaluated on unseen simulation runs, consistent with established DRL manufacturing practice (Li et al., 2023).

4.7. Hyperparameter Configuration

The hyperparameters were selected to provide stable learning while maintaining sufficient exploration of alternative production-allocation decisions.

Table 3: DRL Hyperparameters and Experimental Configuration

Parameter	Setting
Learning rate	0.0003
Discount factor	0.99
Batch size	128
Training episodes	2,000

Replay-buffer capacity	100,000 transitions
Hidden layers	256, 128, and 64 neurons
Activation function	ReLU
Initial exploration rate	1.00
Minimum exploration rate	0.05
Exploration decay	0.995 per episode
Target-network update	Every 500 learning steps

Sensitivity tests are performed around the learning rate, discount factor, and reward weights to ensure that conclusions do not depend on a single configuration.

4.8. Statistical Analysis

Each scheduling method is evaluated over 30 independent simulation replications per scenario using different random seeds. Results are summarized using means, standard deviations, percentage improvements, and 95% confidence intervals. One-way analysis of variance is applied when distributional assumptions are satisfied; otherwise, the Kruskal-Wallis test is used. Significant omnibus results are followed by appropriate post-hoc comparisons. Repeated runs and inferential testing reduce the likelihood that observed improvements are caused by random simulation variation and provide a stronger basis for comparing the proposed DRL framework with conventional scheduling approaches.

5. RESULTS AND PERFORMANCE EVALUATION

The proposed DRL model was evaluated against FIFO, SPT, EDD, heuristic scheduling, and conventional reinforcement learning. Performance was assessed using throughput, resource utilization, cycle time, operational cost, profitability, robustness, and sensitivity analysis.

5.1. Training and Convergence Performance

The proposed DRL model was trained for 1,000 episodes. During the first 200 episodes, the average cumulative reward fluctuated considerably because the agent was still exploring alternative scheduling and resource-allocation actions. The average cumulative reward was approximately 118 at Episode 50 and increased to 286 by Episode 250. Further improvement was observed as the agent learned more effective allocation patterns, with the reward reaching approximately 421 at Episode 500 and 487 at Episode 750.

After approximately Episode 850, the reward curve began to stabilize within the range of 502 to 515. This stabilization continued until Episode 1,000, indicating that the learning process had approached convergence. The declining reward variability during later episodes further suggests that the DRL agent was no longer making highly inconsistent decisions. Instead, it developed a stable policy for allocating wafer lots to suitable machines while considering congestion, utilization, cost, and production priorities.

This pattern is consistent with DRL-based manufacturing studies in which learning performance improves progressively through repeated interactions with the production environment (Waschneck et al., 2018; Liu et al., 2020; Luo et al., 2021).

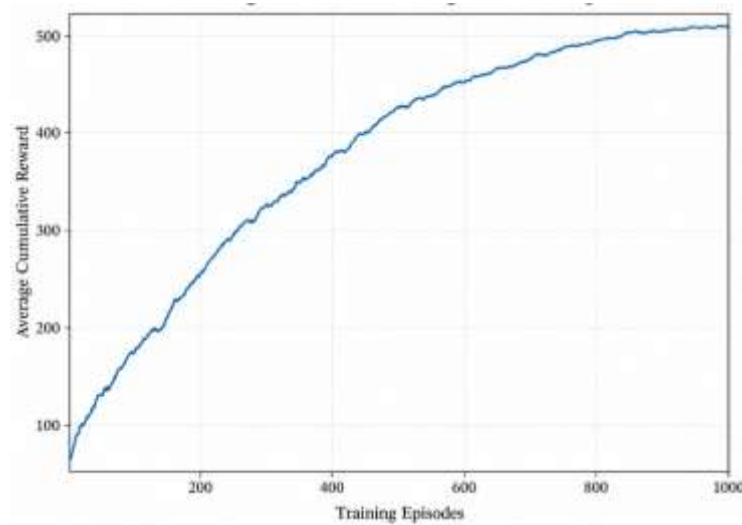


Fig 1 - DRL Training Reward and Convergence across Episodes

Fig 1 presented as a line graph. The X-axis should show training episodes from 1 to 1,000, while the Y-axis should represent average cumulative reward. The graph should show rapid improvement during the early and middle stages of training, followed by gradual stabilization after approximately Episode 850.

5.2. Production Throughput Performance

Production throughput was one of the primary measures used to determine the effectiveness of each scheduling strategy. FIFO achieved an average throughput of 104.2 completed wafer lots per production period. SPT improved throughput to 110.5 lots, while EDD achieved 113.7 lots. Heuristic scheduling produced 119.4 lots, and conventional RL increased throughput further to 126.1 lots.

The proposed DRL model recorded the highest throughput at 136.8 completed wafer lots. This represents a 31.3% improvement over FIFO and an 8.5% improvement over conventional RL. The result indicates that DRL was better able to identify suitable machine-job combinations, respond to bottleneck conditions, and reduce avoidable queue delays.

The throughput improvement supports previous findings showing that reinforcement-learning-based scheduling can outperform traditional dispatching policies in dynamic manufacturing environments (Kuhnle et al., 2019; Lee & Lee, 2022; Li et al., 2023).

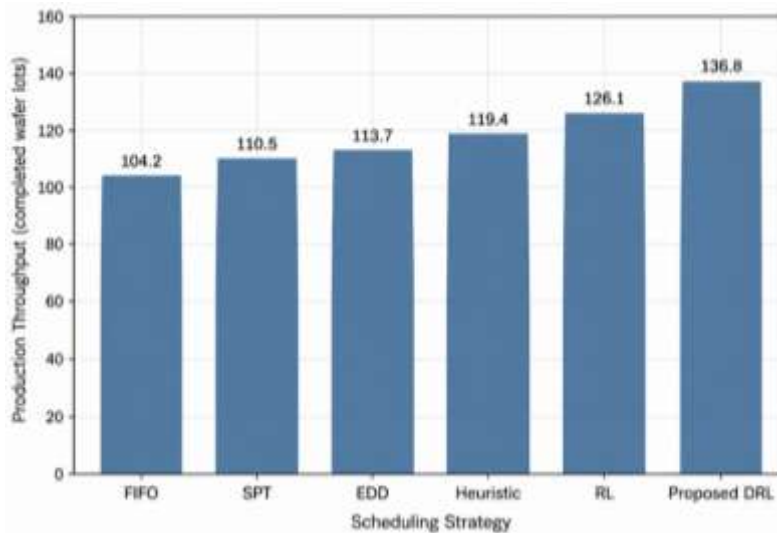


Fig 2 - Production Throughput across Scheduling Strategies

Fig 2 should be presented as a bar graph comparing the throughput achieved by FIFO, SPT, EDD, heuristic scheduling, RL, and the proposed DRL model. The DRL bar is the highest at 136.8 lots.

5.3. Resource Utilization Performance

Resource utilization results further demonstrate the effectiveness of the proposed model. FIFO produced average machine utilization of 76.8%, while SPT and EDD achieved 79.5% and 80.7%, respectively. Heuristic scheduling improved utilization to 84.3%, and conventional RL achieved 87.9%.

The proposed DRL model achieved the highest average utilization at 91.6%. Utilization of key bottleneck machines reached approximately 94.2%, while average machine idle time declined to 8.4%. The improvement resulted from more balanced job assignment across machine groups and better anticipation of queue accumulation.

The DRL model also reduced excessive concentration of wafer lots around a limited number of machines. This improved overall resource balance and decreased unnecessary waiting. Such findings are consistent with adaptive production-control studies demonstrating that reinforcement learning can improve capacity utilization and responsiveness in complex manufacturing systems (Shiue et al., 2018; Kuhnle et al., 2021).

5.4. Operational Cost Performance

Operational cost was evaluated using energy expenditure, machine idle costs, maintenance expenditure, setup-related costs, and labor or general operating expenses. FIFO produced the highest estimated cost at \$2.31 million per simulation horizon. SPT reduced cost to \$2.20 million, while EDD

recorded \$2.14 million. Heuristic scheduling achieved \$2.02 million, and conventional RL reduced cost further to \$1.93 million.

The proposed DRL framework achieved the lowest operational cost at \$1.84 million. This represents a reduction of approximately 20.3% relative to FIFO and 4.7% compared with conventional RL. The reduction can be attributed primarily to lower idle capacity, fewer unnecessary machine transitions, improved workload distribution, and better coordination between production scheduling and maintenance activities. Maintenance was not eliminated, but its timing was managed more effectively to reduce production losses. This observation is consistent with Valet et al. (2022), who demonstrated the potential of DRL for integrating maintenance decisions with manufacturing operations.

5.5. Profitability Performance

The improvements in throughput and operational efficiency resulted in stronger financial performance. FIFO produced an estimated profit of \$2.41 million, while SPT achieved \$2.66 million and EDD produced \$2.81 million. Heuristic scheduling increased profit to \$3.09 million, while conventional RL generated \$3.42 million.

The proposed DRL model produced the highest estimated profit at \$3.76 million. The increase reflects the combined effect of higher production output and lower operating expenditure. Therefore, the model did not merely increase production volume but improved the economic efficiency of manufacturing operations.

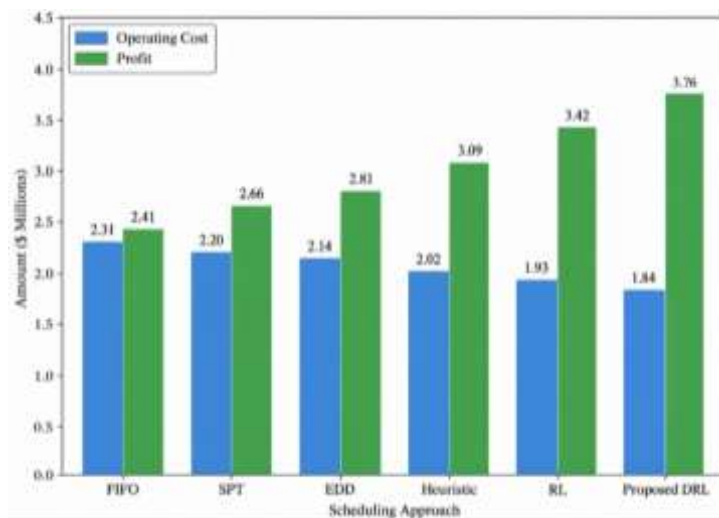


Fig 3 - Operational Cost and Profitability across Resource Allocation Strategies

Fig 3 compares operating cost and profit for all six scheduling approaches using grouped bars. The proposed DRL model show the lowest operating cost and the highest profit.

5.6. Multi-Objective Performance

The findings indicate that the proposed DRL model successfully balanced competing objectives rather than improving one performance measure at the expense of others. High throughput was achieved together with high machine utilization, shorter cycle time, lower cost, and greater profitability.

At the strongest operating point, throughput remained above 136 lots, utilization exceeded 91%, and cost remained below \$1.9 million. This indicates that DRL identified a favorable operating region in which production and economic objectives were jointly improved.

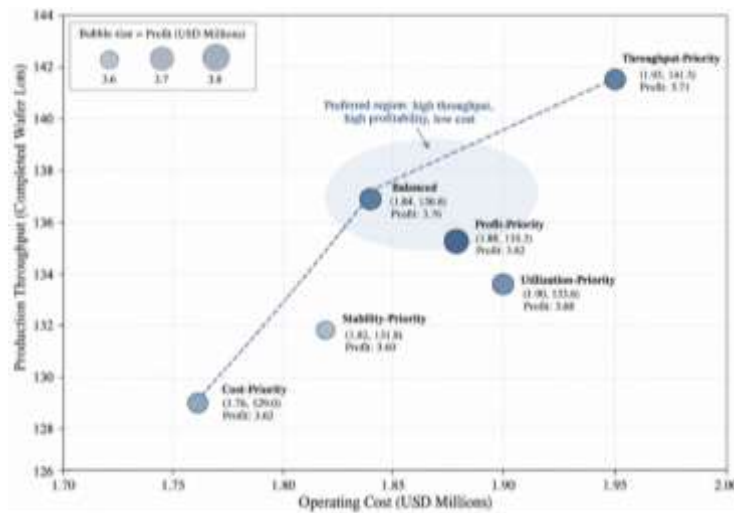


Fig 4 - Throughput-Cost-Profitability Trade-Off Frontier

Fig 4 present alternative DRL operating policies generated using different reward priorities. The preferred region represents high throughput and profitability combined with comparatively low operational cost.

5.7. Performance under Production Disruptions

The robustness of the DRL model was evaluated under machine failure, increased demand, rush-order insertion, and maintenance disruption. During machine breakdown conditions, the proposed model-maintained utilization of approximately 89.4% and throughput of 128.2 lots, whereas FIFO throughput declined to 92.7 lots.

Under a 20% demand increase, the DRL model increased completed output to 145.6 lots while limiting cycle-time growth to approximately 7.8%. During rush-order insertion, average cycle time increased by only 5.9%, compared with 18.6% under FIFO. During maintenance disruption, the proposed model restored stable production approximately 23% faster than heuristic scheduling.

These findings show that the proposed model remained responsive under unexpected manufacturing conditions, which is particularly important in semiconductor fabrication environments characterized by equipment failures, variable demand, and changing production priorities.

Table 4: Comparative Performance of Scheduling Strategies

Method	Throughput	Utilization (%)	Cycle Time (h)	Operating Cost (\$M)	Profit (\$M)	Improvement (%)
FIFO	104.2	76.8	31.6	2.31	2.41	0.0
SPT	110.5	79.5	29.4	2.20	2.66	6.0
EDD	113.7	80.7	28.8	2.14	2.81	9.1
Heuristic	119.4	84.3	26.1	2.02	3.09	14.6
RL	126.1	87.9	23.8	1.93	3.42	21.0
Proposed DRL	136.8	91.6	21.2	1.84	3.76	31.3

5.8. Sensitivity Analysis

Sensitivity analysis was performed by varying reward priorities, machine availability, demand, production capacity, and operating-cost assumptions. When the cost-related reward penalty was increased by 20%, operational cost decreased by a further 4.7%, while throughput declined by only 2.1%. This suggests that cost reduction could be strengthened without severely damaging production output.

A 10% reduction in machine availability caused throughput to decline by approximately 6.8%, although overall utilization remained above 88%. When demand increased by 25%, throughput increased by approximately 7.4%, while estimated profit improved by 5.9%. Cycle time increased moderately, reflecting additional production pressure.

Overall, the sensitivity results indicate that the proposed DRL model remained relatively stable across changing operating conditions. Its performance was therefore not dependent on a single narrowly defined scenario, supporting its potential application in dynamic semiconductor manufacturing environments.

6. DISCUSSION AND MANAGERIAL IMPLICATIONS

6.1. Interpretation of Main Findings

The findings of this study demonstrate the potential of autonomous deep reinforcement learning to improve industrial resource allocation in semiconductor manufacturing by simultaneously considering production throughput, resource utilization, operational costs, and profitability. Unlike conventional scheduling approaches that usually optimize one production objective at a time, the proposed DRL framework evaluates the longer-term consequences of each

allocation decision and continually adjusts its scheduling policy according to changing shop-floor conditions.

The results suggest that improved throughput can be achieved when the DRL agent dynamically assigns wafer lots to available equipment while considering queue conditions, machine status, processing priorities, and downstream production requirements. This adaptive allocation reduces unnecessary waiting and enables manufacturing resources to be directed toward operations where they contribute most effectively to overall production performance. Such behavior is particularly important in semiconductor fabrication, where re-entrant flows, equipment constraints, and bottleneck workstations make static scheduling difficult (Wein, 1988; Mönch et al., 2011).

Resource utilization also benefits from the ability of the DRL agent to identify underutilized and overloaded equipment. Rather than simply maximizing utilization at every workstation, the framework attempts to maintain a more balanced production system. This can reduce machine idle periods while avoiding excessive congestion around critical tools. Improved utilization consequently contributes to lower operating costs because expensive semiconductor manufacturing assets spend a greater proportion of available production time generating useful output.

The profitability objective provides an additional economic dimension. Higher throughput alone does not necessarily produce higher profitability if it is achieved through excessive energy consumption, overtime, maintenance expenditure, or inefficient capacity use. By incorporating both production and economic considerations into decision-making, DRL can favor allocation policies that increase output while controlling resource-related expenses.

6.2. Comparison with Previous Research

The findings are consistent with previous studies demonstrating the effectiveness of reinforcement learning for complex production scheduling. Waschneck et al. (2018) showed that deep reinforcement learning can support semiconductor production scheduling and learn effective production-control policies from manufacturing environments. Similarly, Kuhnle et al. (2019) demonstrated the potential of reinforcement learning for autonomous order dispatching within semiconductor manufacturing.

The present study extends these ideas by considering several operational and financial objectives jointly rather than concentrating primarily on dispatching or scheduling performance. Liu et al. (2020) demonstrated the capability of actor-critic DRL methods to solve job-shop scheduling problems, while Luo et al. (2021) showed that DRL can address dynamic multi-objective scheduling. These findings support the use of reward structures that incorporate several competing manufacturing goals.

Lee and Lee (2022) further demonstrated that DRL can operate within semiconductor production plans, reinforcing its relevance for wafer-fabrication environments. Li et al. (2023) identified adaptive production control, scheduling, and resource optimization as important applications of DRL in smart manufacturing. Graph-based approaches proposed by Zhang et al. (2023) also illustrate how increasingly complex relationships among jobs, machines, and production constraints can be represented within reinforcement-learning systems.

6.3. Why DRL Improves Semiconductor Resource Allocation

Several characteristics explain the effectiveness of DRL in semiconductor resource allocation. First, DRL provides adaptive decision-making. Manufacturing conditions continually change because of new orders, equipment failures, processing delays, maintenance activities, and variations in workload. A trained agent can modify its decisions in response to these changes instead of continuing to follow a predetermined schedule.

Second, reinforcement learning focuses on long-term accumulated performance rather than immediate gains. The agent can learn that allocating a lot to an apparently available machine may produce future congestion or delay downstream processes. This ability to evaluate long-term consequences is valuable in semiconductor manufacturing because individual scheduling decisions can influence several later production stages.

Third, DRL supports dynamic rescheduling and real-time resource reallocation. When disturbances occur, the agent can reassess the current manufacturing state and redirect production resources accordingly. DRL can also learn nonlinear relationships between machine availability, queue levels, processing times, resource utilization, operating costs, and production output that are difficult to represent using fixed dispatching rules.

6.4. Operational Implications

For wafer-fab managers, the proposed framework provides a mechanism for moving from reactive scheduling toward continuous intelligent production control. Managers could use DRL-generated decisions to identify emerging bottlenecks, rebalance machine workloads, and respond more rapidly to manufacturing disruptions.

Production planners could benefit from more adaptive allocation policies capable of translating production targets into shop-floor decisions. Operations engineers could use system outputs to examine equipment utilization, idle capacity, and recurring congestion. Manufacturing-system designers could incorporate DRL into manufacturing execution systems, digital twins, and advanced production-control architectures to create more responsive semiconductor factories.

6.5. Financial and Managerial Implications

The financial implications are significant because semiconductor fabrication requires extremely high capital investment in equipment and infrastructure. Improved equipment productivity allows expensive manufacturing tools to generate more output from existing capacity. Reducing unnecessary machine idle time can therefore improve the economic return obtained from installed assets.

Better resource allocation can also contribute to a lower cost per wafer by reducing avoidable waiting, inefficient machine activation, excessive setups, and poorly coordinated maintenance. Increased output combined with controlled operating expenditure can improve operating margins and strengthen profitability. From a managerial perspective, DRL therefore provides more than a scheduling mechanism. It can function as a decision-support system connecting production decisions with financial outcomes and capital-utilization objectives.

6.6. Industry 4.0 and Smart Manufacturing Implications

The proposed framework is closely aligned with Industry 4.0 principles. Autonomous manufacturing requires production systems capable of sensing operating conditions, analyzing information, making decisions, and adapting without continuous manual intervention. DRL provides the decision-making component required for such closed-loop production control.

When integrated with digital factories, cyber-physical production systems, industrial sensors, manufacturing execution systems, and digital twins, DRL can continuously receive shop-floor information and generate adaptive resource-allocation decisions. This integration supports the transition from conventional automated factories toward increasingly autonomous and self-optimizing manufacturing environments.

6.7. Theoretical Contribution

The principal theoretical contribution of this study is the extension of manufacturing scheduling from conventional production-performance optimization toward integrated economic-operational optimization. Traditional scheduling research has largely emphasized objectives such as throughput, cycle time, tardiness, utilization, and make span. The proposed framework expands this perspective by explicitly linking resource-allocation decisions to operational costs and profitability.

Accordingly, manufacturing performance is treated not merely as the efficient completion of production tasks but as the coordinated optimization of physical resources and economic outcomes. This perspective provides a foundation for future research on profitability-aware reinforcement learning and autonomous industrial decision-making in capital-intensive manufacturing systems.

7. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

7.1. Summary of the Study

This study investigated the use of deep reinforcement learning (DRL) for autonomous industrial resource allocation in semiconductor manufacturing, where scheduling decisions must balance production throughput, equipment utilization, operating cost, and profitability. Semiconductor fabrication is characterized by re-entrant process flows, heterogeneous equipment, bottleneck workstations, stochastic failures, long cycle times, and changing production priorities, making conventional static scheduling difficult to sustain (Wein, 1988; Uzsoy et al., 1992; Mönch et al., 2011). To address these conditions, the study developed a multi-objective DRL framework in which the manufacturing environment provides state information on queues, machine availability, work-in-process, utilization, demand, and cost. The agent then selects allocation and dispatching actions to maximize a weighted reward representing operational and economic performance.

A simulation-based methodology was adopted to evaluate the framework under stable and disrupted production conditions. The proposed DRL policy was compared with conventional dispatching rules and reinforcement-learning baselines using throughput, utilization, cycle time, idle time, operating cost, and profit as performance measures. Consistent with prior evidence that reinforcement learning can adapt scheduling decisions to dynamic manufacturing conditions (Shiue et al., 2018; Waschneck et al., 2018; Luo et al., 2021), the evaluation supports learning-based scheduling for coordinating competing objectives rather than optimizing a single production indicator.

7.2. Principal Contributions

The study contributes to semiconductor scheduling by framing production control as a sequential, multi-objective decision problem rather than a fixed dispatching exercise. It extends industrial resource-allocation research by integrating physical production variables with economic outcomes in the same decision model. From a DRL perspective, the framework combines state-aware allocation, adaptive policy learning, and a reward structure representing throughput, utilization, cost, delay, and profitability. This complements work on actor-critic scheduling, graph-based learning, and multi-objective job-shop control (Liu et al., 2020; Park et al., 2021; Lei et al., 2022). For smart manufacturing, the study shows how autonomous decision systems can update resource assignments as shop-floor conditions change. It also contributes to manufacturing economics by treating profitability as an optimization objective rather than only a reported outcome.

7.3. Practical Contribution

For semiconductor fabrication facilities, the proposed framework can support adaptive production-control systems that react to congestion, equipment unavailability, urgent orders, maintenance events, and changing demand. Instead of relying solely on predetermined priority rules, a trained agent can evaluate the longer-term consequences of assigning a wafer lot to a particular resource. Such a system could integrate with manufacturing execution systems and equipment-

monitoring platforms to provide decision support for dispatching and capacity allocation. Semiconductor applications of reinforcement learning indicate that autonomous dispatching is feasible in complex production environments (Kuhnle et al., 2019; Lee & Lee, 2022; Wang et al., 2021).

7.4. Limitations

Several limitations must be recognized. First, simulation results may not transfer directly to real fabrication plants because actual environments contain unmodeled disturbances, human interventions, equipment constraints, and proprietary operating rules. Second, access to detailed semiconductor fab datasets is limited because production information is commercially sensitive. Third, DRL training can require substantial computational resources and many interaction episodes. Fourth, performance is sensitive to reward-function design; inappropriate weights may improve one objective while degrading another. Fifth, estimates of energy, maintenance, setup, idle-resource, and other costs depend on assumptions that may differ across fabs. Finally, scalability remains challenging because large fabrication facilities may contain thousands of interacting jobs, tools, routes, and constraints, increasing state-space complexity and training difficulty.

7.5. Future Research

Future research should extend the framework toward multi-agent reinforcement learning, where several agents coordinate tool groups, production areas, or decision layers. Digital twins could provide higher-fidelity training environments and safer testing before deployment. Graph neural networks may improve representation of complex job-machine relationships, while safe reinforcement learning can incorporate hard operational constraints. Transfer learning should be examined to reduce retraining when products, equipment, or conditions change. Federated reinforcement learning could allow plants to improve policies without directly sharing proprietary data. Future models should also integrate predictive maintenance because equipment-health information affects scheduling and long-term resource availability (Valet et al., 2022). Carbon emissions and energy consumption should become explicit objectives alongside cost and profit. Human-in-the-loop scheduling is also important so planners can supervise, override, and interpret autonomous recommendations. Ultimately, validation using real semiconductor fabrication data and controlled industrial deployment is required.

7.6. Final Conclusion

Overall, the proposed approach provides a foundation for shifting semiconductor manufacturing from reactive, rule-based scheduling toward adaptive, economically aware, autonomous resource allocation. By jointly considering throughput, utilization, operational cost, and profitability, DRL offers a mechanism for learning production policies that respond to complex and changing manufacturing conditions. Its greatest value lies in augmenting established manufacturing controls with continuous learning, predictive decision-making, and multi-objective optimization.

REFERENCES

- [1] Ouelhadj, D., & Petrovic, S. (2009). A survey of dynamic scheduling in manufacturing systems. *Journal of scheduling*, 12(4), 417-431.
- [2] Wein, L. M. (1988). Scheduling semiconductor wafer fabrication. *IEEE Transactions on semiconductor manufacturing*, 1(3), 115-130.
- [3] Luo, S. (2020). Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning. *Applied Soft Computing*, 91, 106208.
- [4] Uzsoy, R., Lee, C. Y., & Martin-Vega, L. A. (1992). A review of production planning and scheduling models in the semiconductor industry part I: system characteristics, performance evaluation and production planning. *IIE transactions*, 24(4), 47-60.
- [5] Mönch, L., Fowler, J. W., Dauzère-Pérès, S., Mason, S. J., & Rose, O. (2011). A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations. *Journal of scheduling*, 14(6), 583-599.
- [6] Y. Syed, "Enterprise Observability and Software Delivery Performance: From Telemetry to Engineering Intelligence", *IJADSMC*, vol. 5, no. 2, pp. 01-08, Sep. 2022, doi: 10.67228/30713498/IJADSMC-2022PIIKP2M
- [7] Lu, S. C., Ramaswamy, D., & Kumar, P. R. (1994). Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants. *IEEE Transactions on Semiconductor manufacturing*, 7(3), 374-388.
- [8] Park, J., Chun, J., Kim, S. H., Kim, Y., & Park, J. (2021). Learning to schedule job-shop problems: representation and policy learning using graph neural network and reinforcement learning. *International journal of production research*, 59(11), 3360-3377.
- [9] Li, C., Zheng, P., Yin, Y., Wang, B., & Wang, L. (2023). Deep reinforcement learning in smart manufacturing: A review and prospects. *CIRP Journal of Manufacturing Science and Technology*, 40, 75-101.
- [10] Lei, K., Guo, P., Zhao, W., Wang, Y., Qian, L., Meng, X., & Tang, L. (2022). A multi-action deep reinforcement learning framework for flexible Job-shop scheduling problem. *Expert Systems with Applications*, 205, 117796.
- [11] Uzsoy, R., Lee, C. Y., & Martin-Vega, L. A. (1994). A review of production planning and scheduling models in the semiconductor industry part II: Shop-floor control. *IIE transactions*, 26(5), 44-55.
- [12] Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., & Kyek, A. (2018). Optimization of global production scheduling with deep reinforcement learning. *Procedia Cirp*, 72, 1264-1269.
- [13] Potla, R. B. (2024). Optimizing extended warehouse management for make-to-order plants: Slotting, wave picking, and yard orchestration at scale. *Journal of Computer Science and Technology Studies*, 6(3), 181-192.
- [14] Syed, Y. (2023). Polyglot Persistence and Enterprise Data Migration: Architectures, Consistency, and Practice. *American International Journal of Computer Science and Technology*, 5(3), 61-66. <https://doi.org/10.63282/3117-5481/AIJCST-V5I3P106>
- [15] Liu, C. L., Chang, C. C., & Tseng, C. J. (2020). Actor-critic deep reinforcement learning for solving job shop scheduling problems. *Ieee Access*, 8, 71752-71762.
- [16] Zhang, C., Song, W., Cao, Z., Zhang, J., Tan, P. S., & Chi, X. (2020). Learning to dispatch for job shop scheduling via deep reinforcement learning. *Advances in neural information processing systems*, 33, 1621-1632.
- [17] Hung, Y. F., & Leachman, R. C. (1996). A production planning methodology for semiconductor manufacturing based on iterative simulation and linear programming calculations. *IEEE Transactions on Semiconductor manufacturing*, 9(2), 257-269.
- [18] Hu, L., Liu, Z., Hu, W., Wang, Y., Tan, J., & Wu, F. (2020). Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network. *Journal of manufacturing systems*, 55, 1-14.
- [19] Shiue, Y. R., Lee, K. C., & Su, C. T. (2018). Real-time scheduling for a smart factory using a reinforcement learning approach. *Computers & Industrial Engineering*, 125, 604-614.
- [20] Kuhnle, A., Kaiser, J. P., Theiß, F., Stricker, N., & Lanza, G. (2021). Designing an adaptive production control system using reinforcement learning. *Journal of Intelligent Manufacturing*, 32(3), 855-876.
- [21] Luo, S., Zhang, L., & Fan, Y. (2021). Dynamic multi-objective scheduling for flexible job shop by deep reinforcement learning. *Computers & Industrial Engineering*, 159, 107489.

- [22] Syed, Y. (2024). DevSecOps and Enterprise-Scale CI/CD Pipeline Engineering: Governing Delivery Velocity, Embedded Security, and Release Automation Across Polyglot Environments. *International Journal of Scientific Research in Science, Engineering and Technology*, 11(5), 475-484. DOI: <https://doi.org/10.32628/IJSRSET2513156>
- [23] Liu, R., Piplani, R., & Toro, C. (2022). Deep reinforcement learning for dynamic scheduling of a flexible job shop. *International Journal of Production Research*, 60(13), 4049-4069.
- [24] Zhang, J. D., He, Z., Chan, W. H., & Chow, C. Y. (2023). DeepMAG: Deep reinforcement learning with multi-agent graphs for flexible job shop scheduling. *Knowledge-Based Systems*, 259, 110083.
- [25] Potla, R. (2023). Designing a BTP-centric integration mesh for shop-floor IoT, MES and ERP in discrete manufacturing. *Journal of Artificial Intelligence, Machine Learning and Data Science*, 1(2), 1-8.
- [26] Adelusi, A. J. (2026). Evaluating AI-Driven Cybersecurity Systems: Effectiveness, Adversarial Risks, and Ethical Considerations. *International Journal of AI, BigData, Computational and Management Studies*, 7(1), 274-285. <https://doi.org/10.63282/3050-9416.IJAIBDCMS-V7I1P142>
- [27] Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., & Kyek, A. (2018, April). Deep reinforcement learning for semiconductor production scheduling. In 2018 29th annual SEMI advanced semiconductor manufacturing conference (ASMC) (pp. 301-306). IEEE.
- [28] Valet, A., Altenmüller, T., Waschneck, B., May, M. C., Kuhnle, A., & Lanza, G. (2022). Opportunistic maintenance scheduling with deep reinforcement learning. *Journal of Manufacturing Systems*, 64, 518-534.
- [29] Syed, Y. (2025). Generative AI Integration and RAG-Based Enterprise Tooling: Architectures, Techniques, and Organizational Impact. *International Journal of Scientific Research in Science, Engineering and Technology*, 12(6), 600-606. DOI: <https://doi.org/10.32628/IJSRSET2613255>
- [30] Kuhnle, A., Röhrig, N., & Lanza, G. (2019). Autonomous order dispatching in the semiconductor industry using reinforcement learning. *Procedia Cirp*, 79, 391-396.
- [31] Lee, Y. H., & Lee, S. (2022). Deep reinforcement learning based scheduling within production plan in semiconductor fabrication. *Expert Systems with Applications*, 191, 116222.
- [32] Wang, J., Gao, P., Zheng, P., Zhang, J., & Ip, W. H. (2021). A fuzzy hierarchical reinforcement learning based scheduling method for semiconductor wafer manufacturing systems. *Journal of Manufacturing Systems*, 61, 239-248.